The Role of Protein Arginine Methylation in T-lymphocyte Activation



Vincent Geoghegan

Keble College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Abstract

T-lymphocytes are an essential cell type of the adaptive immune system. Due to their importance in immune responses and disorders, the molecular mechanisms leading to T-lymphocyte activation have been the subject of extensive research which has translated into important therapeutic developments. Early signalling events involving tyrosine phosphorylation are well characterised. However, later events involving other post-translational modifications are less well understood. Several studies have provided evidence suggesting a role for protein arginine methylation in T-lymphocyte activation. Arginine methylation is an essential post-translational modification in mammals and yet has not been extensively studied. No large scale analysis of arginine methylation sites has been performed. To gain insight into the role of protein arginine methylation in T-lymphocyte activation, the aims of this work were to:

1. Establish whether levels of arginine methylation are altered during T-lymphocyte activation

2. Use mass spectrometry based proteomics to identify arginine methylated proteins in the T-lymphocyte proteome

3. Further characterise an arginine methylated protein important to T-lymphocyte activation

Arginine methylation was found to be induced after long term (>20 hours) stimulation of primary T-lymphocytes. Large increases in the main protein arginine methyltransferase, PRMT1, were also observed. Enrichment and labelling methods were developed to detect arginine methylated peptides from T-lymphocytes by mass spectrometry. This resulted in the identification of 265 unique arginine methylation sites in 141 proteins. 204 of the methylation sites were novel and 103 of the proteins had not previously been described as arginine methylated. Individual arginine methylation sites were characterised before and after activation of T-lymphocytes, with some sites showing significant changes in abundance.

Among the novel arginine methylated proteins discovered were Dynamin II, WASp and WIPF1. These proteins are involved in re-organisation of the actin cytoskeleton at the immunological synapse formed between a T-lymphocyte and an antigen presenting cell. The functional consequences of the arginine methylation sites in WASp were characterised. WASp is essential for T-lymphocyte activation and some initial evidence showed that one of the arginine methylation sites is important for WASp activation.

Acknowledgements

I thank everyone in the Acuto lab for providing a supportive environment in which I could develop as a scientist and carry out this work. Adrian Gray and Kelly Molder ensured that the lab ran smoothly. I thank Hannes Eichner for all his help on the project in the later stages. It has been a pleasure to share an office with Wolfgang Paster from whom I have learnt a great deal and I also thank him for reading chapters of my thesis. I thank Thomas Uhlmann who supervised me day-to-day and taught me about mass spectrometry based proteomics. I am grateful for all his support and the time and effort he put in to train me. Without his attention and ideas the project would not have succeeded. It has been enjoyable working with Georgius Efstathiou and Kristin Katsch during the later period of my time in the lab and I thank Kristin Katsch for reading a chapter of my thesis.

Benjamin Thomas in proteomics has been an invaluable source of advice and information. I have learnt a great deal from him and his ideas were instrumental in the success of this project. Thank you for reading the largest chapter of my thesis. Thank you also to Gabriela Ridlova and Svenja Hester who worked very hard to ensure acquisition of good quality data. The Sir William Dunn School of Pathology has been an excellent environment in which to carry out research. I hope that it will continue to provide an outstanding environment for the education of scientists. Thank you to Simon Hunt and Peter Cook who provided much useful advice and support at many times during my thesis.

Ginger and Keiko Stainton have helped me to concentrate during long days of writing. Ginger Stainton in particular would often visit me to make sure I was working and shout at me if I wasn't. I was always pleased to hear him padding into the room to check on me.

Kirsty Stainton has been an endless source of good advice, good ideas and nutritional and moral support. I found our scientific discussions were always stimulating. Thank you for reading chapters of my thesis. I am extremely grateful for our time together without which my life would have been much duller.

Most of all I thank Oreste Acuto who has overseen the whole project and was willing to invest time and money to take me on as a student and train me to become a scientist. Thank you for all your support and guidance. Our scientific discussions were of great value to me and I hope that the investment you made was worthwhile.

This work was funded by the Biotechnology and Biological Sciences Research Council.

Contents

1 Introduction **1**

1.1 T-lymphocytes 1
1.2 Post translational modifications 7
1.3 Protein arginine methylation 9
1.4 Functional effects of arginine methylation 12
1.5 Phenotypes of PRMT knockout mice 14
1.6 Properties and regulation of the PRMTs 20
1.7 Reversibility of arginine methylation 26
1.8 Arginine methylation and transcription 28
1.9 Arginine methylation in RNA processing 31
1.10 The role of arginine methylation in determining the subcellular localisation of proteins 34
1.11 Arginine methylation in signal transduction 37
1.12 Arginine methylation in DNA repair 40
1.13 Arginine methylation in T-lymphocytes 42
1.14 Current state of knowledge of arginine methylation 45
1.15 Research objectives 47

2 Materials and Methods **48**

2.1 Culture of cell lines 48
2.2 Isolation of primary human PBMCs or CD4+ T-lymphocytes . 48
2.3 Flow cytometry analysis 51
2.4 Stimulation of cells with bead bound antibodies 52
2.5 Stimulation of cells with superantigen loaded Raji APCs . . . 53
2.6 Western blotting 53
2.7 SILAC labelling of cells 57
2.8 Immunoprecipitation of arginine methylated proteins 60
2.9 Extraction of protein for protease digestion 61
2.10 Strong cation exchange (SCX) chromatography of peptides . . 63
2.11 OFFgel isoelectric focusing of peptides 64

2.12 Desalting of peptides from strong cation exchange and OFFgel isoelectric focusing . . . 65
2.13 Weak cation exchange (WCX) chromatography . . . 66
2.14 Desalting of peptides prior to hydrophilic interaction liquid chromatography . . . 68
2.15 Hydrophilic interaction liquid chromatography (HILIC) . . . 70
2.16 Mass spectrometry data acquisition . . . 71
2.17 Mass spectrometry data analysis . . . 71
2.18 Cloning and mutagenesis of the WASp gene . . . 74
2.19 Transient transfection of Jurkat Clone 20 cells . . . 76
2.20 Pull-down of Strep-tagged WASp . . . 77

3 Induction of arginine methylation during T-lymphocyte activation 78
3.1 Introduction . . . 78
3.1.1 T-lymphocyte activation . . . 78
3.1.2 Signalling from the TCR and CD28 co-receptor . . . 79
3.1.3 T-lymphocyte activation *in vivo* . . . 81
3.1.4 Transcriptional regulation in T-lymphocyte activation and differentiation . . . 83
3.1.5 Post-transcriptional regulation in T-lymphocyte activation and differentiation . . . 84
3.1.6 Alternative splicing in T-lymphocyte activation and differentiation . . . 86
3.1.7 Aims . . . 89
3.2 Results . . . 90
3.2.1 Short term stimulation of T-lymphocytes . . . 90
3.2.2 Long term stimulation of T-lymphocytes . . . 91
3.2.3 Long term stimulation of CD4+ T-lymphocytes . . . 96
3.3 Discussion . . . 100

4 Discovering arginine methylated proteins in T-lymphocytes 108
4.1 Introduction . . . 108
4.1.1 Techniques for the discovery of arginine methylation sites 108
4.1.2 Mass spectrometry based proteomics . . . 111
4.1.3 Ionization . . . 112
4.1.4 Mass analysers . . . 114
4.1.5 Fragmentation of peptides . . . 118
4.1.6 Interpretation of mass spectra . . . 120
4.1.7 Detecting post-translational modifications by mass spectrometry . . . 124

4.1.8 Identification of arginine methylation sites by mass spectrometry . . . 125
4.1.9 Using mass spectrometry to identify arginine methylated proteins in T-lymphocytes . . . 130
4.1.10 Aims . . . 132
4.2 Results . . . 133
4.2.1 Applying SILAC to primary T-lymphocytes . . . 133
4.2.2 Using an anti-methylarginine antibody (7E6) to enrich for arginine methylated peptides . . . 135
4.2.3 Alternative strategies for enriching arginine methylation sites prior to mass spectrometry . . . 136
4.2.4 Enrichment by strong cation exchange chromatography 138
4.2.5 Enrichment by isoelectric focusing . . . 140
4.2.6 Enrichment by weak cation exchange chromatography . 141
4.2.7 Enrichment by hydrophilic interaction liquid chromatography . . . 148
4.2.8 Using a combination of separation methods . . . 152
4.2.9 Analysis of arginine methylation sites discovered . . . 153
4.2.10 Reliability of identified methylation sites . . . 161
4.2.11 Improving heavy methyl SILAC . . . 163
4.2.12 Occupancy of arginine methylation sites . . . 165
4.2.13 Comparing arginine methylation sites in unstimulated and stimulated T-lymphocytes . . . 169
4.3 Discussion . . . 193
4.3.1 Labelling primary cells . . . 194
4.3.2 Enriching for arginine methylation . . . 195
4.3.3 Alternative approaches to enriching for arginine methylation sites . . . 201
4.3.4 Functional consequences of arginine methylation . . . 204
4.3.5 Changes in arginine methylation during T-lymphocyte activation . . . 209

5 Regulation of WASp activity by arginine methylation 212
5.1 Introduction . . . 212
5.1.1 WASp . . . 212
5.1.2 The role of WASp in T-lymphocyte function . . . 213
5.1.3 Regulation of WASp . . . 217
5.1.4 Aims . . . 220
5.2 Results . . . 221
5.2.1 Conservation of methylation sites . . . 221
5.2.2 PRMT responsible for WASp methylation . . . 222

5.2.3 Role of WASp methylation sites in interactions with binding partners 223
5.2.4 Role of WASp methylation sites in activation of WASp 225
5.3 Discussion 227

6 Discussion **233**
6.1 Future developments 235

List of Figures

1.1 T-lymphocyte development 2
1.2 Role of T-lymphocytes during an immune response 5
1.3 Post-translational modifications and proteome complexity . . . 9
1.4 Types of arginine methylation 11
1.5 The PRMTs 21
1.6 The link between adenosine deaminase deficiency and methylation 43

3.1 The immunological synapse between an antigen presenting cell and a T-lymphocyte 82
3.2 Short term stimulation of Jurkat Clone 20 T-lymphocytes . . . 92
3.3 Short term stimulation of PBMCs 93
3.4 Long term stimulation of Jurkat T-lymphocytes 94
3.5 Long term activation of PBMCs 95
3.6 Increases in levels of PRMT1 and PRMT5 in PBMCs 97
3.7 Increases in ubiquitination after stimulation of PBMCs 98
3.8 Stimulation of CD4+ T-lymphocytes 99

4.1 Modules used for mass spectrometry based proteomics 116
4.2 Fragmentation of peptides according to Roepstorff and Fohlman nomenclature 119
4.3 A typical mass spectrometry based proteomics workflow . . . 122
4.4 Heavy methyl SILAC 128
4.5 Example precursor and fragmentation spectra of an arginine methylated peptide 129
4.6 Labelling of primary human T-lymphocytes 134
4.7 Immunoprecipitation of arginine methylated proteins 135
4.8 Physico-chemical properties of known tryptic arginine methylated peptides 137
4.9 Strong Cation Exchange chromatography as a method to enrich for methylated peptides 139

4.10 OFFGEL isoelectric focusing to enrich for arginine methylated peptides . . . 141
4.11 Weak Cation Exchange based separation of basic peptides . . 142
4.12 Separation of peptides above and below an isoelectric point cut-off using WCX . . . 144
4.13 Enrichment of arginine methylated peptides using WCX at varying binding pH . . . 145
4.14 Amino acid analysis of peptides from WCX . . . 147
4.15 Hydrophilic Interaction Liquid Chromatography for enrichment of arginine methylated peptides . . . 149
4.16 Physico-chemical properties of chymotryptic arginine methylated peptides . . . 151
4.17 Analysis of arginine methylation sites identified in this study . 154
4.18 Gene ontology analysis of arginine methylated proteins identified 155
4.19 Overview of all methylated proteins identified . . . 157
4.20 Sequence logo representation of all arginine methylation sites discovered . . . 158
4.21 Amino acid frequencies at arginine methylation sites . . . 159
4.22 Average hydropathy scores for amino acids in PRxxxP motifs 161
4.23 Effect of methyl SILAC on FDR . . . 162
4.24 Isomethionine methyl SILAC . . . 164
4.25 Comparison of heavy and isomethionine methyl SILAC . . . 166
4.26 Occupancy of arginine methylation sites . . . 167
4.27 Labelling to compare arginine methylation in unstimulated and stimulated cells . . . 170
4.28 Changes in arginine methylation upon T-lymphocyte activation 175

5.1 Signalling from the TCR . . . 215
5.2 Regulation of WASp . . . 218
5.3 Conservation of WASp . . . 222
5.4 Screening for PRMT responsible for WASp methylation . . . 223
5.5 Effect of methylation on WASp binding partners . . . 224
5.6 Importance of WASp methylation sites for WASp phosphorylation . . . 225
5.7 Quantification of effect of methylation sites on WASp phosphorylation . . . 226

List of Tables

1.1 Summary of PRMT characteristics 16

2.1 Antibodies used for western blotting 56
2.2 Strong cation exchange chromatography profile 64
2.3 Chromatography profile for desalting peptides 69
2.4 Chromatography profile for HILIC 71
2.5 Primers used to clone WASp 74
2.6 PCR method used for cloning 74
2.7 General PCR reaction mix 75
2.8 Colony PCR reaction mix 75

4.1 Arginine methylation sites identified by the 7E6 antibody . . . 136
4.2 Physico-chemical properties of tryptic and chymotryptic peptides 152
4.3 Significant arginine methylation motifs found by motif-x . . . 160
4.4 Summary of alterations in arginine methylation upon T-lymphocyte activation 178
4.5 Methylation sites identified in this study 192

1. Introduction

1.1 T-lymphocytes

Thymus-dependent lymphocytes (T-lymphocytes) are a key cell type of the adaptive immune system in higher vertebrates. They recognise antigens derived from pathogens via the T-cell receptor (TCR), a cell surface expressed protein complex able to specifically recognise antigenic peptides bound to Major Histocompatibility Complex (pMHC). Development of T-lymphocytes begins in the bone marrow when a hematopoietic stem cell divides to produce a common lymphoid progenitor. Those lymphoid progenitors destined to become T-lymphocytes then migrate to the thymus, an organ located just above the heart. When lymphoid progenitors enter the thymus they do not express the TCR or any other markers present on mature T-lymphocytes. Within the thymus, lymphoid progenitors differentiate into mature T-lymphocytes in a process guided by interactions with epithelial cells of the thymic stroma (Figure 1.1). Lymphoid progenitors first differentiate into double negative thymocytes which do not express the TCR or either of the co-receptors CD4 and CD8. A pre-TCR is then expressed together with both CD4 and CD8 making the cell a double positive thymocyte. Antigen independent signalling from the pre-TCR provides survival signals to the developing thymocyte. The

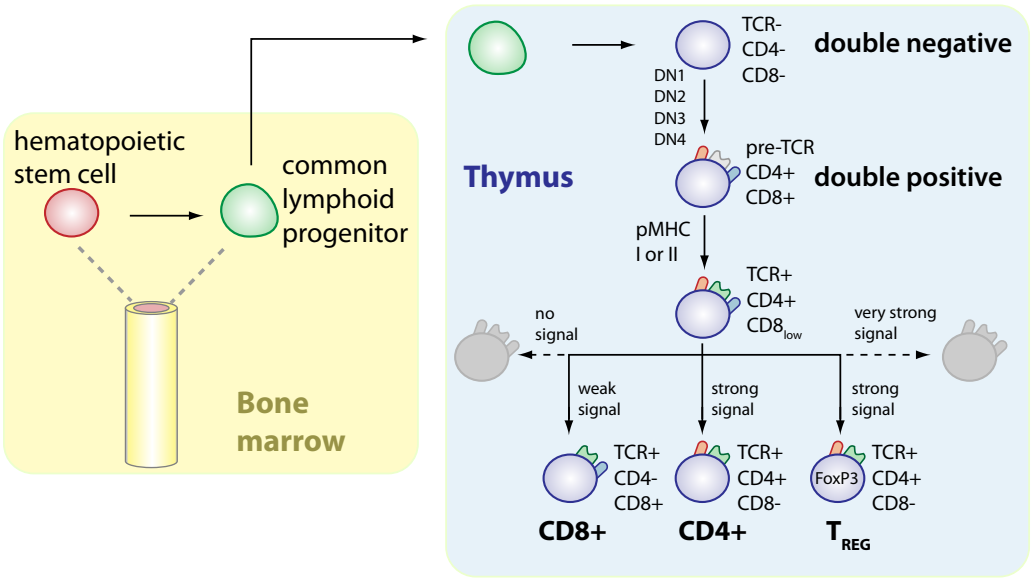


Figure 1.1: T-lymphocyte development. In the bone marrow a hematopoietic stem cell divides to produce a common lymphoid progenitor. Some lymphoid progenitor cells migrate to the thymus. Once in the thymus, development of T-lymphocytes is guided by interactions with epithelial cells. Thymocytes pass through 4 double negative stages (DN1-4) after which pre-TCR, CD4 and CD8 are expressed. These double positive thymocytes interact with epithelial cells expressing peptide-MHC (pMHC) class I or II leading to initiation of TCR signalling and downregulation of CD8. Signal intensity from subsequent interactions with pMHC determine the fate of the cell. No signalling leads to death. A weak signal leads to development of a CD8+ single positive T-lymphocyte. Strong signalling leads to a CD4+ T-lymphocyte. A stronger signal generates a regulatory T-lymphocyte expressing FoxP3. Very strong signals lead to death. Once mature, T-lymphocyte exit the thymus and enter secondary lymphoid tissue.

co-receptors CD4 and CD8 bind different classes of MHC molecule. CD4 binds MHC class II, present on professional antigen presenting cells of the immune system, while CD8 binds MHC class I, which is expressed by nearly all cells in the body. TCR diversity is generated through DNA rearrangement generating many possible permutations of the gene segments coding for the TCR. Due to the large number of possible permutations of gene seg-

ments, the potential number of unique T-cell receptor specificities has been estimated at 10^{13}. Once this rearrangement has taken place, the pre-TCR is replaced by a fully functional TCR and the thymocyte can then undergo selection based on the specificity of its TCR. Cells in the thymus present self peptides bound to MHC. If the TCR on a thymocyte recognises such self peptides with low or intermediate affinity, the thymocyte will undergo positive selection and become either a CD4 or CD8 single positive T-lymphocyte. If the TCR recognises the self peptide with high affinity, it may develop into a regulatory T-lymphocyte (T_{REG}) or be eliminated by negative selection.

Mature CD4 and CD8 single positive T-lymphocytes circulate through the secondary lymphoid organs via the bloodstream. When they encounter a peptide-MHC complex on an antigen presenting cell (usually residing in the lymphoid organs) which can be specifically recognised by the TCR, the T-lymphocytes become activated (Figure 1.2). Activation of a T-lymphocyte requires strong sustained signalling through the TCR and also signalling through the co-receptor CD28. Activation of T-lymphocytes leads to marked changes in cell morphology and behaviour. T-lymphocytes become larger, more metabolically active and begin to proliferate rapidly. This differentiation process is driven by activation and up-regulation of transcription factors and changes to the alternative splicing of mRNA. Many genes become up-regulated including those coding for cytokines, which are secreted signalling proteins of the immune system. One of the first cytokines to be produced by activated T-lymphocytes is interleukin-2 (IL-2), an important growth factor driving the proliferation of activated T-lymphocytes.

Cytokines secreted by antigen presenting cells, together with the nature of

the signalling from the TCR, determine how the activated T-lymphocyte differentiates. CD4 T-lymphocytes can differentiate into any one of four major subsets currently known, T_H1, T_H2, T_H17 and T_{REG} cells. The precise conditions under which each T-lymphocyte subset arises has not been completely elucidated. It appears that the cytokine IFN-γpromotes differentiation into T_H1 cells, IL-25 may lead to T_H2 differentiation and TGF-βpromotes differentiation to T_H17 and T_{REG} cells. Each of the T-lymphocyte subsets produces a distinct set of cytokines upon activation and responds to different cytokines and migratory signals. Consequently, the sub-types act on different cells of the immune system and have different roles as part of the immune response. T_H1 cells, characterised by secretion of IFN- γ, are responsible for the destruction of pathogenic micro-organisms within tissues through the activation of macrophages, NK cells, dendritic cells and CD8 T-lymphocytes. T_H2 cells, characterised by secretion of IL-4, recognise peptides presented by B-lymphocytes and provide signals (cytokines) leading to activation and differentiation of these cells into antibody secreting plasma cells. In addition, T_H2 co-ordinate eosinophils, basophils and mast cells to provide protection from larger extracellular pathogens which cannot be phagocytosed, such as parasitic worms. T_H17 cells, identified by secretion of IL-17, recruit neutrophils to local acute inflammatory responses against extracellular bacteria and fungi not dealt with by the Th1/Th2 response. Regulatory T-lymphocytes (T_{REG}), secrete IL-10 and suppress the activation of other T-lymphocytes to prevent damage to self by an excessively strong immune response. Differentiation of CD8 T-lymphocytes upon activation has been less well characterised. CD8 T-lymphocytes provide protection against viruses

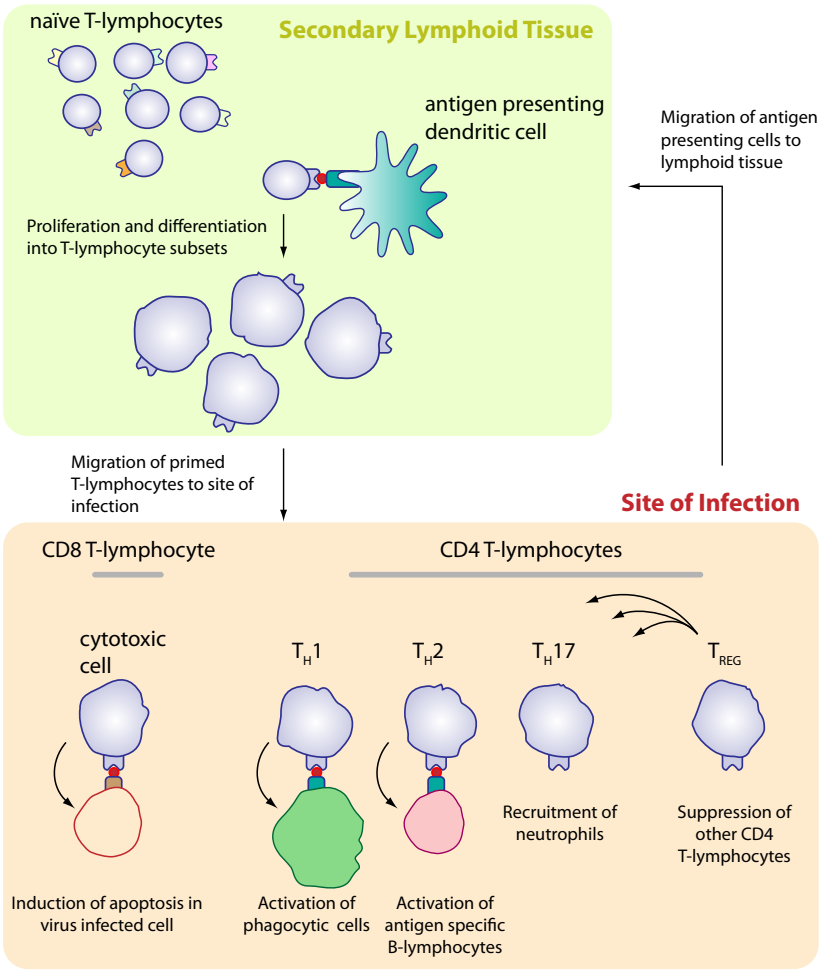


Figure 1.2: Activation and differentiation of T-lymphocytes during an immune response. The immune response begins with cells of the innate immune system recognising foreign antigens on the surface of a pathogen. This leads to activation of antigen presenting cells and inflammation at the site of infection. The inflammatory response increases the permeability of blood vessels so that plasma leaks out at the site of infection. This in turn causes increases in the flow from the site of infection to the lymphatic system resulting in rapid migration of antigen presenting cells to secondary lymphoid tissue. Naïve T-lymphocytes bearing T-cell receptors specific for pathogen derived peptides are activated and proliferate and differentiate into distinct sub-types depending on signals from the antigen presenting cell. These primed T-lymphocytes then migrate to the site of infection where they carry out their specific role.

and intracellular proliferating bacteria. Such pathogens remain inside cells and therefore cannot readily be recognised and processed for antigen presentation. However CD8 T-lymphocytes are able to recognise pathogen derived peptides bound to MHC class I on infected cells. When CD8 T-lymphocytes become activated, they differentiate into IFN-γ producing cells. IFN-γ increases antigen presentation by infected cells, facilitating recognition by CD8 T-lymphocytes. Activated CD8 T-lymphocytes act directly on the infected cell, delivering cytotoxic proteins thereby eliminating the compromised cell.

Only 1 in 100,000 T-lymphocytes will be able to recognise any given antigen. In order to mount an effective response therefore, the activated T-lymphocyte divides rapidly leading to a large expansion (10,000-50,000 fold) in cell numbers. This expansion, or priming, takes place in the secondary lymphoid tissue when a T-lymphocyte specifically detects pathogen derived antigen on the surface of a dendritic cell. Once naïve T-lymphocytes have differentiated into effector T-lymphocytes, they migrate to the sites of infection where they secrete cytokines and co-ordinate the immune response according to their subtype.

Activation of T-lymphocytes also causes up-regulation of the pro-apoptotic protein Bim, leading to activation induced cell death. This is a mechanism for limiting the T-lymphocyte response and means that only a small proportion of the antigen specific T-lymphocytes remain once the infection is cleared. The remaining cells are memory T-lymphocytes which occur in two types, effector memory T-lymphocytes (T_{EM}) and central memory T-lymphocytes (T_{CM}). T_{EM} cells can respond immediately to any re-exposure to the same antigen and rapidly differentiate into effector T-lymphocytes, while T_{CM} cells

first have to proliferate and then differentiate to produce more effector cells upon re-exposure.

In order for a T-lymphocyte to become activated, the information that the TCR has recognised a cognate pMHC complex must be transmitted to the nucleus leading to an alteration in the genes transcribed. There are several ways in which this information is conveyed. Triggering of the TCR causes a signal which begins as a cascade of tyrosine phosphorylation on numerous downstream proteins (described in more detail in Section 3.1.2). The tyrosine phosphorylation signal then diverges into other types of phosphorylation (serine and threonine) as well changes in concentrations of small molecules, for example release of Ca^{2+} from the endoplasmic reticulum. The increase in $Ca^(2+)$ activates calmodulin which in turn activates the phosphatase calcineurin. Calcineurin dephosphorylates the NFAT transcription factors allowing them to enter the nucleus where they lead to transcription of genes necessary for activation (Oh-hora and Rao, 2008). Thus, during T-lymphocyte activation, post-translational modifications such as tyrosine phosphorylation play a crucial role in transmitting signals from the TCR to the nucleus.

1.2 Post translational modifications

After the human genome was sequenced, it became apparent that it contained fewer genes than many had expected given our behavioural and biological complexity. Although the exact number of genes is still unknown, several

studies have estimated the number to be between 20,000 and 25,000. This number is not significantly more than species of lower apparent complexity, such as *Caenorhabditis elegans* which has around 20,000 genes, and is significantly less than other species of lower apparent complexity, for example *Oryza sativa* (rice) which has 41,000 genes (Sterck et al., 2007). This has given rise to the N-value paradox (Claverie, 2001), which states that there is poor correlation between the number of genes and the complexity of organisms. While there is no objective measure of complexity, it is clear that it is not so important how many genes are present in the genome, rather it is what these genes code for that matters. Genes which encode proteins capable of modifying the products of other genes should lead to rapid expansion of molecular diversity in the proteome (Figure 1.3). This molecular diversity is important for the evolution and function of organisms as it is more likely there will be an appropriately specialised protein to carry out any required task.

Proteins can be co- or post-translationally modified. There are at least 280 different types of post-translational modification (PTM) described in UniProt (www.uniprot.org). PTMs expand the diversity of physico-chemical properties offered by amino acid residues, thus leading to great variety in the proteome. This molecular diversity offered by PTMs can be regulated, generating signalling pathways so that information can be transmitted into and out of the cell. PTMs are therefore a major determinant of organism complexity and their study can lead to the discovery of new signalling pathways. It is difficult to predict modification sites from studying the genome, instead they must be directly analysed to aid in our understanding of how

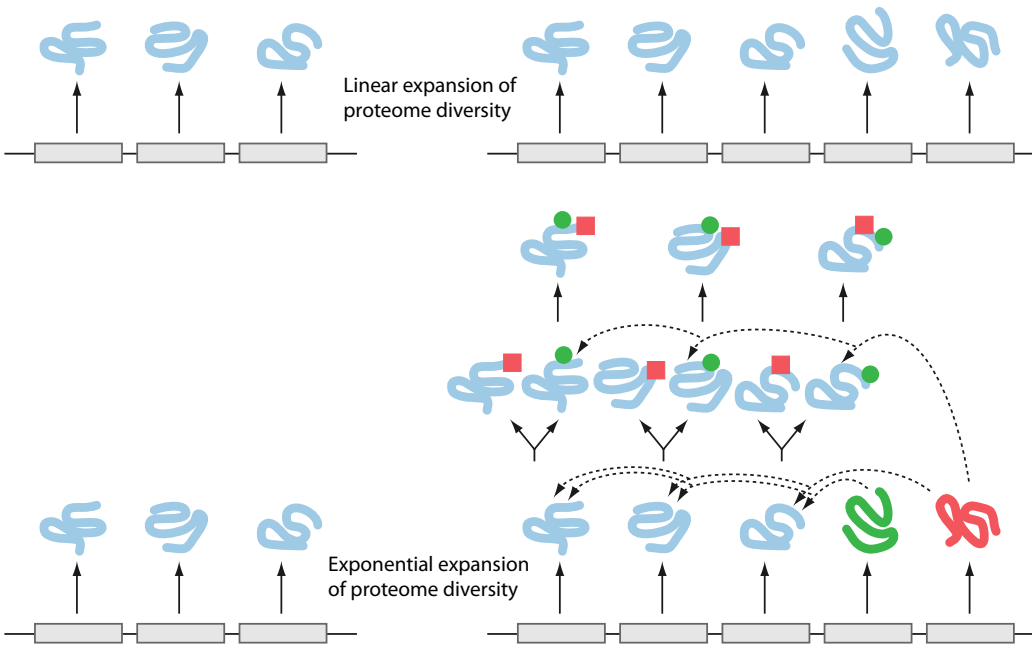


Figure 1.3: Exponential increase in proteome diversity through post-translational modifications (PTMs). A linear increase in proteome diversity is the result of simple expansion of the genome. An exponential expansion of proteome diversity is achieved by expanding the proteome with genes coding for proteins able to modify other proteins either directly as in the case of PTMs, or indirectly in the case of alternative splicing or editing of mRNA.

cells function at a molecular level.

1.3 Protein arginine methylation

The methylation of arginine residues is a PTM discovered over 40 years ago (Paik and Kim, 1967), and is thought to be the most prevalent form of protein methylation (Najbauer et al., 1967). 4 different forms of methylated arginine have been described: ω-N^{G}-monomethylarginine (MMA), ω-N^{G},

N^{G}-symmetric dimethylarginine (SDMA), ω-N^{G}, N'^{G}-asymmetric dimethylarginine (ADMA) and δ-N^{G}-monomethylarginine, the latter having only been observed in yeast (Zobel-Thropp et al., 1998)(Figure 1.4). ADMA accounts for 50% of methylarginine found in mammals, with SDMA and MMA present at 20-50% of ADMA levels. MMA is generally regarded as a reaction intermediate. In mammals, the enzymes responsible for the methylation of arginine are the *p*rotein a*r*ginine *m*ethyl*t*ransferases (PRMTs), a group of enzymes which catalyse the transfer of a methyl group from the cofactor *S*-adenosylmethionine (SAM) to the guanidinium group of arginine, generating *S*-adenosyl-L-homocysteine as a reaction side product. PRMT activity was first described in 1967 (Paik and Kim, 1967) and the first genes to code for PRMTs were described in 1996 (Lin et al., 1996). To date, 9 members of the PRMT family are known (PRMT1-9), all characterized by the presence of seven beta strand motifs common to many methyltransferases (Katz et al., 2003), as well as "double E" and "THW" motifs which are unique to the PRMT family and important for SAM binding and catalysis (Cheng et al., 2005). A further two members of the arginine methyltransferase family, FBX010 and FBX011, are proposed to exist based on weak sequence similarity to PRMTs1-9, however the activity of these gene products has not been confirmed.

Arginine methyltransferases are subdivided according to the type of methylated arginine they generate, no enzyme has so far been found to form both ADMA and SDMA. Type I PRMTs catalyse the formation of ADMA, Type II the formation of SDMA, Type III the formation of MMA and Type IV enzymes result in the formation of δ-*N*-monomethylarginine (RMT2 in yeast).

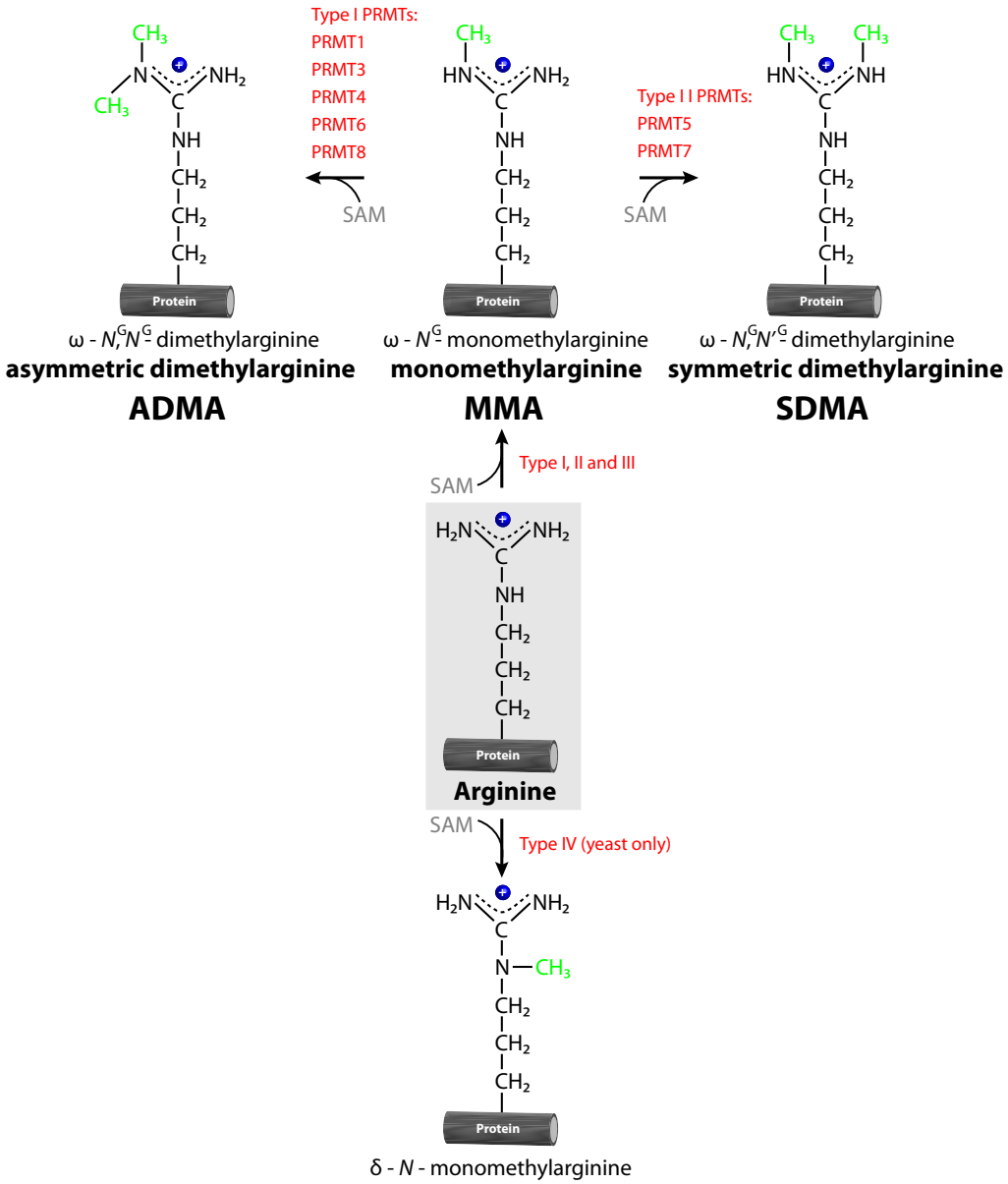


Figure 1.4: Protein arginine methylation. Currently four products resulting from methylation of arginine are known. The protein arginine methyltransferases (PRMTs) use *S*-adenosyl methionine (SAM) as the methyl group donor. Type I PRMTs form asymmetric dimethylarginine, Type II PRMTs form symmetric dimethylarginine, Type III PRMTs form monomethylarginine only and Type IV enzymes formδ-monomethylarginine.

PRMTs 1, 2, 3, 4, 6, 8 are Type I enzymes, PRMTs 5, 7 are Type II and PRMT7 is also a type III enzyme, generating MMA as the final product (Bedford and Clarke, 2009; Lakowski and Frankel, 2009). The methyltransferase activity of PRMT9 is currently unknown.

1.4 Functional effects of arginine methylation

Arginine is the most basic of the 20 standard amino acids, with a pK value for the guanidino side chain of 12.48 (Voet and Voet, 2004). The positively charged planar guanidinium group of arginine can act as a hydrogen bond donor for up to 5 hydrogen bonds (Borders et al., 1994). Due to these distinct properties, arginine residues are often found to interact with negatively charged molecules such as nucleic acids and commonly form single or double hydrogen bonded salt bridges with aspartate or glutamate residues. Arginine is thus a key functional residue at interacting interfaces of proteins. Although the methylation of arginine does not alter its charge at physiological pH, for each methyl group added, a potential hydrogen bond donor is lost. In addition, methylation will increase the size and hydrophobicity of the arginine side chain. Methylation can therefore disrupt interactions with arginine by removing hydrogen bonds, or it can lead to enhanced interactions by increasing hydrophobicity and promoting stacking with DNA/RNA bases (Hyun et al., 2008). Specific examples include arginine methylation of Sam68 which disrupts binding to SH3 domains and methylation of PIWIL1 which is required for interaction with SND1 (Liu et al., 2010). In principle,

up to 5 methyl groups could be added to the guanidino side chain of arginine, however only mono- and dimethylated products have ever been observed.

One type of protein domain, the Tudor domain, has so far been shown to selectively recognise all three methylated forms of arginine found in mammals. The Tudor domain is a 60 amino acid domain comprised of an antiparallel β sheet containing 5 β strands, forming a barrel like shape. Methylated lysine or arginine residues are recognised by an aromatic pocket (Chen et al., 2011). Around 30 proteins are thought to contain Tudor domains in the mammalian genome and can be divided into two groups depending on whether they bind methylated arginine or methylated lysine, however, the specificity of domains for methylated arginine or lysine is not clear. It has been suggested that methylarginine binding Tudor domains have a narrower aromatic cage, thus favouring the planar guanidinium group of arginine over the butylammonium group of lysine (Liu et al., 2012). In the case of other domains shown to recognise methylarginine, specificity is not absolute. For example, the WD40 domain of WDR5, which was known to bind methylated lysine 4 of histone 3, has also been shown to bind symmetrically dimethylated arginine in histone 3 (Migliori et al., 2012).

Three Tudor domain containing proteins have been shown to specifically recognise methylated arginine, SMN (Survival motor neuron protein), TDRD3 (Tudor domain containing protein 3) and SPF30 (Splicing factor 30 kDa). SMN appears to favour SDMA although it has been shown to bind ADMA generated by PRMT4, while TDRD3 preferentially binds ADMA (Liu et al., 2012; Cheng et al., 2007). The affinities for all three methylarginine binders is rather low, and binding generally occurs to methylarginine within

arginine/glycine rich (GAR) motifs, although some studies have shown that TDRD3 and SMN can sometimes recognise methylarginine outside of the normal GAR context (Liu et al., 2012), for instance TDRD3 has been shown to recognise methylated arginine in histone tails (Yang et al., 2010). The output from the binding of these methylarginine "readers" is poorly understood. In the case of TDRD3, binding to methylated histone tails results in recruitment to promotors and transcriptional activation (Yang et al., 2010).

1.5 Phenotypes of PRMT knockout mice

Ablating gene function in mice has proven to be a useful tool to assess the importance of a protein to development and function of a whole mammalian organism (Table 1.1). This approach has been especially useful in the field of immunology, where the effect of ablating or altering gene products on an entire immune system can be studied. The first PRMT to be knocked out in a mouse model was PRMT1 (Pawlak et al., 2000). This resulted in embryonic lethality shortly after implantation, with the mice failing to develop beyond embryonic day 6.5 (E6.5). Interestingly, embryonic stem (ES) cells derived from PRMT1 null mice could be cultured and grew at a similar rate to wild type ES cells. However, this could be due to the fact that hypomorphic alleles of PRMT1 were generated where levels of PRMT1 are reduced but not eliminated so that sufficient PRMT1 remains to carry out essential cellular functions (Côté et al., 2003). In *Saccharomyces cerevisiae*, creation of a HMT1 (the main PRMT in yeast) null mutant does not result in any obvious

defect in cellular viability (Henry and Silver 1996), suggesting that arginine methylation is important to cellular differentiation and that differentiated cells might be more sensitive to levels of PRMT compared to pluripotent cells. In support of this, PRMT1 is essential for the viability of mouse embryonic fibroblasts (MEFs), contrasting the situation in ES cells (Yu et al., 2009). Measurement of ADMA levels in PRMT1 deficient cells from PRMT1-/- mice showed that PRMT1 is responsible for generating just over 50% of ADMA produced in the cell and accounts for 85% of methyltransferase activity. In another study, PRMT1 was immuno-depleted from RAT1 fibroblast cells, levels of methylation were assayed and it was demonstrated that PRMT1 was responsible for 85% of the methylation activity in these cells (Tang et al., 2000). Thus it appears that PRMT1 is the major arginine methyltransferase.

PRMT1 seems to have an important role in maintenance of genome integrity. This is clearly illustrated by a conditional knockout of PRMT1 in mouse embryonic fibroblasts which gives rise to spontaneous DNA damage and chromosomal aberrations such as structural re-arrangements and polyploidy. A PRMT1 deficiency in these cells results in failure to activate the G2/M and G1/S checkpoints in response to DNA damage (Yu et al., 2009). Deletion of PRMT2 expression in mice led to no obvious phenotypic change under normal conditions. Mice were viable and reproduced, with no observable malignancies over 12 months (Yoshimoto et al., 2006). However, PRMT2 null mice did display an abnormal response to vascular injury, as this led to more rapid cellular proliferation leading to abnormal intimal hyperplasia. The doubling times of PRMT2 null MEFs were not different to wild type cells, however PRMT2-/- cells did enter S phase earlier than wild-

PRMT	Localisation	Knockout mouse	Notes
PRMT1	cytoplasm + nucleus	early embryonic lethality	Responsible for majority of arginine methyltransferase activity
PRMT2	nucleus	abnormal response to vascular injury	Only weak activity has been demonstrated for this enzyme
PRMT3	cytoplasm	embryos are smaller	Associated with the ribosome
PRMT4	cytoplasm + nucleus	late embryonic lethality	Ablation in mice causes block in early thymocyte development
PRMT5	cytoplasm + nucleus	early embryonic lethality	Generates SDMA
PRMT6	nucleus	mice are normal	Automethylated
PRMT7	cytoplasm + nucleus	partial lethality	Generates SDMA
PRMT8	membrane	decreased IgG levels	N-terminally myristoylated
PRMT9	-	-	Not yet characterised

Table 1.1: Summary of PRMT characteristics.

type cells after serum exposure. Another study showed that NF-kB activity was upregulated in PRMT2-/- MEF cells, and that these cells showed an increased susceptibility to apoptosis compared to wild type cells (Ganesh et al., 2006). A later study on PRMT2-/- mice demonstrated that PRMT2 is a regulator of the inflammatory response after injection of Lipopolysaccharide (LPS) into mice. PRMT2-/- mice produced elevated concentrations of the inflammatory cytokines TNF-α and IL-6, likely due to an upregulation of NF- κ B dependent transcription (Dalloneau et al., 2011).

PRMT3 knock out mice are 75% the size of wild type embryos at E13.5, but reach a normal size in adulthood. Since PRMT3 is a ribosomal associated protein and methylates ribosome proteins such as 40S ribosomal protein S2, this size defect could be the result of aberrant translation of certain transcripts (Swiercz et al., 2007).

PRMT4 (also known as CARM1-coactivator-associated arginine methyltransferase 1) has also been genetically targeted in mice, resulting in late embryonic lethality or death shortly after birth (Yadav et al., 2003). PRMT4-/- embryos are 60% the size of wild type, and the lungs do not inflate after birth. In addition, the loss of even one copy of the PRMT4 gene in PRMT4+/- mice is enough to cause some lethality between birth and weaning. A further study revealed that total numbers of all major subsets of thymocytes in thymi from PRMT4-/- mice were substantially reduced, giving an overall 5-10 fold reduction in total cell numbers from PRMT4-/- thymi compared to wild type (Kim et al., 2004). The absence of PRMT4 causes a block in early thymocyte development at the DN1 to DN2 developmental stage, leading to an increased proportion of DN1 cells and accounting for the reduction in all subsequent

thymocyte developmental stages. Although peripheral T-lymphocytes were not analysed, since thymocytes are the progenitors of mature T-lymphocytes, a reduction in peripheral T-lymphocytes might also be expected in PRMT4-/- mice.

The block in T-lymphocyte development caused by loss of PRMT4 indicated that this arginine methyltransferase could be critical for proper T-lymphocyte function. TARPP (thymocyte cyclic AMP-regulated phosphoprotein) was identified as a substrate for PRMT4, its expression correlates with rearrangement of the TCR genes, being mostly restricted to thymocyte progenitors (Kisielow et al., 2001). This provided a possible molecular link between loss of PRMT4 and the blockade in early T-lymphocyte development. However the actual function of TARPP is unknown and the block in thymocyte development could instead have been caused by a defect in thymic epithelial cells which are essential for this process. Indeed, PRMT4 is a ubiquitously expressed enzyme with a role in general transcription processes, its absence therefore likely affects numerous proteins in numerous cell types and its importance is unlikely to be restricted to thymocytes.

A PRMT5 null mouse is lethal at an early embryonic stage, with no normal structures being observed at E6.5 (Tee et al., 2010). Unlike for PRMT1-/- mice, no PRMT5 null embryonic stem cells could be derived as the inner cell mass of pluripotent cells within the blastocyst failed to develop at E3.5. Thus PRMT5 appears to be essential for maintenance of pluripotency.

PRMT6 has recently been ablated in mice, leading to no obvious phenotype. MEFs were isolated from these PRMT6-/- mice and showed an increase in cellular senescence, thought to be due to elevated levels of p53

(Neault et al., 2012).

Results of targeting the remaining PRMTs (PRMTs 7-9) have not been published. However a search of the mouse resources portal from the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/mouseportal/) reveals that some of the other PRMTs have been knocked down in mice. PRMT7 null mice show partial lethality before weaning and erythrocyte micronuclei were observed, suggesting chromosomal instability and PRMT8 null mice had decreased IgG1 levels but were otherwise normal.

While gene targeting in mice has proven very useful, perturbing complex systems by deleting important components and analysing the results can often only reveal very general effects and it is difficult to precisely pinpoint which molecular pathways have been directly affected. Moreover, in the case of the PRMT knockout mice the possibility that the effects observed are due to a methyltransferase independent activity of the PRMTs cannot be ruled out. To address this concern, a PRMT4 enzyme dead knockin was created (Kim et al., 2009). Mice homozygous for the enzyme dead PRMT4 showed the same phenotype as the PRMT4 knockout mice, demonstrating that the lack of arginine methyltransferase activity is responsible for the defects observed in the knockout. Interestingly though, transcriptional reporter assays showed that the enzyme dead form of PRMT4 possessed some residual activity.

1.6 Properties and regulation of the PRMTs

The PRMT family of enzymes likely arose very early on during eukaryotic evolution since homologues have been found in unicellular eukaryotes such as *Plasmodium* but not in the prokaryote or archaebacterial lineages (Krause et al., 2007). PRMT1 and PRMT5 are likely to have been the first PRMTs to appear as they are found in nearly all eukaryotes (apart from *Giardia lamblia* where interestingly no PRMTs are found). In humans, all PRMTs are ubiquitously expressed with the exception of PRMT8, which is only present in neurons in the brain. Ranging in molecular weight from 40kDa to 80kDa, the PRMTs share features at the sequence and structural level (Figure 1.5). All share a conserved catalytic domain consisting of 310 amino acids which contains 5 characteristic motifs known as motif I, post-I, II, III and a THW loop. Motif I, post-I, II and III form part of the SAM binding pocket (Zhang et al., 2000). The structures of all arginine methyltransferases known to date (Hmt1, PRMT1,3,4,5) show a conserved core which can be broadly divided into three regions: the methyltransferase domain, a seven stranded β-barrel and a dimerization arm. Several PRMTs are active as homo-oligomers and can be automethylated, dimerisation may be required for processivity since the monomethylated product could be released to the second enzyme in the dimer for the final methylation (Cheng et al., 2005). Outside the conserved catalytic core, the PRMTs have variable N-termini which may determine the specificity of substrate recognition. Supporting this is evidence from a study on *S. cerevisiae* Rmt1 showing that a single mutation in the N-terminus alters substrate specificity (McBride et al., 2000).

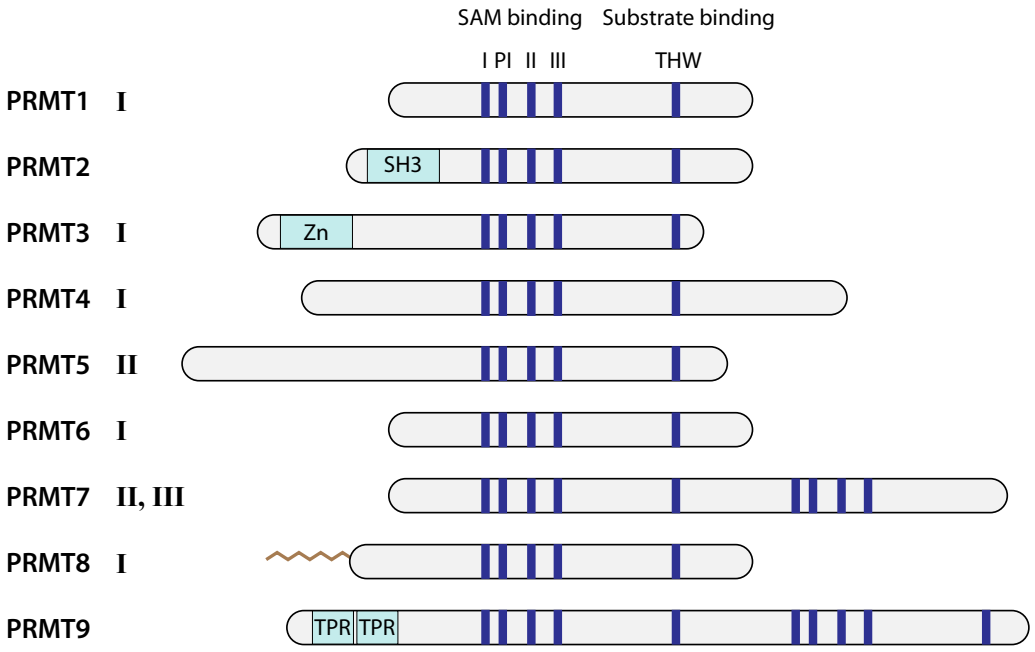


Figure 1.5: The human protein arginine methyltransferases (PRMTs). Nine PRMTs are currently known sharing conserved sequence elements (blue bars). All have conserved *S*-adenosyl methionine (SAM) binding domains consisting of motifs I, post-I (PI), II and III. The THW loop is found within the substrate binding region. The PRMTs have variable N-termini. PRMT2 contains an N terminal SH3 domain, and PRMT3 an N-terminal Zinc finger domain. PRMT8 is N-terminally myristoylated and PRMT9 contains two tetratricopeptide repeats (TPR).

PRMT1 is a 40kDa protein, but gel filtration of PRMT1 in cell extracts showed that it forms complexes with sizes from 275 to 450kDa, representing oligomeric states or the presence of other proteins necessary for the correct localisation and regulation of activity. In the cell, PRMT1 localises to both the cytoplasm and nucleus, depending on the splicing form, and is able to methylate arginine in a variety of sequence contexts, showing a preference for arginine in glycine arginine rich (GAR) regions. The crystal structure of PRMT1 in complex with *S*-adenosyl-L-homocysteine and a GAR motif has

been solved, revealing the presence of 3 peptide binding channels explaining the ability of PRMT1 to recognise a variety of sequence contexts. Three binding proteins, BTG1, TIS2/BTG2 and hCAF1 regulate PRMT1 activity. BTG1 and TIS2/BTG2 lead to enhancement of activity, whilst the effect of hCAF1 is dependent on the methylation substrate (Robin-Lespinasse et al. 2007). PRMT2 has also been shown to bind to PRMT1 and lead to augmented activity of the latter, thus complexes between PRMTs appear to be important for regulating activity (Pak et al., 2011).

PRMT2 is distinguished from the other PRMTs by the presence of an SH3 domain at the N-terminus. It appears to be a predominantly nuclear localised enzyme. Very weak activity for PRMT2 was claimed in one study, which also showed evidence that PRMT2 was a type I arginine methyltransferase, producing MMA and ADMA, and being able to methylate histone 4. However one cannot rule out the presence of a co-purifying methyltransferase (Lakowski and Frankel, 2009).

PRMT3 is a cytosolically localised enzyme whose distinguishing feature is the presence of a zinc finger domain at its N-terminus which is necessary for recognition of its main substrate, the ribosomal protein rpS2. Only very low activity has been shown for this methyltransferase (Herrmann et al., 2009). Activity of PRMT3 is inhibited by the tumor suppressor DAL-1.

PRMT4 is localised to either the cytoplasm or nucleus depending on the cell line and stage of the cell cycle. PRMT4 is itself subject to automethylation on R551 and phosphorylation on serine. PRMT4 was observed to exist almost exclusively in a form automethylated on R551 when expressed in HEK293T cells. Automethylation is not necessary for methyltransferase

activity, however it does appear to be required for the proper transcriptional activation and splicing of certain target genes (Kuhn et al., 2011). The phosphorylation of PRMT4 on S217 during mitosis appears to be a major switch for controlling PRMT4 activity, leading to abolishment of SAM binding and relocation to the cytoplasm (Feng et al., 2009). This switch may prevent PRMT4 modifying cytoplasmic proteins during mitosis, or it may simply be that PRMT4 activity is not required because transcriptional activity is low during mitosis. Furthermore, the phosphorylated serine is conserved in all type I PRMTs, indicating that this could be a requlatory mechanism for these other methyltransferases. PRMT4 is a component of the nucleosomal methylation activator complex, a complex of at least 10 proteins including chromatin remodelling factors. As part of this complex, PRMT4 preferentially methylates nucleosomal histones over free histones. As a recombinant protein however, the reverse is true, suggesting that much like PRMT1, PRMT4 is subject to regulation by binding proteins (Xu et al., 2004).

The best characterised type II enzyme, PRMT5, is present in both the nucleus and cytoplasm. The crystal structure of PRMT5 was recently solved providing structural insight into the mechanism responsible for the specificity of SDMA production. It appears that a phenylalanine residue (Phe379) extends towards one side of the methyl-accepting arginine residue, thus sterically restricting the acceptance of methyl groups to the opposite side resulting in formation of SDMA. Altering Phe379 to methionine led to relaxed specificity, with both ADMA and SDMA being formed (Sun et al., 2011). In the cytoplasm, PRMT5 forms a complex (termed the methylosome) with plCln and WD45 which is involved in snRNP biogenesis (Friesen et al. 2001),

PRMT7 also appears to act together with PRMT5 in this role (Gonsalvez et al., 2007). In the nucleus, the function of PRMT5 is regulated by binding proteins such as COPR5 and hSWI/SNF chromatin remodelling proteins.

PRMT6 is a nuclear localised enzyme which also undergoes automethylation, although the significance of this is unclear. PRMT6 has been the subject of studies on the enzymatic mechanism of arginine methylation. An initial investigation suggested that PRMT6 methylates substrates via an ordered mechanism, releasing the monomethylated intermediate followed by SAH, meaning processive methylation is not possible (Lakowski and Frankel, 2009). However a subsequent study has challenged this, claiming that a complex comprising of enzyme, SAM, substrate-MMA can form after the first methylation without release of the monomethylated intermediate, thus making processive methylation possible (Obianyo and Thompson, 2012). *In vitro*, PRMT6 has limited processivity, however, *in vivo* the presence of binding proteins may alter this.

Although substrates containing monomethylated arginine are generally considered as reaction intermediates, PRMT7 has been shown to possess type III methyltransferase activity, with monomethylated substrates being the end product (Zurita-Lopez et al., 2012). Whether these monomethylated end products exist *in vivo* or are further methylated by another PRMT is not clear. In fact PRMT7 does function with PRMT5 in the methylation of Sm proteins, so it is possible that PRMT7 generates the monomethylated substrate which is then dimethylated by PRMT5. PRMT7, along with PRMT9 possesses two catalytic domains indicating a duplication event at some point during evolution, however what this contributes to enzyme behaviour is not

known.

PRMT8 is very similar in sequence to PRMT1 (80% identity), however it has the unique feature of being N-terminally myristoylated and is localised to the plasma membrane in neurons through this modification. Despite being similar to PRMT1, the activity of PRMT8 is much lower due to inhibition by the unique N-terminal domain. Removal of the N-terminal domain leads to a ten fold increase in arginine methyltransferase activity (Sayegh et al., 2007). The N-terminal domain is also the site of automethylation and is proline rich, enabling it to bind to SH3 domain containing proteins including PRMT2. The inhibitory effect of the N-terminal domain may mean that the *in vivo* function of PRMT8 depends on binding of other factors leading to conformational change or cleavage thereby releasing inhibition.

The final known member of the human PRMT family, PRMT9, was discovered at the same time as PRMT8 by searching for proteins containing the four conserved amino acid motifs (I, post-I, II and III). Although this enzyme appears to have all the features of an arginine methyltransferase it has not yet been characterised (Lee et al., 2005).

Two further members of the PRMT family are proposed to exist, the F box only proteins FBXO11 and FBXO10. FBXO11 was discovered by searching the human genome for proteins containing the SAM binding motif (motif I) (Cook et al., 2006). The primary structure of FBXO11 is very different to PRMTs1-9 and the secondary structure is also predicted to be different. This has led to the hypothesis that FBXO11 arose by convergent evolution. Weak arginine methyltransferase activity was shown for FBXO11, it was observed to catalyse the formation of MMA, ADMA and SDMA. This enzyme could

therefore be the first PRMT to generate both ADMA and SDMA. However in one experiment a FLAG-tagged construct was used and the anti-FLAG antibody is known to recognise PRMT5 meaning that the preparation was almost certainly contaminated with PRMT5 (Cook et al., 2006; Nishioka and Reinberg, 2003). FBXO10 was discovered through a search for homologues of FBXO11, however the methyltransferase capacity of this protein has not been assayed (Krause et al., 2007). Further arginine methyltransferases which are not related to the PRMTs may be discovered, this was the case for lysine methyltransferases for example where 8 subfamilies of enzyme are currently known (Zhang et al., 2012).

1.7 Reversibility of arginine methylation

Many post-translational modifications such as phosphorylation, acetylation and lysine methylation are reversible processes due to the existence of enzymes responsible for removing the modification. Currently however, no enzyme capable of demethylating arginine residues has been found. MMA residues can be deiminated to citrulline by PAD4 (a peptidylarginine deiminase), however no enzyme has been found capable of converting citrulline to arginine, therefore this cannot be considered as a true demethylation reaction (Li et al., 2010). Moreover, peptidylarginine deiminases are restricted in their expression and cannot act on ADMA residues, making this group of enzymes unlikely candidates for the role of arginine demethylases (Pahlich et al., 2005). Enzymes from the Jumonjii C (JmjC) domain containing group of proteins

are thought to be more likely to be capable of carrying out arginine demethylation. JmjC domain containing proteins have been shown to demethylate lysine residues in histones, and one member of this family, Jmjd6, was initially thought to be an arginine demethylase partly based on mass spectrometric analysis of methylated peptides incubated with the enzyme which showed that methyl groups were being removed from arginine (Chang et al., 2007). However the analysis demonstrated that oxidation on lysine residues was also taking place, an observation which was explained by later work which failed to confirm the demethylating activity of Jmjd6 and instead revealed that the main activity this enzyme possesses is the hydroxylation of lysine residues in RNA splicing factors (Webby et al., 2009). The confusion between a demethylase and a hydroxylase is understandable since hydroxylation is a mechanism for demethylation.

More recently, a screen for factors involved in control of transcription by light during seed germination in *Arabidopsis thaliana*, identified two JmjC domain containing proteins, JMJ20 and JMJ22 (Cho et al., 2012). Evidence was presented indicating that *in vitro* JMJ20 was able to remove methylation at histone 4 arginine 3. However, detection of demethylation relied on the use of an antibody whose binding could have been reduced by hydroxylation at a nearby lysine. Thus the question of whether an arginine demethylase exists is still open. Given the fact that arginine methylation is an important component of the histone code which is dynamic, it seems likely that arginine demethylases do exist and will be discovered in the near future.

1.8 Arginine methylation and transcription

Arginine methylation has been shown to play a role in numerous cellular processes such as transcription, RNA processing, signal transduction and localisation of proteins. It affects such processes by altering protein-protein or protein-nucleic acid interactions.

Histones were the first proteins found to harbour methylated arginine and after many decades of research this modification is now a moderately well characterised component of the histone code. The N-terminal tails of Histone 2A (H2A), Histone 3 (H3) and Histone 4 (H4) have been shown to contain methylarginine at several sites, H2A(R3, R11, R29), H3(R2, R8, R17, R26) and H4(R3). H2A was recently shown to be methylated at R11 and R29 by PRMT6, however only the latter site was observed *in vivo* (Waldmann et al., 2011). Some evidence was provided showing H2AR29me2a (asymmetric dimethylation at R29) was enriched at genes known to be repressed by PRMT6. A better characterised histone arginine methylation site is H3R2, which is asymmetrically methylated by PRMT6 (H3R2me2a) thereby preventing the MLL1 complex methylating H3K4, a marker for actively transcribed genes. As a result of this cross-talk between arginine methylation and lysine methylation, there is a negative correlation between H3K4me3 and H3R2me2a (Kirmizis et al., 2007; Guccione et al., 2007). Additionally, many proteins which bind the H3K4me3 mark are blocked if H3R2 is also methylated. Methylation of H3R2 is therefore linked to transcriptional repression. However the classification of histone methylarginines according to whether they are activating or repressing marks is not clear cut. PRMT6 is

also known to be a coactivator of nuclear receptors and a genome wide CHIP-seq study of histone methylated lysine and arginine marks did not show an enrichment of H3R2 in the promoters of either active or silent genes (Barski et al., 2007).

H3R2 is also subject to symmetric dimethylation (H3R2me2s) by PRMT5 and PRMT7, which blocks binding of RBBP7, a common component of co-repressor complexes and enhances binding of WDR5, a component of coactivator complexes which appears to be a reader of symmetrically methylated arginine in histone tails (Migliori et al., 2012). H3R2me2s is found in euchromatic regions of the genome and is associated with active transcription. H3R2me2a on the other hand reduces binding of WDR5 and is found in heterochromatic regions of the genome. Thus it seems that symmetric or asymmetric methylation of the same arginine can have opposite functional outputs. H3R8 and H4R3 are also sites of symmetric dimethylation by PRMT5, in this case however these marks are both linked to transcriptional repression (Pal et al., 2004; Wang et al., 2008).

Methylation of H3R17 by PRMT4 and H4R3 by PRMT1, PRMT6 and PRMT8 is generally associated with active promoters. Indeed PRMT4 and PRMT1 have both been identified as transcription coactivators, PRMT4 acts with transcription factors such as p53, NF-κB, PPARγ and c-Fos, while PRMT1 acts with YY1, p53 and RUNX1. The recruitment of both PRMT1 and PRMT4 to an estrogen receptor target promoter (pS2) has been analysed at high resolution by ChIP (Metivier et al. 2003). This demonstrated that recruitment was cyclical, each wave of PRMT1/4 recruitment lasting 40 minutes. Interestingly, although H3R17 methylation followed the same

period as PRMT1/4 and ER association, H4R3 methylation occurred at half the frequency, persisting for two ER association cycles. The significance of this is unclear, but this data clearly supports the existence of an arginine demethylase acting to remove methylation between cycles.

Arginine methylation plays a fundamental role in the regulation of transcription as it is an important part of the histone code. The exact mechanisms by which the deposit of methyl groups onto arginines in histones leads to transcriptional silencing/activation is largely unclear. Some progress has been made with the identification of a SDMA reader (WDR5) and an ADMA reader (TDRD3)(Yang et al., 2010), however the sequence of events downstream requires more study. Using histone peptides containing methylated arginine to isolate binders could lead to identification of novel readers, however it is likely that methylarginines within histone tails are not interpreted in isolation, but rather in the context of several other modifications. In fact there are many examples of cross-talk between arginine methylation and other modifications that make up the histone code (Wu et al., 2012; Feng et al., 2011).

A potentially more direct role for arginine methylation in the regulation of transcription was discovered by a study showing that the carboxy terminal domain (CTD) of RNA polymerase II (RNAPII) is methylated at R1810 by PRMT4 (Sims et al., 2011). The CTD of RNA polymerase II serves to recruit other proteins which are necessary for the coordination of transcription and RNA processing. Methylation of arginine within the CTD appears to take place before transcription initiation and this modification is recognised at least *in vitro* by TDRD3. However, while the *in vitro* evidence for

methylation of the CTD is strong, *in vivo* evidence was gathered using an antibody raised against a methylated CTD peptide whose reactivity seems to be affected by the phosphorylation status of a nearby serine (Ser5) thus confounding the results. In addition it is unclear what effect methylation of the CTD might have, since a non-conservative mutation (R1810A) was used to assess the functional significance of methylation at this site. Further work will be required to confirm these findings.

1.9 Arginine methylation in RNA processing

The majority of arginine methylated proteins identified so far are implicated in RNA processing, in fact the heterogenous nuclear ribonucleoproteins (hnRNPs) which are a family of RNA binding proteins and ensure correct transport and processing of pre-mRNAs, are major targets for PRMTs (PRMT1 in particular) and account for 65% of the ADMA in the cell nucleus. The proportion of these proteins existing in the methylated form is high, with 12% of the arginine residues in total hnRNPs existing as methylarginine (Boffa et al., 1977). Many hnRNPs contain glycine arginine rich (GAR) motifs known as RGG boxes which are methylated by PRMT1 (Liu and Dreyfuss, 1995). These RGG boxes are also RNA binding domains, making it likely that arginine methylation affects the affinity of hnRNPs for RNA. While this has not been shown for the hnRNPs, the methylation of other RNA binding proteins has been investigated and several studies have found that methylation of RNA binding proteins reduces affinity for RNA. For instance arginine

methylation of REF, an mRNA export adaptor, weakens its binding to RNA enabling the transfer of the mRNA to TAP and export from the nucleus. Methylation of the HIV-1 Rev protein by PRMT6 similarly reduced the interaction of Rev with viral RNA, but here this disrupts transport leading to reduced export of viral RNA from the nucleus. Work in *S. cerevisiae* showed that arginine methylation by Hmt1 was necessary for the dissociation of the mRNA export factor Npl3 from Tho2, a transcription elongation factor. In the absence of this methylation, the dissociation of the complex of RNA binding proteins formed during transcription is not regulated, leading to reduced export of mRNA (Yu et al., 2004).

Transcription and mRNA processing are tightly coupled and arginine methylation plays a role in the coordination of these two events. Several splicing factors are methylated by PRMT4, including CA150, a protein which links transcription to splicing. The methylation of CA150 in a proline glycine methionine (PGM) motif is necessary for the recruitment of SMN, a key factor in assembly of the spliceosome. Furthermore, experiments on both endogenous CD44 and a CD44 minigene showed that PRMT4 affected splicing outcome in an activity dependent manner (Cheng et al., 2007). The RNA binding proteins HuR and HuD, involved in stabilising mRNAs containing AU and U rich sequences are both methylated by PRMT4. HuR becomes methylated in response to LPS stimulation of macrophages, and methylation of HuD was shown to disrupt its association with RNA. It is thought that arginine methylation of HuR and HuD might prevent these proteins from protecting certain mRNAs coding for proteins involved in cell cycle arrest, thus causing cellular proliferation (Fujiwara et al., 2006).

Components of spliceosomes are known to be arginine methylated. Three of the seven Sm proteins which make up the snRNPs that are core components of spliceosomes are subject to symmetric dimethylation on arginine by PRMT5 and PRMT7. In human cells, the methylation of the C-terminal tails of Sm proteins is required for assembly of snRNPs (Gonsalvez et al., 2007). Methylation of the C-termini of Sm proteins leads to the interaction with the Tudor domain containing SMN. The SMN-Sm protein association is required for the efficient assembly of snRNPs. However this is not the case in *Drosophila melanogaster*, where it was shown that methylation is dispensable for snRNP biogenesis (Gonsalvez et al., 2008). In yeast, Sm proteins lack the methylated region completely (Roy et al., 1995), suggesting that during evolution the appearance of the RG rich C-termini and subsequent development as targets for modification by PRMTs has led to an increase in functional outputs dependent on this region. The critical role that arginine methylation plays in alternative splicing via the methylation of Sm proteins was highlighted by two studies that identified mutations in *PRMT5* causing a lengthening of the period in circadian rhythms of leaf movement and gene expression in *Arabidopsis thaliana* (Sanchez et al., 2010; Hong et al., 2010). In *Arabidopsis*, PRMT5 regulates the alternative splicing of a core clock gene, *PRR9*, loss of PRMT5 leads to increased inclusion of an intron and a reduction in production of the full length *PRR9* mRNA. This effect on alternative splicing appeared to extend to numerous other mRNAs, including that coding for Rubisco activase, whose circadian oscillations in alternative splicing were disrupted in PRMT5 mutant plants. In addition, PRMT5 expression is itself regulated by the circadian clock. This data cor-

relates with work showing that splicing could be inhibited by an antibody recognising methylated Sm proteins and that splicing efficiency is reduced in hypomethylated nuclear extracts (Boisvert et al., 2002). Thus, a likely molecular explanation for the effects seen in PRMT5 mutant plants is that the reduced methylation of spliceosomal Sm proteins reduces recognition of some 5' splice sites, although no direct evidence for this hypothesis has been provided in the *Arabidopsis* studies and the effects observed could be due to altered histone methylation leading to altered expression of other proteins involved in splicing.

1.10 The role of arginine methylation in determining the subcellular localisation of proteins

Arginine methylation is known to affect the subcellular localisation of many RNA binding proteins by affecting nuclear-cytoplasmic shuttling. This effect was first shown in yeast cells lacking the arginine methyltransferase Hmt1, where a nuclear accumulation of the hnRNPs Npl3 and Hrp1, both methylated by Hmt1, was observed (Shen et al., 1998). Conversely, overexpression of Hmt1 led to increased nuclear export of Npl3. There are also some examples of the opposite scenario, where blocking arginine methylation leads to cytoplasmic rather than nuclear accumulation of proteins. For instance inhibiting methylation or deficiency of PRMT1 leads to cytoplasmic accu-

mulation of the RNA binding protein Sam68 (Côté et al., 2003). Similarly, nuclear accumulation of the higher molecular weight form of FGF-2 was inhibited by blocking methylation (Pintucci et al., 1996), in this case a general methyltransferase inhibitor was used so the effect cannot be attributed directly to methylation on arginine, however FGF-2 was later shown to be a substrate for PRMT5 (Bruns et al., 2009).

It is worth noting that proteins whose localisation is altered by inhibition of methyltransferase activity are themselves arginine methylated and that nuclear-cytoplasmic shuttling of the total cellular protein pool is not affected by blocking arginine methylation. This makes it likely that the effect on subcellular localisation is due to the methylarginine within these proteins affecting binding to partners rather than a general pathway for nuclear-cytoplasmic transport being susceptible to arginine methylation. Supporting this notion, evidence for the direct impact of methylarginine on protein localisation was provided by experiments where RNA Helicase A protein, pre-methylated by PRMT1 or not, was cytoplasmically microinjected into HeLa cells (Smith et al., 2004). Nuclear import of RNA Helicase A which was not pre-methylated could be inhibited by a methylase inhibitor. However, methylating RNA Helicase A with PRMT1 prior to injection meant that nuclear import was not blocked by the methylase inhibitors. Methylation of Piwi family proteins (MILI, MIWI, MIWI2) by PRMT5 leads to recognition and binding by Tudor domain containing proteins (TDRD1, TDRD2, TDRD4, TDRD6, TDRD7 and TDRD9). Association of Tudor domain containing proteins is essential for the correct localisation and therefore function of Piwi family proteins. Specifically it was demonstrated that MIWI2 and TDRD1

were co-localised to the cytoplasmic nuage, and localisation of MIWI2 was disrupted in prospermatogonia from TDRD1 mutant mice, this was accompanied by a loss of secondary piRNAs antisense to transposons and evidence of reduced silencing of transposons (Vagin et al., 2009).

Another mechanism by which methylation on arginine could alter localisation of proteins is by affecting the interaction of a substrate protein with its nuclear import receptor. Poly(A) binding protein, PABPN1, which regulates poly(A) tail length on mRNA, contains numerous arginine residues targeted for asymmetric dimethylation by PRMT1, PRMT3 and PRMT6. In fact a total of 13 arginine residues within the C-terminal domain can be methylated. *In vitro* at least, arginine methylation does not affect binding of PABPN1 to poly(A) RNA (Kühn et al., 2003), but it does lower the affinity of PABPN1 for transportin, an import receptor involved in shuttling RNA binding proteins into the nucleus (Fronz et al., 2011). Therefore PABPN1 could be transported into the nucleus, once there it may become methylated preventing binding and exit via transportin. However no direct evidence for this mechanism through the use of methylase inhibitors or PRMT knockdown has been provided.

1.11 Arginine methylation in signal transduction

Post-translational modifications are used to rapidly transmit a signal from a receptor into the cell where the information can be transformed into a functional output. Phosphorylation is often used in this role as it imparts a large physico-chemical change on the modified protein (negative charge) and is readily reversible, enabling on/off switching. Methylation on arginine might seem an unusual choice for this role as it is a relatively stable PTM and is perhaps a more subtle modification as it does not alter the charge of the residue. Nonetheless, there is some evidence that arginine methylation may play a role in signal transduction. PRMT1 was found to interact with the cytoplasmic tail of a type I interferon receptor (IFNAR1) in a yeast two hybrid screen, and knocking down PRMT1 made cells more resistant to growth inhibition by IFN, although no additional *in vivo* data was shown demonstrating PRMT1 interaction with IFNAR1 in mammalian cells (Abramovich et al., 1997). In neuronal PC12 cells, NGF stimulation leads to an increase in PRMT1 activity and accompanying increases in the levels of arginine methylated proteins after 12-24 hours of treatment (Cimato et al., 2002). On a much faster time scale, stimulation of T-lymphocytes by antigen presenting cells (with or without superantigen), led to increases in PRMT activity and elevated arginine methylation of numerous proteins within 10-30 minutes of activation (Blanchet et al., 2005). In B-lymphocytes, the B-cell antigen receptor is reported to be monomethylated by PRMT1, this was suggested to

negatively regulate downstream signalling (Infantino et al., 2010). However, binding of the antibody against the monomethylation site may be influenced by the nearby phosphorylation of the tyrosine in the ITAM motif, potentially complicating the western blot results.

More detailed studies into the role of arginine methylation in signal transduction have shown that arginine methylation by PRMT5 can modulate the output from Mitogen activated protein kinase (MAPK) signalling cascades. MAPK signalling consists of a relay of protein phosphorylation beginning at a growth factor receptor and resulting in phosphorylation and activation of transcription factors by a MAPK. EGFR is one such receptor which was shown to be methylated on arginine by PRMT5 (Hsu et al., 2011). Mutagenesis of the methylation site and knockdown of PRMT5 indicate that methylation positively modulates phosphorylation at a nearby tyrosine, leading to recruitment of a phosphatase and consequent attenuation of signalling. Proteins downstream of EGFR in the MAPK signalling pathway are also subject to arginine methylation. Stimulation of PC12 cells by EGF was shown to lead to PRMT5 dependent increases in CRAF and BRAF methylation (Andreu-Pérez et al., 2011). Methylation of these RAF family kinases leads to their degradation and therefore decreased activation and output from the MAPK ERK1/2. The decreased phosphorylation output from ERK1/2 leads to proliferation of PC12 cells, whilst inhibiting methylation leads to increased signalling amplitude and duration from ERK1/2, resulting in cellular differentiation. Thus these two studies indicate a down-modulation of MAPK signalling by arginine methylation.

Cross-talk between arginine methylation and phosphorylation is also seen

in the PI3K-Akt signalling pathway. The Forkhead box O (FOXO1) transcription factor is methylated at R248 and R250 by PRMT1 in response to oxidative stress. These methylated arginines form part of the motif phosphorylated by Akt. As a consequence, Akt phosphorylation of S253 is blocked and FOXO1 accumulates in the nucleus, increasing oxidative stress induced apoptosis. Another Akt-phosphorylated protein, BCL-2 antagonist of cell death (BAD), is also subject to methylation by PRMT1 within the Akt phosphorylation motif. This similarly blocks phosphorylation by Akt leading to pro-apoptotic signalling. In *C. elegans* the Forkhead transcription factor DAF-16 is subject to methylation by PRMT1, blocking phosphorylation by Akt. This leads to expression of longevity related genes and knocking out PRMT1 function in *C. elegans* reduces lifespan. Another example of the involvement of arginine methylation in apoptotic signalling pathways is the methylation of apoptosis signal regulating kinase 1 (ASK1). ASK1 is a MAP kinase kinase kinase whose activation by cellular stresses eventually leads to activation of caspases and apoptosis. In this apoptotic signalling pathway, arginine methylation appears to act as a negative regulator. ASK1 is methylated at R78 and R80 by PRMT1, enhancing interaction with thioredoxin, thereby preventing ASK1 from activating downstream apoptotic signalling pathways.

In estrogen receptor α(ERα) signalling, arginine methylation is required for the formation of a cytoplasmic ERα signalling complex. Methylation of ERα at R260 by PRMT1 takes place within 5 minutes of initiating estrogen signalling. Inhibiting this methylation through knockdown of PRMT1 prevents ERα binding to the p85 subunit of PI3K and the tyrosine kinase Src.

Thus arginine methylation of ERα is required for formation of a cytoplasmic signalling complex leading to activation of downstream Akt and MAPK signalling pathways. Interestingly, inhibiting Src kinase activity with PP1 prevented the methylation of ERα, revealing in this case a dependence of arginine methylation on phosphorylation. A further interesting observation was the down regulation of ERα methylation after 15 minutes of signalling, pointing towards a mechanism for demethylation, although there was some evidence that this could be due to proteasomal degradation (Le Romancer et al., 2008).

1.12 Arginine methylation in DNA repair

DNA repair is an essential process for the maintenance of genome integrity. Lesions induced in DNA, such as double strand breaks, represent a danger to the cell as they can lead to chromosomal rearrangement. The Mre11 complex, consisting of Mre11, Rad50 and NBS1 is a central component of the DNA damage response and can sense double strand breaks (Stracker and Petrini, 2011). Mre11 is recognised by an ADMA specific antibody, Asym25 and can be methylated *in vitro* by PRMT1 (Boisvert et al., 2005). MALDI-TOF analysis of the mass of intact tryptic peptides from Mre11 immunoprecipitated with Asym25 suggested the presence of 9 dimethylated arginine residues. Arginine methylation is thought to be important for the 3'-5' exonuclease activity of Mre11 as mutating all arginines within the glycine arginine rich domain to lysine diminished exonuclease activity. Knocking down PRMT1

was shown to cause cells to enter into S-phase in spite of induced DNA damage and introduction of methylated Mre11 complex restored this checkpoint. However the experiment with non-methylated Mre11 was not performed. Further work showed that treating cells with methylase inhibitors prevented localisation of Mre11 to sites of DNA damage (Boisvert et al., 2005).

Single strand breaks or damaged bases are repaired by the Base excision repair (BER) pathway. There is evidence that arginine methylation may also be involved in this DNA repair pathway. DNA polymerase β is able to remove damaged bases and fill in short single stranded DNA gaps. It was demonstrated that DNA pol β is methylated at R83 and R152 by PRMT6 and this methylation is augmented upon treatment of cells with the DNA alkylating agent MMS. Mutation of both methylation sites to lysine impaired the ability of DNA pol β to carry out BER. Finally *in vitro* assays showed that methylation of DNA pol β increases its affinity for DNA and processivity. DNA pol β was also shown to be methylated by PRMT1 at R137, this had no effect on the BER activities of DNA pol β, but it blocked interaction with PCNA (El-Andaloussi et al., 2007), although the consequences of this are unclear. By contrast, arginine methylation of FEN1, an endo and exonuclease involved in DNA repair, enhances its interaction with PCNA. Methylation of FEN1 by PRMT5 at R192 in response to oxidative stress leads to blockage of phosphorylation which facilitates interaction with PCNA. This in turn ensures the localisation of FEN1 to the nucleus where DNA repair can be carried out.

Other proteins involved in DNA repair have been shown to be regulated by PRMT5. p53 is an important DNA damage responsive transcription fac-

tor and is methylated by PRMT5 in a region necessary for oligomerization and nuclear localisation. Arginine methylation was required for correct localisation and influenced the transcriptional outcome of the p53 response. A methylation defective mutant of p53 was unable to cause G1 arrest but was still able to induce apoptosis (Jansson et al., 2008). Rad9 is also thought to be methylated by PRMT5, although *in vivo* evidence of methylation was based on the use of an antibody. Mutation of the proposed methylation sites led to increased sensitivity of cells to DNA damage by hydroxyurea (He et al., 2011). However the continued use of the anti-FLAG antibody in many studies on PRMT5 is a cause for concern as this antibody has been shown to specifically recognise PRMT5.

1.13 Arginine methylation in T-lymphocytes

The first indication that methylation may be particularly important for T-lymphocyte function was shown in studies on the effects of adenosine deaminase (ADA) deficiency (Kredich and Martin 1977). A deficiency in ADA activity causes severe combined immune deficiency (SCID). ADA is important for the breakdown of purines, in the absence of this enzyme deoxyadenosine accumulates, in turn generating excess dATP in the cell. Excess dATP inhibits ribonucleotide reductase, an enzyme important for DNA replication and repair. This appears to be particularly toxic to T and B lymphocytes. The reason for the particular toxicity seen in lymphocytes was suggested to be due to the elevated levels of adenosine and deoxyadenosine inhibiting the

activity of S-adenosylhomocysteine hydrolase (SAHase) (Figure 1.6). As a result, levels of SAH become elevated which is a potent feedback inhibitor of SAM dependent methyltransferases (Gaspar et al., 2009). A study of SAM metabolism in primary human peripheral blood mononuclear cells showed that in resting cells SAM turnover is 3-5 fold higher compared to non-hepatic tissues. Stimulation of these cells with phytohemagglutinin (PHA) which induces proliferation, led to a 10-16 fold increase in SAM utilization. Increases in SAM metabolism were noticeable after 2 hrs of stimulation and peaked at 30-40 hrs of stimulation (German et al., 1983). Further evidence for the

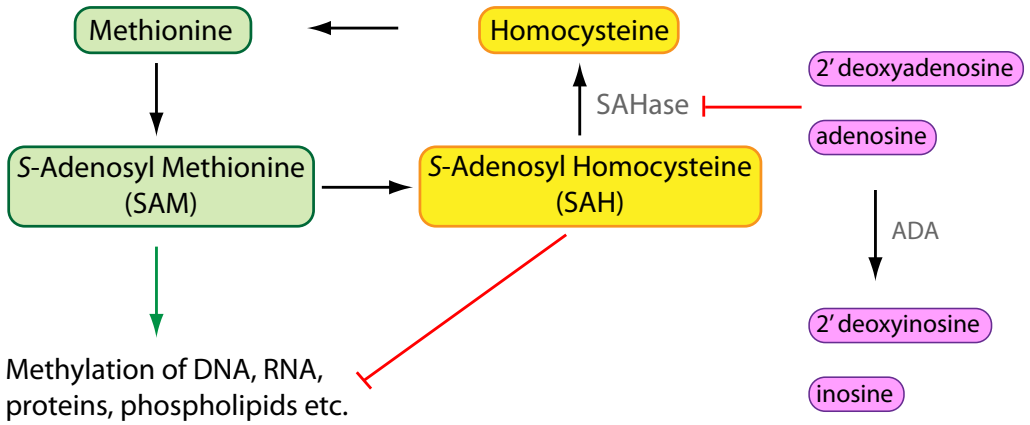


Figure 1.6: The link between adenosine deaminase (ADA) deficiency and methylation. Severe combined immune deficiency (SCID) is a lethal impairment of T and B lymphocytes. One cause of SCID is a defect in adenosine deaminase leading to accumulation of adenosine and deoxyadenosine both of which inhibit *S*-adenosyl homocysteine hydrolase (SAHase). This leads to accumulation of *S*-adenosyl homocysteine, a potent feedback inhibitor of SAM-dependent methyltransferases.

importance of methylation in T-lymphocytes was gathered through the use of an irreversible inhibitor of S-adenosylhomocysteine hydrolase which led to inhibition of T-lymphocyte but not B-lymphocyte proliferation after stimulation. Secretion of IL-2 by T-lymphocytes was also inhibited. *In vivo*,

inhibiting SAHase impaired the T-lymphocyte but not the B-lymphocyte response (Wolos et al., 1993). A later study made use of a less toxic reversible inhibitor of SAHase. Pre-treatment of mouse CD4 T-lymphocytes with this compound impaired signalling from the TCR. By contrast, the same experiment on CD8 T-lymphocytes led to an enhancement of signalling (Lawson et al., 2007). This data suggests that CD4 and CD8 T-lymphocytes differ in their responses to inhibition of methylation.

The finding that inhibiting SAHase impairs T-lymphocyte function is only indirect evidence for the importance of arginine methylation in these cells since lysine, DNA, RNA and other methylation reactions will also be affected. More direct evidence that arginine methylation in particular is important to T-lymphocyte function was shown by Richard et al. who knocked down PRMT5 expression in the Jurkat T-lymphocyte cell line (Richard et al., 2005). This led to a reduction in Nuclear Factor of Activated T-cells (NFAT) and IL-2 promoter driven gene expression, as a consequence IL-2 secretion was impaired. A ChIP assay provided weak evidence that stimulation led to association of symmetrically dimethylated proteins to the IL-2 promoter, thus suggesting a possible mechanism linking PRMT5 to IL-2 production. NFAT refers to any one of a family of four transcription factors activated by calcium after triggering of the TCR. This family of transcription factors links TCR signalling to the transcription of many genes important for T-lymphocyte activation, such as those coding for cytokines. NFAT interacting protein 45 (NIP45) binds to NFAT and enhances its ability to activate transcription. NIP45 is methylated at its N-terminus by PRMT1, increasing the affinity for NFAT. Although the exact site of methylation was not discovered, the

N-terminus of NIP45 was shown to be important for the production of IL-4 after stimulation and PRMT1 was shown to positively regulate the IFN-γ and IL-4 promoters (Mowen et al., 2004). This finding was corroborated through the use of a PRMT inhibitor which was shown to inhibit the production of IFN-γ and IL-4 from stimulated T_H1 and T_H2 cells (Bonham et al., 2010).

These observations, together with the findings that arginine methylation is induced upon T-lymphocyte activation and T-lymphocyte development is blocked in PRMT4 null mice all point towards an important role for arginine methylation in T-lymphocyte function.

1.14 Current state of knowledge of arginine methylation

Mass spectrometry based proteomics is the only technique suitable for high throughput analysis of post translational modifications. While several in depth investigations of other post-translational modifications such as phosphorylation and acetylation have been published, no large scale study of arginine methylation has ever been carried out. Previous attempts to discover arginine methylation sites by mass spectrometry either failed to specifically identify any sites, or led to a small number of sites being identified. This was due to a lack of suitable reagents (antibodies) and techniques for enriching arginine methylation sites prior to mass spectrometry. Consequently, only a relatively small number of arginine methylation sites are known. A search of

the UniProt database returns 137 arginine methylation sites in 59 proteins, this compares to over 9000 phosphorylation sites listed in over 7000 proteins. Even less is known about arginine methylation in T-lymphocyte activation, in fact only three proteins have been shown to be arginine methylated in T-lymphocytes, TARPP, NIP45 and Vav1 (Mowen et al., 2004; Kisielow et al., 2001; Blanchet et al., 2005). Moreover, all previous proteomic studies of arginine methylation have been carried out in transformed cell lines where arginine methylation may be dysregulated. There is therefore scope to significantly extend the understanding of this post-translational modification.

1.15 Research objectives

The main aim of this thesis was to gain an understanding of the role of arginine methylation in T-lymphocyte activation. This required the completion of three main research objectives:

1. Discover when arginine methylation takes place during the course of T-lymphocyte activation (Chapter 3).

2. Develop methods for the enrichment and identification of arginine methylation sites by mass spectrometry (Chapter 4).

3. Follow up novel arginine methylated proteins known to be important for T-lymphocyte activation and analyse the functional effect of arginine methylation (Chapter 5).

By developing methods and carrying out a large scale characterisation of arginine methylation, it is also possible to increase understanding of the role of arginine methylation in general cellular processes and gain insight into exactly how methylation on arginine affects protein-protein interactions.

2. Materials and Methods

2.1 Culture of cell lines

Jurkat Clone 20 cells were maintained at 37°C in a 5% CO_2 atmosphere in RPMI 1640 medium (PAA), supplemented with 10% triple 0.1µm sterile filtered fetal calf serum (FCS, HyClone, Thermo Scientific).

2.2 Isolation of primary human PBMCs or CD4+ T-lymphocytes

Materials

Buffer B1:

100ml Dulbecco's PBS (GIBCO)

100µl FCS

400µl 1M EDTA pH 8.0 (GIBCO)

Dynal CD4 negative isolation kit (Invitrogen), containing antibody mix (raised in mouse) for depletion of all cells other than CD4+ T-lymphocytes and mag-

netic beads coated with goat anti-mouse IgG.

50ml pouches of defibrinated whole blood were obtained from NHS Blood and Transplant.

All steps and solutions were at room temperature unless stated otherwise.

1. The 50ml defibrinated whole blood was diluted 1:1 with Dulbecco's PBS
2. Each 25ml of this diluted whole blood was carefully overlaid onto 15ml of Ficoll (GE Healthcare) in a 50ml centrifuge tube using a 25ml pipette
3. Blood/Ficoll layers were centrifuged at 400g for 40 minutes, ensuring centrifuge brake was off
4. Most of the uppermost yellow layer (blood plasma containing platelets) was carefully removed and discarded using a 25ml pipette
5. Using a 1ml pipette, the whitish material (peripheral blood mononuclear cells - PBMCs) that had collected at the interface with the Ficoll was carefully removed and collected in a 50ml centrifuge tube
6. Collected PBMCs were diluted 3 times in PBS and centrifuged at 100g for 10 minutes
7. Supernatant was removed and pellet was carefully resuspended in 20ml buffer B1. This was centrifuged at 100g for 10 minutes

8. Supernatant was removed and pellet was carefully resuspended in 20ml buffer B1 and centrifuged at 400g for 10 minutes. This step was repeated, during which a 50µl aliquot of the resuspended pellet was taken to count the cells. For isolation of PBMCs, the cell pellet after the final centrifugation was resuspended at a concentration of 1-2$\times 10^6$ cells/ml in RPMI 1640 medium supplemented with 10% FCS

9. For negative isolation of primary human CD4+ T-lymphocytes, the cell pellet was resuspended carefully in buffer B1 + 20% FCS and placed in a 15ml centrifuge tube. The antibody mix was added and the tube rotated at 4°C for 20 minutes. For every 200$\times 10^6$ PBMCs, the cell pellet was resuspended in 1.6ml B1 + 400µl FCS + 120µl antibody mix

10. For 200$\times 10^6$ PBMCs, 8ml B1 was added. The tube was mixed well and centrifuged for 8 minutes at 300g

11. For 200$\times 10^6$ PBMCs, 250µl Dynal beads were washed twice in an equal volume of B1

12. The cell pellet was resuspended in B1 (for 200$\times 10^6$ PBMCs 2.8ml B1), pre-washed Dynal beads were added and cells were rotated for 15 minutes

13. Bead-bound cells were resuspended by gently pipetting 5 times using a 1ml pipette

14. Cell suspensions were transferred to 1.5ml microcentrifuge tubes, placed on a magnet and left for 2 minutes. Supernatant was removed, taking

care to leave a small amount in the bottom of the tube to prevent removal of some of the beads. Beads were washed in 1ml B1, placed on the magnet and supernatant removed as before

15. Supernatant from the beads (containing the purified hCD4+ cells) was centrifuged for 8 minutes at 400g. The pellet was resuspended in RPMI 1640 supplemented with 10% FCS. Cells were rested overnight at a concentration of 1-2$\times 10^6$/ml. Purity of isolated CD4+ T-lymphocytes was assessed by flow cytometry and ranged from 85-95% pure.

2.3 Flow cytometry analysis

Materials

Staining buffer:

PBS with 2% FCS and 0.05% NaN_3 (ice cold)

0.5-1$\times 10^6$ cells were used for each antibody staining.

1. Cells were centrifuged for 2 minutes at 400g at RT to pellet. Supernatant was removed. Pellet was resuspended gently in 900μl staining buffer

2. Cells were pelleted again for 2 minutes at 400g at RT. Supernatant was removed and pellet was resuspended gently in staining buffer (105μl per antibody stain carried out). For each antibody stain, 100μl of the resuspended cells was placed in a 1.5ml microcentrifuge tube

3. 1µl antibody was added to each tube. For CD3 and CD4 staining, PE conjugated antibodies (BD Pharmingen) were used at a 1:100 dilution. PE conjugated IgG was included as a control (BD Pharmingen). Cells were mixed gently by pipetting and left covered on ice for 30 minutes

4. 1ml of staining buffer was added to each tube and cells were centrifuged for 2 minutes at 400g

5. Supernatant was removed and cells were gently resuspended in 1ml staining buffer. Cells were pelleted again for 2 minutes at 400g

6. Supernatant was removed and cells were resuspended in 500µl staining buffer and transferred to flow cytometry tubes for analysis

2.4 Stimulation of cells with bead bound antibodies

1. 2×400µl of protein G coated magnetic beads (4µm diameter, Kisker biotech) were washed twice in 800µl PBS

2. 4µg of anti-CD3 antibody (OKT3) and 4µg of anti-CD28 antibody (CD28.2) were added to 400µl of beads, 8µg control IgG (Santa Cruz Biotechnology) was added to the control beads. Antibodies were left to bind for 1 hour at room temperature

3. Beads were washed twice in 800µl PBS, control and stimulating beads were each resuspended in 400µl PBS

4. In the longer term stimulations, for each stimulation time point, 10-20×10^6 cells were resuspended in 10ml RPMI 1640. 100µl of control or stimulating beads was added (bead:cell, 4:1). Penicillin and streptomycin was also added at 100 units/ml and 100µg/ml respectively. For shorter time courses (30 minutes), 5×10^6 cells were stimulated in 100µl RPMI and 25µl were taken out at each time point. Cells were harvested by centrifugation at the appropriate time point and frozen on dry ice

2.5 Stimulation of cells with superantigen loaded Raji APCs

1. Ratio of Raji APC to stimulating cells was 1:1. 10×10^6 Raji cells were centrifuged and resuspended in 100µl RPMI.

2. Staphylococcal Enterotoxin A, B and C were added to Raji cells so the concentration of each was 2µg/ml. Cells were left at 37°C for 2 hours for superantigen to bind

3. Cells were stimulated as above for short time courses

2.6 Western blotting

Materials

RIPA buffer:

50mM Tris pH 7.5

125mM NaCl

1mM EDTA

1% NP-40

0.5% Deoxycholate

0.1% SDS

SDS loading buffer:

200mM Tris pH 6.8

8% SDS

200mM DTT

40% glycerol

0.2% Bromophenol blue

5% β-mercaptoethanol

Transfer buffer:

190mM glycine

50mM Tris

0.05% SDS

20% methanol

Running buffer:

25mM Tris

192mM glycine

0.1% SDS

Primary antibody solution:

3% BSA

0.2% Tween

0.1% NaN_3

Secondary antibody solution:

3% BSA

0.2% Tween

0.01% SDS

0.1% NaN_3

1. Every 10×10^6 cells was lysed in 45μl RIPA buffer (containing protease inhibitors (Complete, Roche) and Na_3VO_4 at 1mM). Cells were lysed by vortexing with RIPA and leaving on ice for 10 minutes.

2. Lysate was centrifuged for 20 minutes at 20,000g at 4°C

3. SDS loading buffer was diluted 4× in the lysate supernatant. This was then boiled for 5 minutes

4. Lysates were loaded on a 10% polyacrylamide gel, together with molecular weight markers (Precision Plus dual color, Bio-Rad). Lysate from 0.5-1.0$\times10^6$ cells was loaded per well.

5. Gel was run at 140V in running buffer until the marker had just exited the gel

Antibody	Dilution	Source
4G10	1:5000	Millipore
actin	1:5000	Sigma
DMA (7E6)	1:500	Abcam
Asym24	1:500	Millipore
GAPDH	1:5000	Santa Cruz Biotechnology
Sym10	1:500	Millipore
PRMT1	1:1000	Upstate
PRMT1 (MAT-B12)	1:1000	Santa Cruz Biotechnology
PRMT5	1:1000	Upstate
PRMT4	1:1000	Upstate
Zap-70	1:5000	Cell Signalling Technologies
Ubiquitin	1:1000	Santa Cruz Biotechnology
HuR	1:1000	Santa Cruz Biotechnology
GFP	1:1000	Santa Cruz Biotechnology
WASp	1:1000	Santa Cruz Biotechnology
Fyn	1:1000	Santa Cruz Biotechnology
WIPF1	1:1000	Santa Cruz Biotechnology
Grb2	1:1000	Santa Cruz Biotechnology
Lck	1:1000	Santa Cruz Biotechnology
WASp pY291	1:500	Abcam

Table 2.1: Antibodies used for western blotting.

6. Gel was removed and placed on a nitrocellulose membrane (LI-COR) in between two pieces of filter paper pre-soaked in transfer buffer. Proteins were transferred at 18V for 1 hour

7. Membrane was blocked for 15 minutes using blocking buffer (LI-COR)

8. Membrane was incubated with primary antibody solution for 1.5 hours (or overnight at 4°C)

9. Membrane was washed 3× with PBS 0.1% Tween 20

10. Membrane was incubated with secondary antibody solution for 30 minutes

11. Membrane was washed 3× with PBS 0.1% Tween 20

12. Infra-red dye conjugated secondary antibodies were detected with an Odyssey imaging system (LI-COR)

2.7 SILAC labelling of cells

Materials

Dialysed FCS (GIBCO)

SILAC RPMI 1640 without Arginine, Lysine or Methionine (GIBCO or Dundee cell products)

Stock amino acid solutions, all from Cambridge Isotope Laboratories except

labelled L-Methionine from Sigma Isotec

0.4M L-Arg	84.3mg/ml
0.8M L-Lys	146mg/ml
0.4M Arg10(13C6, 15N4)	88.3mg/ml
0.2M L-Methionine	29.8mg/ml
0.2M L-Methionine(13CD3)	30.65mg/ml
0.2M L-Methionine(13C4)	30.65mg/ml

1. To 500ml of SILAC RPMI 1640, 357µl of the required 0.4M stock arginine (final concentration 0.1mM), 137µl of the required 0.8M stock lysine (final concentration 0.219mM) and 250µl of the appropriate methionine (final concentration 0.1mM) were added

2. Media was sterile filtered using a 0.22µm filter. 50ml dialysed fetal calf serum was added

Jurkat Clone 20 cells

1. 5 million Jurkat Clone 20 cells were centrifuged for each SILAC condition required, at 400g for 5 minutes

2. Supernatant was removed, cell pellet was resuspended in 20ml of appropriate pre-warmed SILAC medium

3. Cells were maintained at a concentration of 2.5-5.0$\times 10^5$/ml in 20ml for 3 cell doublings, after which SILAC medium was added until at least 6 cell doublings had occurred. At this point cells could be harvested for experiments

Primary human T-lymphocytes

1. On the day prior to isolation of T-lymphocytes from whole blood, a small cell culture flask was coated with anti-CD3 antibody (OKT3, Adrian Gray), by using 5ml of a 10µg/ml OKT3 solution in PBS. Flasks were left overnight at 4°C for antibody to adsorb to surface

2. Primary T-lymphocytes were isolated as described previously, except 20 million lymphocytes were resuspended in 10ml of the SILAC labelling medium

3. The anti-CD3 solution was removed from the flasks and primary lymphocytes were added together with CD28.2 antibody (BioLegend) at 1µg/ml, 50ng/ml Interleukin-2 (IL-2, AbD serotec) and penicillin + streptomycin (PAA)

4. Primary lymphocytes were cultured under the same conditions as Jurkat Clone 20 cells, with additional SILAC medium (together with CD28.2 antibody and IL-2) being added to maintain the cells at a concentration of 0.5-1.0$\times 10^6$/ml

5. After 7-9 days, cells were centrifuged at 400g for 5 minutes and either:

- To achieve near complete incorporation of SILAC label, 100×10^6 lymphocytes were resuspended in 100ml of the SILAC medium in a fresh flask not containing anti-CD3 antibody, CD28.2 antibody or IL-2. After resting for 2-3 days, cells were transferred to a fresh flask coated with anti-CD3 antibody with addition of CD28.2, IL-2 and penicillin + streptomycin and grown for a further 7-9 days
- Or for harvesting of cells, batches of 25 million were washed twice in 1ml PBS and frozen on dry ice

2.8 Immunoprecipitation of arginine methylated proteins

1. 40µl of protein G sepharose was washed 3× in 200µl PBS
2. Beads were resuspended in 400µl PBS and split into two tubes. Mouse IgG (40µg) was added to one and 40µl 7E6 antibody (Abcam) to the other. Antibodies were left to bind for 1 hour at room temperature
3. Beads were washed twice in 200µl 0.2M sodium borate
4. Antibodies were cross-linked for 45 minutes at room temperature using 20mM dimethyl pimelimidate
5. Beads were washed once in 200µl triethanolamine, resuspended in triethanolamine and blocked for 2 hours at room temperature
6. Beads were washed once in PBS. 100×10^6 light and heavy labelled cells were each lysed in 200µl modified RIPA buffer (see below). Light and

heavy lysates were mixed in a 1:1 protein ratio. Lysate was divided and placed with IgG and 7E6 beads

7. Immunoprecipitation was performed overnight at 4°C
8. Beads were washed 3× in modified RIPA
9. Proteins were eluted in 40µl triethylamine for 30 minutes. 2µl 1M Tris pH 6.8 was then added to neutralise pH prior to loading on a gel. Beads were boiled in NuPage sample buffer (Invitrogen)

2.9 Extraction of protein for protease digestion

Materials

Urea lysis buffer:

8M urea in 20mM Tris pH 8

1. Each pellet of 25 million cells was lysed by addition of 75µl of urea lysis buffer and subsequently shaking the cells at 850 rpm at room temperature for 15 minutes
2. Lysate was centrifuged at 20,000g for 30 minutes at room temperature
3. Supernatant was removed and placed in a 1.5ml microcentrifuge tube. 600µl of ice cold acetone (pre-chilled at -20°C) was added and proteins left to precipitate at -20°C for 1 hour

4. Precipitated proteins were pelleted at 3000g for 10 minutes at 4°C

5. Supernatant was discarded and pellet was washed with 500µl ice cold acetone. Proteins were centrifuged again at 3000g for 10 minutes at 4°C

6. Supernatant was discarded and pellet was allowed to air dry in a laminar flow hood until excess acetone had evaporated

7. Protein pellet was resuspended by addition of 75µl urea lysis buffer and shaking the tubes at 850rpm at room temperature until all protein was solubilised

8. Concentration of protein was measured using the Bradford dye binding assay

9. Protein extracts were mixed in the appropriate protein ratios. Proteins were reduced with 5mM dithiothreitol for 30 minutes at room temperature

10. Proteins were alkylated with 15mM iodoacetamide for 30 minutes at room temperature. From this point onwards, tubes were kept covered with foil

11. Urea in the protein extracts was diluted to 1.5M with 20mM Tris pH8

12. $CaCl_2$ was added to a concentration of 1mM

13. Trypsin or chymotrypsin was added at a w/w ratio of protease:protein of 1:50-100

14. Proteins were digested overnight at 37°C with shaking at 850rpm

2.10 Strong cation exchange (SCX) chromatography of peptides

Materials

Buffer A:

50mM KH_2PO_4

20% acetonitrile

pH 2.7

Buffer B:

50mM KH_2PO_4

20% acetonitrile

1M KCl

pH 2.7

SCX chromatography was carried out using an ÄKTApurifier (GE Healthcare) at a flow rate of 1ml/min

1. Tryptic peptides (5-10mg) from protein extraction and protease digestion were acidified with 0.1% Trifluoroacetic acid (TFA) and clarified by centrifuging for 10 minutes at 20,000g

Column volumes	% Buffer B
2	0
2.1	7
12	7
20	35
24	100
29	100
29.1	0
36	0

Table 2.2: Strong cation exchange chromatography profile (established by Thomas Uhlmann) used to enrich arginine methylated peptides.

2. Peptides were applied to a 1ml SCX column (Resource S, GE Healthcare) in 5ml Buffer A (See Table 2.2 for chromatography conditions). Bound peptides were eluted with an increasing gradient of Buffer B

3. Peptides with charge $<+3$ were eluted by increasing the concentration of Buffer B to 7%. Peptides with charge $\geq+3$ were then eluted by increasing the concentration of Buffer B to 35%

4. In the final section of the chromatography, Buffer B was changed to pH 7 and the proportion of Buffer B was increased to 100% to ensure elution of all bound peptides

5. Fractions were dried down and resuspended in 0.1% TFA for desalting according to the protocol below

2.11 OFFgel isoelectric focusing of peptides

1. 1mg of tryptic peptides was diluted so that the final concentration of Tris was <10mM

2. Peptides were loaded onto an IPG strip (pH 7-11, 13cm GE Healthcare) together with ampholytes (pH 7-11, GE Healthcare) according to the manufacturer's instructions

3. Peptides were focused in solution using an Agilent 3100 OFFgel fractionator (Agilent Technologies) for 20kVh up to a maximum voltage of 8000V

4. After focusing, the solution containing the peptides was removed using a pipette, taking care not to touch the gel on the IPG strip. Peptides were also extracted by immersing the cathode paper pad in 0.1% TFA for 1-2 hours followed by brief sonication. Peptides were then desalted according to the protocol below

2.12 Desalting of peptides from strong cation exchange and OFFgel isoelectric focusing

Solutions required:

100% acetonitrile

50% acetonitrile 0.5% acetic acid

0.1% TFA

0.5% acetic acid

60% acetonitrile 0.1% TFA

1. For each sample to be desalted, two discs approximately 3mm in diameter were punched out from a larger disc of C18 (Empore Octadecyl C18, Sigma). This gave a binding capacity for peptides of around 2μg. These small discs were then packed into a 200μl pipette tip

2. C18 packed tips were inserted into a vacuum manifold and conditioned with 100μl 100% acetonitrile, followed by 100μl 50% acetonitrile 0.5% acetic acid and finally 100μl 0.1% TFA

3. Peptides were bound to C18

4. Peptides were washed with 100μl 0.1% TFA, followed by 100μl 0.5% acetic acid

5. Peptides were eluted with 30μl 60% acetonitrile 0.1% TFA and dried down prior to mass spectrometry analysis

2.13 Weak cation exchange (WCX) chromatography

Materials

Buffer pH 7:

7mM KH_2PO_4

30% acetonitrile

Buffer pH 8:

10mM Tris

Buffer pH 9:

10mM Tris

Buffer pH 10:

10mM CAPS (N-cyclohexyl-3-aminopropanesulfonic acid)

Buffer pH 11:

10mM CAPS

Buffer pH 12:

7mM KH_2PO_4

Buffer B:

7mM KH_2PO_4

1M KCl

pH 2.7

WCX was performed at a flow rate of 1-2ml/min

1. 5-10mg of tryptic peptides were diluted in 6ml of a binding buffer of the appropriate pH (pH selection based on the desired pI cut-off)

2. The WCX solid phase extraction (SPE) cartridge (6cc/500mg, Waters) was conditioned by passing through 6ml of 80% acetonitrile followed by 6ml water

3. The WCX SPE cartridge was then equilibrated by applying the binding buffer of the appropriate pH for 30 minutes

4. A blank run was performed by applying 12ml of the binding buffer with the desired pH (above) followed by 6ml Buffer B

5. The cartridge was re-equilibrated as above

6. Peptides were applied to the cartridge and flow-through collected using polypropylene tubes

7. The cartridge was washed with 12ml of binding buffer

8. Bound peptides were eluted with 6ml Buffer B

9. Flow-through and eluate fractions were lyophilized to remove acetonitrile. Samples were sent to Alta Biosciences for amino acid analysis after desalting

2.14 Desalting of peptides prior to hydrophilic interaction liquid chromatography

Materials

Buffer A:

Column volumes	% Buffer B
0	0
7.1	0
12	100
16	100
16.1	0
20	0

Table 2.3: Chromatography profile for desalting peptides (established by Thomas Uhlmann).

0.1% TFA

Buffer B:

95% acetonitrile 0.1% TFA

Peptides were desalted using an ÄKTA purifier (GE Healthcare) at a flow rate of 0.5ml/min

1. Peptides were applied to a C18 column (μRPC C2/C18 ST 4.6/100, GE Healthcare) and desalted according to the chromatographic profile (established by Thomas Uhlmann) in Table 2.3.

2. Fractions containing peptides (as revealed by absorbance at 215nm) were pooled and freeze dried overnight

2.15 Hydrophilic interaction liquid chromatography (HILIC)

Materials

Buffer A:

0.1% Formic acid

Buffer B:

95% acetonitrile 0.1% formic acid

HILIC was carried out using an Ettan-LC system (GE Healthcare) fitted with an autosampler A-905

1. 1mg of desalted, freeze-dried peptides were resuspended in 100µl Buffer B and 40µl of this was applied to a HILIC column (ZIC HILIC 150mm × 4.6mm, 3.5µm, 200Å, column volume 2.493ml, Merck). HILIC was performed according to the chromatographic profile outlined below.

2. Peptides eluting from 60% to 40% buffer B (hydrophilic fractions) were collected and every two fractions were pooled. These were dried down for analysis by mass spectrometry

Column volumes	% Buffer B
0	90
4.5	90
14.5	40
17	40
17.1	90
20.1	90

Table 2.4: Chromatography profile for HILIC (established by Thomas Uhlmann).

2.16 Mass spectrometry data acquisition

Peptides were separated on a 15cm by 75µm inner diameter picotip analytical column (New Objective) packed with Reprosil-Pur C18-AQ phase, 3µm bead (Dr. Maisch), run using an Ultimate 3000 nano HPLC (Dionex) in direct injection mode coupled to either a LTQ XL Orbitrap mass spectrometer or a Q Exactive mass spectrometer (Thermo Electron). A 120 minute or 180 minute gradient was used to separate peptides and each sample was injected 2-3 times.

The LTQ XL mass spectrometer was used in a "Top 5" data dependent acquisition mode and the Q Exactive mass spectrometer in a "Top 10" mode. Precursor scans were performed at a resolving power of 60,000 from which either 5 precursor ions (LTQ Orbitrap) or 10 precursor ions (Q Exactive) were selected and fragmented. Charge state +1 ions were rejected.

2.17 Mass spectrometry data analysis

MS/MS peak lists were converted to mzXML format using ReAdW version 4.4.1 (LTQ XL Orbitrap data) or to mgf format using ProteoWizard

msconvert release 3.0.3535 (Q Exactive data). Both tools were used with default parameters except that zlib compression of spectral data was enabled. Data was uploaded to the central proteomics facilities pipeline (CPFP at: https://cpfp-master.molbiol.ox.ac.uk/cpfp_demo/auth/login) (Trudgian et al., 2010). Data was searched against a concatenated target and reversed decoy version of the IPI human sequence database (EBI), versions 3.78 and 3.79 containing 86,702 and 86,635 target protein sequences respectively. Enzyme setting was trypsin or chymotrypsin allowing for up to 3 miscleavages. Carbamidomethyl cysteine was set as a fixed modification. For discovery of arginine methylation sites oxidized methionine, heavy oxidized methionine, monomethylarginine, heavy monomethylarginine, dimethylarginine and heavy dimethylarginine were set as variable modifications. For discovering lysine methylation, oxidized methionine, heavy oxidized methionine, monomethyllysine, heavy monomethyllysine, dimethyllysine, heavy dimethyllysine, trimethyllysine and heavy trimethyllysine were set as variable modifications. Mass tolerances for MS and MS/MS peak identifications were 10ppm and 0.5Da respectively (LTQ XL Orbitrap) or 20ppm and 0.1Da respectively (Q Exactive). Within CPFP, inference of protein identifications from peptide sequences was performed using ProteinProphet. Default parameters were used for protein grouping and the assignment of isoform identifications.

Assignment of a methylation site required identification by searches in the CPFP at 1-2.5% FDR and the presence of a confirming peptides in the precursor spectrum of equal intensity separated by a mass difference introduced by the light/heavy methyl group(s). The actual false positive rate of identification using the heavy methyl SILAC criteria (1.2%) was calculated

on a representative dataset by analysing all false positive hits for arginine methylation for the presence of confirming heavy methyl SILAC pairs. This number was doubled and divided by the total number of hits for arginine methylation.

Modification localisation scoring was performed using the ModLS algorithm which extends the PTMScore method to incorporate automatic specificity expansion (Trudgian et al., 2012). For each variable modification chosen for the database search, all amino acid specificities defined in the Unimod database are considered during localisation. Thus methylation could be assigned with higher probability to a residue other than arginine.

Peptide sequences were uploaded to ProteinCenter (Thermo Scientific) to obtain isoelectric points and GRAVY scores. Charges were calculated using a custom python script. All data plots were generated using custom python scripts and matplotlib (Hunter, 2007).

To calculate significance levels by bootstrapping a custom python script was written. All methylation sites were aligned on arginine with 5 flanking residues either side. Arginine was then removed. An amino acid was selected from each methylation site at random. This was repeated for the set of methylation sites 100,000 times. The probability of proline occuring with a given frequency in the set of methylation sites could then be calculated. In the case of hydrophobicity, hydropathy scores of randomly selected amino acids were averaged over one round of selection. The probability of a given hydropathy score occuring could then be calculated.

Primer name	Sequence (5'-3')
WASp_F	AGCCTCTAGAAAGCACCATGAGTGG
WASp_R	AGTGGTACCGTCATCCCATTCATCATC
WASp_R338K_F	GGGGTAACAAGGGTAAATCTGGTCC
WASp_R338K_R	GGACCAGATTTACCCTTGTTACCCC
WASp_R357K_F	CCACCAACACCCAAGGGACCC
WASp_R357K_R	GGGTCCCTTGGGTGTTGGTGG
IBA_F	GAGAACCCACTGCTTACTGGC
IBA_R	TAGAAGGCACAGTCGAGG

Table 2.5: Primers used to clone WASp and perform site directed mutagenesis.

Step	Temperature (°C)	Time (minutes)
1.	95	2:00
2.	94	0:15
3.	55	0:30
4.	72	1:30
5.	Go to step 2	30 times
6.	72	7:00

Table 2.6: PCR method used for cloning and mutagenesis of WASp.

2.18 Cloning and mutagenesis of the WASp gene

1. The coding sequence for human WASp was amplified from a cDNA clone (open biosystems) using the WASp_F and WASp_R primers (Table 2.5) according to the PCR program above

2. The PCR product was run on a 1% agarose gel and extracted using a Qiagen gel extraction kit

3. 2μg of the PCR product was digested with KpnI and XbaI (NEB) for 1 hour

Reagent	Volume (µl)
10mM dNTPs	1
Forward primer at 100µM	0.5
Reverse primer at 100µM	0.5
DNA template	100ng
HiFi buffer 10× (Roche)	5µl
Taq polymerase	0.75 (2.6U)
Water	to 50µl

Table 2.7: General PCR reaction mix.

Reagent	Volume (µl)
5× green GoTaq buffer (Promega)	2
25µM $MgCl_2$	0.4
10mM dNTPs	0.2
Forward primer at 100µM	0.25
Reverse primer at 100µM	0.25
GoTaq polymerase	0.1
Water	6.8

Table 2.8: Colony PCR reaction mix.

4. Digested product was run on a 1% agarose gel, purified as above and ligated into pEXPRIBA103, a plasmid containing a C-terminal Strep-tag. 180ng of the WASp gene was ligated with 150ng of pEXPRIBA103 (IBA) at 4°C overnight using T4 DNA ligase (Promega) and ligation buffer

5. Half of the ligation mixture was used to transform chemically competent DH5α*E. coli*

6. Transformed bacteria were plated on ampicillin plates (100µg/ml) and left to grow overnight

7. Colonies were screened by colony PCR (Table 2.8) using IBA_F and IBA_R primers (Table 2.5)

8. Positive colonies were grown up in LB ampicillin (100μg/ml) and larger quantities of plasmid were purified using the Qiagen maxi-prep kit

9. WASp mutants were generated (with Hannes Eichner) by performing splice overlap extension PCR on the WASp-pEXPRIBA103 construct using the mutagenesis primers in Table 2.5 and the PCR method (Table 2.6) and reaction mix above

2.19 Transient transfection of Jurkat Clone 20 cells

1. For each construct being expressed, 20×10^6 Jurkat cells were centrifuged at 400g for 5 minutes. 1ml of supernatant (conditioned medium) was put aside

2. Cells were washed twice in 10ml RPMI 1640

3. 10μg of plasmid was placed in each electroporation cuvette (4mm cuvette, Bio-Rad). WASp-pEXPRIBA103 plasmids were constructed in this work, PRMT-GFP plasmids were a kind gift from Mark Bedford

4. Cells were resuspended in 400μl RPMI and electroporated at 260V with 1 pulse of length 25msec (using a Gene pulser X-cell, Bio-Rad)

5. 800μl of conditioned medium was added to each cuvette and cells were transferred to a flask. Cells were cultured for 1-2 days before harvesting

2.20 Pull-down of Strep-tagged WASp

Materials

Modified RIPA buffer:

50mM Tris pH 7.5

125mM NaCl

1mM EDTA

1% NP-40

0.25% Deoxycholate

1. For each pull-down 10×10^6 Jurkat cells were lysed in 200µl modified RIPA buffer containing phosphatase and protease inhibitors
2. Lysate was cleared by centrifuging at 20,000g for 20 minutes at 4°C
3. For each pull-down, 10µl of Strep-Tactin resin (IBA) was washed 3 times in 200µl PBS
4. Pull-down was carried out for 1 hour at 4°C
5. Resin was pelleted at 400g for 2 minutes, supernatant removed and resin was washed 3 times in modified RIPA buffer at 4°C
6. Proteins were eluted by boiling resin in SDS loading buffer for 10 minutes

3. Induction of arginine methylation during T-lymphocyte activation

3.1 Introduction

3.1.1 T-lymphocyte activation

T-lymphocytes are extremely sensitive, being able to detect as few as 10 antigenic peptides complexed to MHC even in the presence of a vast excess of self-peptide MHC (1×10^6 complexes). When a T-lymphocyte contacts an antigen presenting cell (APC) such as a dendritic cell, the interaction only lasts seconds, during which time the T-lymphocyte must scan the excess of self-peptide-MHC complexes for the presence of rare antigenic peptide-MHC complexes (Miller et al., 2002, 2003). This high selectivity and sensitivity is necessary to avoid mounting a response to self peptides leading to autoimmunity, while still being able to detect the presence of a pathogen.

Within antigen presenting cells, proteins from internalised pathogens or translated viral proteins are cleaved by various proteases (e.g. the immunoproteasome). These peptides are then complexed with MHC molecules and presented on the cell surface. T-lymphocytes detect the presence of these

peptides using the T-cell receptor (TCR), a complex of proteins in the cell surface membrane. The TCR has a modular design, consisting of a recognition component (α and β chains) associated to a signalling component (CD3 complex). The α and β chains form a heterodimer and are each composed of two immunoglobulin domains: a constant (C) region and a highly variable (V) region. The V region is responsible for specific recognition of a peptide presented by MHC (the antigen). The CD3 complex consists of γ, δ, ε and ζ chains and is the starting point for the phosphorylation cascade following TCR triggering (van der Merwe and Dushek, 2011).

3.1.2 Signalling from the TCR and CD28 co-receptor

The precise molecular mechanism by which binding of the TCR to a peptide-MHC leads to activation of the TCR signalling pathway has not been clearly elucidated. One proposed mechanism is that structural changes in the TCR lead to signalling, while another proposes that clustering of the TCR in zones of contact with the APC leads to exclusion of inhibitory phosphatases and initiation of signalling. Whatever the initial mechanism is, the result is that the protein tyrosine kinase Lck is brought into contact with the cytoplasmic tails of the CD3 complex (Figure 5.1). Lck then phosphorylates tyrosine residues within immunoreceptor tyrosine based activating motifs (ITAMs) found in the γ, δ, ε and ζ chains of CD3. A second tyrosine kinase, Zap-70, then binds to the phosphorylated ITAMs and is phosphorylated by Lck. Once Zap-70 has become activated, it phosphorylates the scaffold molecules LAT

and SLP-76, which become associated via the adaptor Gads. The phosphorylated LAT-SLP76 complex recruits numerous proteins leading to activation of many downstream signalling pathways. One important pathway leads to remodelling of the actin cytoskeleton and is initiated when the adaptor Nck binds to phosphorylated SLP-76. Nck in turn recruits Wiskott-Aldrich Syndrome protein (WASp) which directs the polymerisation of actin filaments (Acuto et al., 2008).

The fact that the TCR has recognised an antigenic peptide is thus transmitted into the cell via a cascade of protein tyrosine phosphorylation, ultimately reaching the nucleus and causing alterations in gene transcription. The early wave of tyrosine phosphorylation is readily detectable using an antibody against phospho-tyrosine such as 4G10 (Druker et al., 1989). This is therefore a good example of a post-translational modification acting as an information carrier and directing cellular behaviour in response to a stimulus. However, signalling from the TCR alone is not sufficient to activate a T-lymphocyte, a signal from the co-receptor CD28 is also required, ensuring that T-lymphocyte activation only takes place in the proper context (Acuto and Michel, 2003).

A mature dendritic cell expresses two ligands of CD28 on its surface, B7.1 (CD80) and B7.2 (CD86). When the TCR recognises a peptide-MHC complex on the dendritic cell, the CD28 receptor is also activated by CD80/CD86. This leads to tyrosine phosphorylation of CD28 and recruitment of phosphoinositide 3-kinase (PI3 kinase) (Boomer and Green, 2010). PI3 kinase generates phosphatidylinositol 3, 4, 5 triphosphate (PIP3) at the cell surface membrane which recruits phospholipase C (PLC)-γ1, an enzyme that

hydrolyses phosphatidylinositol 4,5 bisphosphate to the second messengers diacylglycerol and inositol triphosphate. These second messengers then go on to activate pathways leading to gene transcription. Phospholipase C (PLC)-$\gamma 1$ is also activated downstream of signalling from the TCR, thus the CD28 co-receptor and the TCR synergise to generate the appropriate signals leading to T-lymphocyte activation (Michel et al., 2001).

3.1.3 T-lymphocyte activation *in vivo*

Although ligating CD28 and CD3 *in vitro* is sufficient to activate T-lymphocytes leading to proliferation, *in vivo* the T-lymphocyte integrates many signals leading to activation and differentiation into subtypes. For example, the co-receptors CD4/CD8 bind to MHC class II and class I respectively, enhancing signalling. The adhesion molecules LFA-1, CD2 and CD44, which are up-regulated upon activation, promote the interaction between T-lymphocyte and APC, leading to sustained signalling. OX40 and ICOS are also important co-stimulatory molecules upregulated on activated T-lymphocytes. They bind to OX40L and ICOSL respectively which are found on activated dendritic cells, and promote the survival and proliferation of T-lymphocytes (Figure 3.1). Other important signals are induced by cytokines secreted by the dendritic cell. The set of cytokines secreted will depend on the type of pathogen encountered by the dendritic cell, and this will in turn determine the subset of T-lymphocyte generated. Cytokines bind to receptors on T-lymphocytes and activate the JAK-STAT signalling pathway. As a result,

signal transducers and activators of transcription (STATs) transcription factors become phosphorylated and dimerise. STAT dimers then regulate transcription of genes necessary for growth and differentiation (Murphy, 2012; DeFranco et al., 2007). T-lymphocyte activation *in vivo* is characterised by

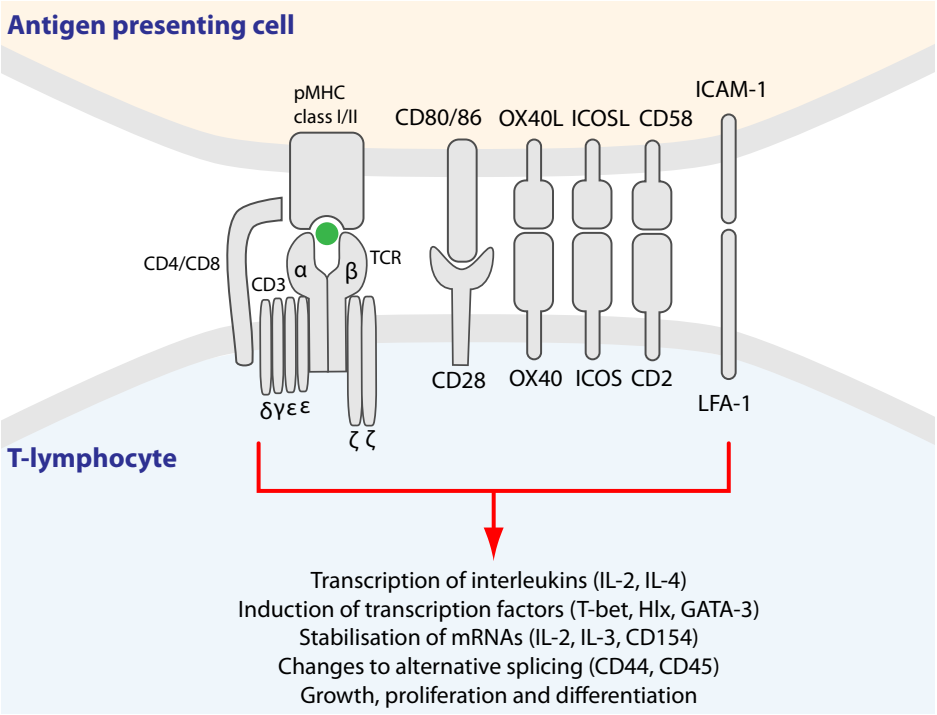


Figure 3.1: The immunological synapse (IS) between an antigen presenting cell (APC) and a T-lymphocyte. The TCR recognises a peptide complexed to MHC. Numerous other co-receptors bind to their ligands on the APC (some of which are shown here), generating an area of close contact and inducing numerous signalling pathways. These signalling pathways lead to upregulation of many genes and growth, proliferation and differentiation of the T-lymphocyte.

three phases (Mempel et al., 2004). In the first phase lasting 8 hours, T-lymphocytes enter the lymph node and make brief contacts with dendritic cells, each contact lasting 5-10 minutes. In the second phase lasting 12 hours, T-lymphocytes make more stable contacts with dendritic cells, with

contact time lasting hours. Finally T-lymphocytes begin to proliferate in the third phase, dividing every 6-8 hours, and make continued brief contacts with dendritic cells.

3.1.4 Transcriptional regulation in T-lymphocyte activation and differentiation

Prior to activation, naïve T-lymphocytes are small quiescent cells with low metabolic activity. DNA is tightly packaged into heterochromatin and transcription and translation take place at low levels. After activation, T-lymphocytes become more metabolically active and upregulate transcription and translation. They enter the G1 phase of the cell cycle and begin to proliferate. These dramatic changes in cell phenotype are driven by alterations in gene transcription and mRNA processing (Rawlings et al., 2011).

The differentiation of activated T-lymphocytes into subsets is directed by distinct transcriptional regulators. Differentiation into T_H1 cells is driven by T-bet which activates transcription from several target genes including IFN-γ and the transcription factor Hlx (Schulz et al., 2009). T-bet recruits chromatin remodifying complexes, including the H3K4me2 methyltransferase complex, leading to a chromatin state permissive for transcription at target genes (Miller et al., 2008, 2010). Expression of IFN-γ and Hlx lead to repression of IL-4 production, thus inhibiting a T_H2 cytokine (Kaneko et al., 2007).

Differentiation to T_H2 occurs due to expression of the GATA-3 transcrip-

tion factor which binds to promoters of T_H2 target genes including IL-4, IL-5 and IL-13 (Zheng and Flavell, 1997). Like T-bet, binding of GATA-3 to promoters leads to chromatin remodelling through the methylation of H3K4. This makes the promoters of target genes accessible to other transcription factors such as NF-AT. In addition to promoting T_H2 differentiation, GATA-3 also suppresses the T_H1 differentiation pathway by blocking Runx3 mediated IFN-γ production and preventing STAT4 and IL-12Rβ up-regulation (Zhu et al., 2006). T_H17 differentiation is driven by the transcription factors RORγt and RORα, while the transcription factor Foxp3 specifies differentiation to T_{REG} cells (Dong, 2011; Feuerer et al., 2009).

3.1.5 Post-transcriptional regulation in T-lymphocyte activation and differentiation

Post-transcriptional control of protein expression by mRNA binding proteins also plays a crucial role in T-lymphocyte activation. Many transiently expressed genes and genes expressed shortly after a stimulus contain AU rich elements (AREs) and GU rich elements (GREs) in the 3' untranslated region, conferring instability and a short half life on the transcript as well as affecting translational efficiency (Shaw and Kamen, 1986; Anderson, 2010). This enables rapid control of protein levels by altering the rate of mRNA degradation and translation. The mRNAs coding for cytokines and immediate-early genes involved in growth such as c-*fos* contain AREs (Stoecklin et al., 1994). Triggering of the TCR and CD28 signalling pathways during T-lymphocyte

activation leads to stabilisation of mRNAs coding for cytokines such as IL-2, IL-3 and IFN-γ (Lindstein et al., 1989). Input from both signalling pathways is required for this effect, triggering only the TCR or CD28 does not alter the stability of these mRNAs. In addition, cytokine mRNAs may also be blocked from being translated, this latter situation occurs in self-reactive T-lymphocytes to prevent cytokines being expressed (Villarino et al., 2011).

A well known example of an RNA binding protein regulating cytokine mRNA translation in T-lymphocytes is Tristetraprolin (TTP) (Ogilvie et al., 2005). TTP binds to the AREs of several mRNAs including those coding for IL-2, IL-6, IFN-γ, and C-FOS. This leads to recruitment of proteins that mediate mRNA decay such as the 5'-3' exoribonuclease XRN1. The activity of TTP is regulated by p38 MAP kinase mediated phosphorylation which causes TTP to be excluded from cytoplasmic foci where mRNAs accumulate, thereby preventing TTP from destabilising transcripts. The activity of p38 can in turn be regulated by upstream pathways such as TCR signalling. Mice lacking TTP suffer from arthritis and dermatitis caused by over expression of pro-inflammatory cytokines (Taylor et al., 1996).

CD154, expressed on activated CD4+ T-lymphocytes and a ligand for CD40 on APCs, is a well documented example of a protein whose expression is post-transcriptionally controlled. During the initial 12 hours of T-lymphocyte activation, CD154 mRNA is relatively unstable with a half life of less than 40 min. This half life increases to over 2 hours after a further 12 hours of stimulation. The increase in stability of the CD154 transcript appears to be due to the assembly of an activation-dependent ribonucleoprotein (RNP) complex at the 3'UTR of the mRNA (Barnhart et al., 2000).

The translation machinery can also be regulated to control protein expression. *In vivo* T-lymphocytes are first primed by dendritic cells in secondary lymphoid tissue and subsequently migrate to sites of infection where they secrete cytokines and co-ordinate the immune response. During the priming phase, the TCR and CD28 signalling pathways are triggered leading to increases in transcription of cytokines but no increases in cytokine protein levels. This uncoupling of transcription and translation is thought to be caused by initiation of the integrated stress response, during which eukaryotic initiation factor 2 α (eIF2α) becomes phosphorylated, inhibiting its activity leading to translational arrest (Scheu et al., 2006). A second stimulation of T-lymphocytes at the site of infection then relieves this phosphorylation so that cytokine mRNAs can be rapidly translated.

3.1.6 Alternative splicing in T-lymphocyte activation and differentiation

As a result of signalling from the TCR and CD28 co-receptor, large scale alterations in alternative splicing take place. A recent study in T-lymphocytes found that 178 exons in 168 genes were differentially included in response to activation, these exons possessed an activation response sequence (ARS) which appears to be required for TCR dependent regulation of splicing (Martinez et al., 2012). This study was carried out in a T-lymphocyte cell line which is constitutively metabolically active. It is likely that a similar analysis in primary human T-lymphocytes would identify more dramatic alterations

in alternative splicing upon activation, since in primary T-lymphocytes, activation results in the cell being transformed from a quiescent state to a highly active proliferating state.

Splicing of CD45 is a well known specific example of activation dependent alternative splicing. CD45 is an abundant transmembrane protein tyrosine phosphatase acting as both a positive and negative regulator of TCR signalling (McNeill et al., 2007). Up to 8 different isoforms of CD45 can be generated as a result of alternative splicing of exons 4-6 (McNeill et al., 2004). Naïve T-lymphocytes express the largest isoform of CD45, whereas activated and memory T-lymphocytes express the shortest isoform. It is thought that the smaller isoform of CD45 leads to attenuation of TCR signalling. The alteration in splicing of CD45 after T-lymphocyte stimulation is dependent on newly synthesised proteins as cycloheximide treatment blocks this switch (Lynch and Weiss, 2000). An RNA binding protein, hnRNPLL, was identified in an RNAi screen as being up-regulated 24 hours after stimulation and necessary for exclusion of exons 4-6 in CD45 mRNA. Knockdown of hnRNPLL in primary CD4+ T-lymphocytes and subsequent exon array analysis revealed that hnRNPLL is a master regulator of alternative splicing and is necessary for the correct processing of numerous mRNAs important in proliferation and differentiation (Oberdoerffer et al., 2008).

Another example of TCR signalling dependent splicing is the processing of CD44 mRNA. CD44 is a cell surface protein important for migration of T-lymphocytes. The CD44 gene contains 10 variable exons which can be excluded or included in varying combinations generating up to 20 different forms of mRNA (Lynch, 2004). In naïve T-lymphocytes all variable exons

are excluded, after 20 hours of stimulation, various isoforms of CD44 are expressed with different exons being included (König et al., 1998). This TCR signalling regulated splicing was shown to be independent of *de novo* protein synthesis, suggesting a direct link between signal transduction and the splicing machinery.

The above examples illustrate the point that activation and differentiation of T-lymphocytes leads to extensive changes in transcriptional regulation and alterations in mRNA processing. Given that the majority of known arginine methylated proteins are involved in transcription and/or mRNA processing and that arginine methylation clearly has an important role in these processes, it is possible that activation of T-lymphocytes may lead to altered levels of arginine methylation. Arginine methylation could therefore play an important role in coordinating changes in transcription/mRNA processing after activation. Indeed, induction of arginine methylation has been observed during neuronal cell differentiation. Cimato *et al* applied nerve growth factor (NGF) to PC12 cells and observed increases in arginine methylation of several nuclear proteins within 5 hours (Cimato et al., 2002). They also found an NGF dependent increase in PRMT1 activity. Mowen *et al* found that activation of T_H1 or T_H2 T-lymphocytes for 24 hours led to increases in PRMT1. Finally, rapid induction of arginine methylation and increases in arginine methyltransferase activity were observed within 10 minutes of stimulating PHA blasts (primary T-lymphocytes which have become activated and proliferated in response to the mitogen phytohemagglutinin) (Blanchet et al., 2005).

3.1.7 Aims

There is some evidence suggesting that arginine methylation is induced upon T-lymphocyte activation. However it is not clear if arginine methylation is induced when naïve T-lymphocytes are stimulated, and whether induction is rapid or only occurs after prolonged stimulation. In addition it is not known whether T-lymphocyte cell lines exhibit any changes in arginine methylation upon activation. The aim of the work in this chapter was to determine whether levels of arginine methylation are altered during T-lymphocyte stimulation by stimulating primary T-lymphocytes and Jurkat T-lymphocytes under various conditions and following levels of arginine methylation by western blot.

3.2 Results

3.2.1 Short term stimulation of T-lymphocytes

In order to determine whether arginine methylation was rapidly induced upon activation as reported by Blanchet et al. (Blanchet et al., 2005), the Jurkat Clone 20 T-lymphocyte cell line was stimulated in solution using beads coated with antibodies against the CD3εsignalling component of the TCR and the CD28 co-receptor. Cells were taken at 0, 2, 10 and 30 minutes post-stimulation, lysed and proteins were separated by SDS-PAGE. Western blotting with an antibody against phospho-tyrosine (4G10) was carried out which showed the stimulation was successful as several proteins became phosphorylated at 2 minutes of activation (Figure 3.2, A). The same samples were then run on SDS-PAGE but western blotting was performed using an antibody against methylated arginine (7E6). Using this antibody no increases in arginine methylation were observed within 30 minutes of stimulation (Figure 3.2, B).

Primary T-lymphocytes were also stimulated to check for any immediate changes in arginine methylation. Peripheral Blood Mononuclear Cells (PBMCs) were isolated from human blood by separation on a Ficoll cushion. PBMCs consist of 60% T-lymphocytes (of which 70% are CD4+ and 30% are CD8+), 15% monocytes/macrophages, 10% B-lymphocytes and 15% natural killer cells (Rowland-Jones and McMichael, 2000). Immediate increases in tyrosine phosphorylation are difficult to observe when stimulating primary T-lymphocytes with antibody coated beads, therefore Raji antigen presenting

cells were used loaded with Staphylococcus enterotoxin A, B and C (SEA, SEB, SEC). These are superantigens that bind to MHC class II and are capable of triggering the TCR without requiring processing into peptides. Raji APCs also express ligands for CD28. By stimulating PBMCs with superantigen loaded Raji APCs induction of tyrosine phosphorylation could be observed (increase in phosphorylation of a protein at 70kDa and 37kDa, Figure 3.3, A). This indicates successful stimulation, however when the samples were run again on SDS-PAGE and western blotting was carried out with the 7E6 anti-methylarginine antibody, no convincing increases in arginine methylation could be seen (Figure 3.3, B).

3.2.2 Long term stimulation of T-lymphocytes

Short term stimulation of T-lymphocytes did not lead to alterations in levels of arginine methylation. Previous studies showed increases in arginine methylation in PC12 cells occurred hours after application of stimulus, suggesting arginine methylation is a later event in cell proliferation and differentiation. Therefore Jurkat cells were stimulated for 24 hours with beads either uncoated, coated with CD3 antibody, coated with CD28 antibody or coated with both CD3 and CD28 antibodies. Cells were lysed and proteins separated by SDS-PAGE. No discernable alterations in arginine methylation could be seen with the 7E6 antibody (Figure 3.4).

Since Jurkat Clone 20 is a leukemic T-lymphocyte cell line, it is highly metabolically active and in a constant state of proliferation. Basal arginine

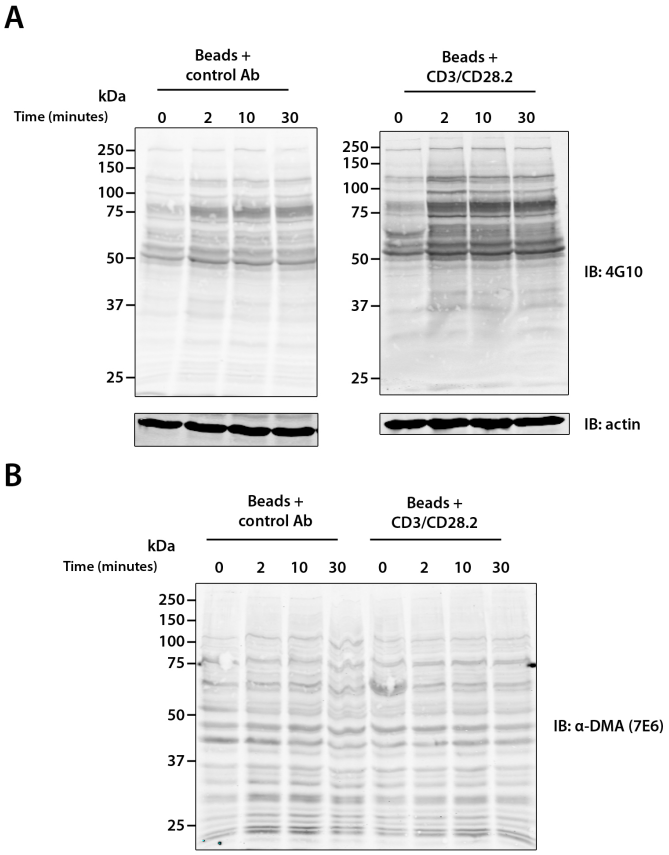


Figure 3.2: Short term stimulation of Jurkat Clone 20 T-lymphocytes does not lead to changes in levels of arginine methylation. Jurkat cells were incubated with beads coated with IgG or CD3 and CD28 antibodies. Samples were taken at 0, 2, 10 and 30 minutes, proteins were separated by SDS-PAGE and immunoblotted with anti-phosphotyrosine (4G10, panel A) or anti-methylarginine (7E6, panel B).

methlylation levels in these cells may therefore be high and changes in levels after stimulation may not occur. PBMCs isolated from human blood were therefore stimulated with antibody coated beads for 48 hours. This led to activation and proliferation of primary T-lymphocytes. Samples were taken at 4, 19, 26 and 43 hours and levels of arginine methylation were assessed by western blotting with three different antibodies against methylarginine, 7E6

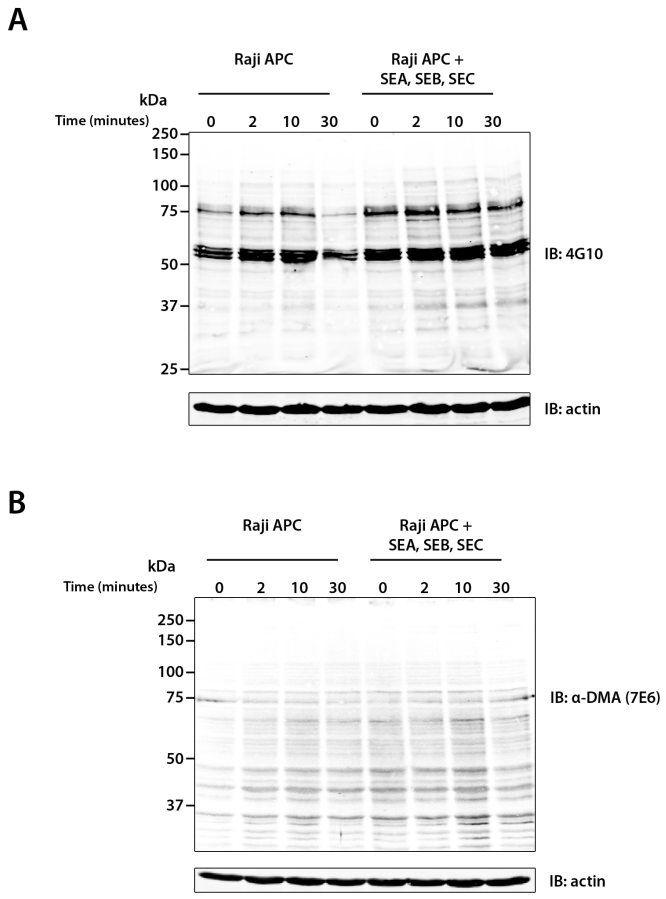


Figure 3.3: Short term stimulation of PBMCs isolated from human blood does not lead to changes in levels of arginine methylation. PBMCs were stimulated with Raji cells either not loaded or loaded with superantigen. Samples were taken at 0, 2, 10 and 30 minutes, proteins were separated by SDS-PAGE and immunoblotted with anti-phosphotyrosine (4G10, panel A) or anti-methylarginine (7E6, panel B).

(Figure 3.5, A), Asym24 (Figure 3.5 B) and Sym10 (Figure 3.5, C). 7E6 is claimed to be specific for mono- or dimethylarginine and recognises asymmetrically or symmetrically dimethylated arginine but not unmodified arginine. It was raised against asymmetrically dimethylated arginine. Asym24 is thought to recognise asymmetrically dimethylated arginine and was raised

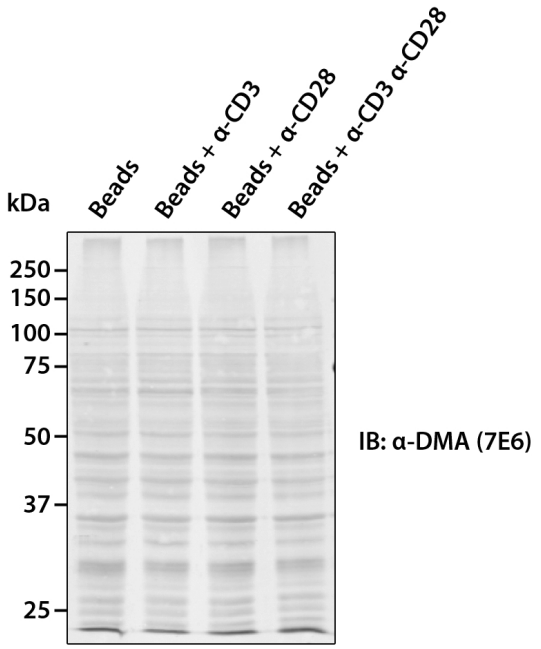


Figure 3.4: Stimulation of Jurkat cells does not lead to changes in arginine methylation even at later time points. Jurkat cells were cultured for 24 hours in the presence of beads either uncoated, coated with CD3 antibody, coated with CD28 antibody or coated with both CD3 and CD28 antibodies. Cells were harvested, lysed and proteins separated by SDS-PAGE. Arginine methylated proteins were detected by immunoblotting with 7E6 antibody.

against the peptide KGR*GR*GR*GR*GPPPPPR*GR*GR*GR*G (R* is an asymmetrically dimethylated arginine). Sym10 recognises symmetrically dimethylated arginine and was raised against a peptide containing four SDMA-glycine repeats.

Immunoblotting with the 7E6 antibody revealed at least 6 proteins (Figure 3.5, A) giving stronger signals after 20 hours of stimulation. Similarly, blotting with the Asym24 antibody showed at least 5 proteins giving stronger signals after 20 hours of stimulation (Figure 3.5, B). The Sym10 antibody revealed fewer proteins compared to the other antibodies, the methylation

state of the few recognised proteins appeared to increase at 26 hours of stimulation (Figure 3.5, C).

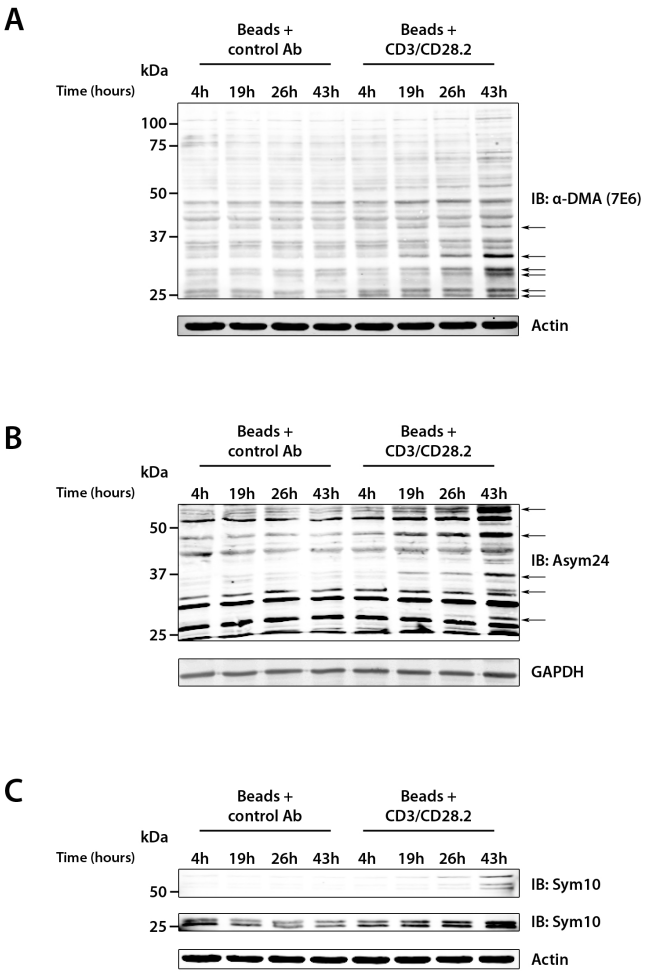


Figure 3.5: Long term activation of PBMCs leads to increases in arginine methylation. PBMCs isolated from human blood were cultured in the presence of beads either coated with a control antibody or CD3 and CD28.2 antibodies. Cells were taken at 4, 19, 26 and 43 hours after stimulation, lysed and analysed by western blotting with the 7E6 antibody **(A)**, Asym24 antibody **(B)** and Sym10 antibody **(C)**. Actin and GAPDH are loading controls. Arrows indicate proteins whose methylation increases as a result of the stimulation.

A previous study had shown up-regulation of PRMT1 after prolonged

stimulation of T_H1 or T_H2 cells (Mowen et al., 2004). Therefore samples from stimulated PBMCs were blotted with two antibodies against PRMT1, a rabbit antibody (Figure 3.6, A) and a mouse antibody (MAT-B12, Figure 3.6, B). Levels of PRMT4 and PRMT5 were also assayed (Figure 3.6, C). Three isoforms of PRMT1 are clearly visible (v1, v2, v3), levels of all three increase from 20 hours after stimulation begins. In addition, higher molecular weight forms of v1 and v2 PRMT1 isoforms are observed with both antibodies. The distance between these higher molecular weight forms is consistent with the addition of a 9kDa modification which could represent ubiquitin or an ubiquitin like modifier. Levels of PRMT4 are not altered upon activation, while levels of PRMT5 show a 2 fold increase (Figure 3.6, C). To check whether ubiquitination in general increases upon T-lymphocyte activation, samples were immunoblotted with an antibody against ubiquitin (Figure 3.7). This revealed that at 20 hours of stimulation there is an increase in ubiquitin signal from proteins >75kDa.

3.2.3 Long term stimulation of CD4+ T-lymphocytes

The previous stimulations had been carried out using PBMCs derived from human blood which are a mixture of cells including T-lymphocytes. In order to confirm that increases in arginine methylation originated from the activated T-lymphocytes, CD4+ T-lymphocytes were purified from human blood and cultured for 48 hours either with beads coated with a control antibody, or with beads coated with the activating CD3 and CD28 antibodies.

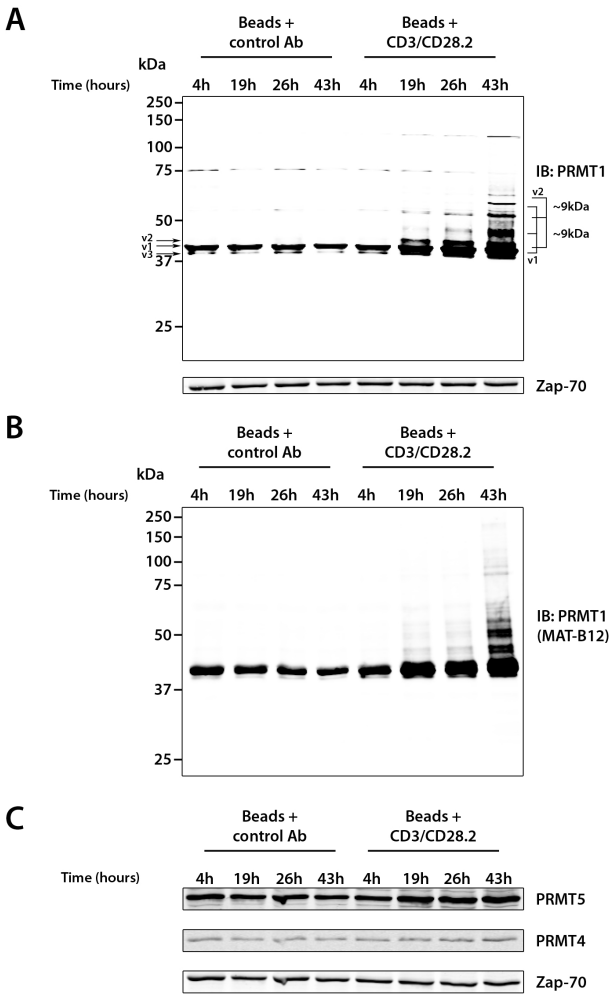


Figure 3.6: Increases in levels of PRMT1 and PRMT5 in PBMCs after 20 hours of stimulation. **(A)** PBMCs isolated from human blood were stimulated with beads coated with CD3 and CD28 antibody and samples taken at the indicated times. Cells were lysed and analysed by western blotting with an antibody against PRMT1 raised in rabbit. Zap-70 is shown as a loading control. **(B)** As in (A) except an antibody against PRMT1 raised in mouse was used. **(C)** As in (A) but samples were immunoblotted with antibodies against PRMT4 and PRMT5.

Cells were lysed and proteins separated by SDS-PAGE. Levels of arginine methylation were assayed with the methyl-specific antibodies Asym24 and Sym10 (Figure 3.8 A and B), confirming that several proteins increase in

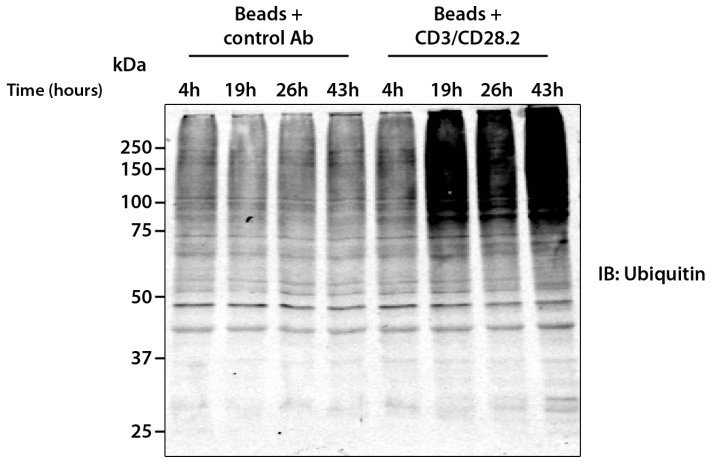


Figure 3.7: Increases in ubiquitination after long term stimulation of PBMCs. Samples from Figure 3.6 A were immunoblotted with an antibody against ubiquitin.

arginine methylation when T-lymphocytes are activated. Levels of PRMT1 also increase and higher molecular weight forms appear, similar to those seen when PBMCs are activated.

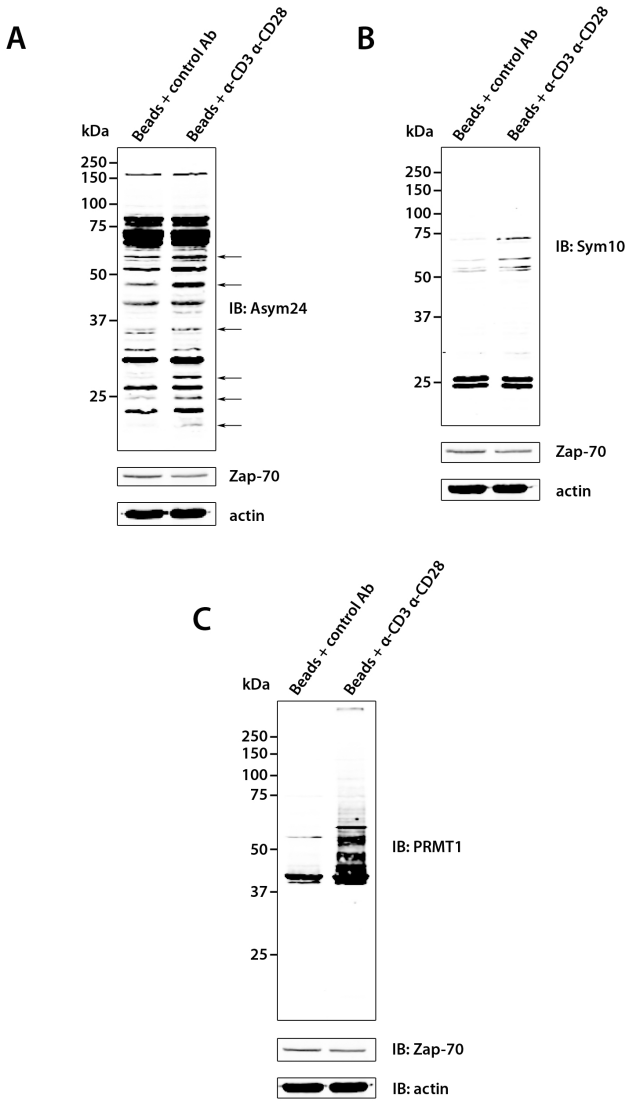


Figure 3.8: Increases in arginine methylation and levels of PRMT1 after stimulation of CD4+ human primary T-lymphocytes. CD4+ human T-lymphocytes were isolated from blood and stimulated with antibody coated beads for 48 hours. **(A)** Immunoblot with antibody against asymmetrically dimethylated arginine, **(B)** Immunoblot with antibody against symmetrically dimethylated arginine, **(C)** Immunoblot with antibody against PRMT1.

3.3 Discussion

The question of whether activation of T-lymphocytes leads to alterations in levels of arginine methylation has thus far not been resolved since the scientific literature offers mixed reports. Mowen et al. showed that activation of T_H1 or T_H2 for 24 hours leads to increases in levels of PRMT1 and PRMT4, and that T_H1 and T_H2 cells had higher basal levels of PRMT1 and PRMT4 compared to T-helper precursor cells (Mowen et al., 2004). However levels of arginine methylation were not assayed in these experiments, although given the increases in PRMT enzymes observed, corresponding increases in arginine methylation are likely. Using the Jurkat T-lymphocyte cell line, Richard et al. showed that after 3 hours of activation no increases in PRMTs or arginine methylation (as revealed using Asym24 and Sym11 antibodies) were observed (Richard et al., 2005). Finally Blanchet et al. demonstrated that re-stimulation of primary T-lymphocytes led to rapid induction of arginine methlyation within 10 minutes.

In order to elucidate under what conditions arginine methylation was altered in T-lymphocytes, both primary T-lymphocytes and Jurkat T-lymphocytes were stimulated for short or longer periods and levels of arginine methylation were determined by western blotting. Stimulation of Jurkat T-lymphocytes for short periods (<30 minutes) did not lead to any increases in arginine methylation (Figure 3.2), corroborating the report by Mowen et al.. Similarly, stimulating primary T-lymphocytes for short periods also did not lead to any increases in arginine methylation. Triggering a strong tyrosine phosphorylation response by activating primary T-lymphocytes is challenging and

often only weak induction of phosphorylation is observed (Figure 3.3, A). Nonetheless an increase in tyrosine phosphorylation was seen indicating successful stimulation, making it unlikely that the failure to see induction of arginine methylation was due to the cells not being stimulated. The lack of arginine methylation response after a short period of stimulation is in contrast to the findings of Blanchet et al.. However the primary cells used in that study were first stimulated and expanded for 9-11 days using phytohemagglutinin (PHA) to generate PHA-blasts (Blanchet et al., 2005). These were then starved of IL-2 and stimulus for 2-3 days before being re-stimulated with antigen loaded APCs. Thus the cells had already been pre-activated with PHA for an extended time before the increase in arginine methylation was demonstrated, making the physiological significance of this finding unclear. Moreover, the specificity of the 7E6 antibody in western blots is questionable since treating cells with a methylase inhibitor (Adox) does not lead to a decrease in signal from the 7E6 antibody (Chen et al., 2004). The antibodies Asym24 and Sym10 have been shown to be more specific (Chen et al., 2004) and it is for this reason that they were used to confirm increases in arginine methylation in the present study.

Blanchet et al. showed that Vav1, a component of the T-cell receptor signalling pathway, was arginine methylated upon T-lymphocyte activation and that this was linked to nuclear translocation (Blanchet et al., 2005). A later study failed to reproduce this finding, instead it seems as though Vav1 may be constitutively arginine methylated, at least as detected by the 7E6 antibody (Lawson et al., 2007).

Stimulation of Jurkat T-lymphocytes for 24 hours also did not lead to any

observable increases in arginine methylation (Figure 3.4). However, stimulation of primary T-lymphocytes for longer time points led to increases in arginine methylation and PRMTs 1 and 5 after 20 hours (Figures 3.5, 3.6, 3.8). It appears that increases in arginine methylation cannot be induced in Jurkat T-lymphocytes upon activation. This may be due to the fact that Jurkat T-lymphocytes are highly metabolically active and may have high basal levels of arginine methylation. They proliferate rapidly and cannot differentiate into subtypes upon activation suggesting that activation and cell division/differentiation are uncoupled (Schneider et al., 1977; Abraham and Weiss, 2004). Unstimulated primary T-lymphocytes on the other hand are quiescent cells in the G_0 phase of the cell cycle. Prolonged stimulation of the TCR and CD28 co-receptor causes cells to enter S-phase after 20 hours of stimulation and the first cell division takes place after 30-40 hours of stimulation (Akbar et al., 1989; Mire-Sluis et al., 1987). The onset of S-phase therefore corresponds to the observed increases in arginine methylation and PRMT1 levels. S-phase and cellular division require synthesis of many new proteins. Transcription, mRNA processing and translation are increased meaning a higher demand for arginine methylation which is involved in all of these processes.

Protein methylation can be rapidly induced and is a highly dynamic process in certain circumstances, for example during bacterial chemotaxis. In chemotaxis binding of attractant leads to dephosphorylation of CheY causing it to dissociate from the receptor leading to reduced tumbling. In addition, the bacterium must also sense changes in attractant concentration (compare current levels with previous levels) which requires memory. This memory

is in the form of methylation of glutamate residues on the attractant receptor which occurs when attractant binds (Li and Stock, 2009). Bacterial chemotaxis relies on phosphorylation to provide signalling and methylation to provide memory. This could be analogous to T-lymphocyte activation, where phosphorylation provides shorter term signalling and arginine methylation is a longer term more sustained signal that imparts a memory on the cell by causing a stable change in transcription and mRNA processing.

The Sym10 antibody is known to recognise by western blot symmetrically dimethylated forms of SmB (30kDa), SmB' (32kDa) and coilin (60kDa) which are important in the formation of the splicing machinery (Boisvert et al., 2002). Proteins with corresponding masses were indeed observed to be increasingly methylated upon T-lymphocyte activation (Figure 3.5, C, 3.8, B). The arginine methylation of transcription factors and mRNA processing factors could lead to establishment of a stable transcriptional and splicing program leading to differentiation of the cell.

Although several reports have suggested that arginine methylation may be involved in proximal TCR signalling, no evidence for that was found here. The findings of Blanchet et al. have not been reproduced in naïve T-lymphocytes. Lawson et al. showed that a methylase inhibitor suppressed TCR signalling in CD4 T-lymphocytes but enhanced it in CD8 T-lymphocytes suggesting a role for arginine methylation in proximal TCR signalling (Lawson et al., 2007). Unfortunately, as with many studies on arginine methylation, the methylase inhibitor used blocks many methylation reactions and could affect the expression of many proteins, making it impossible to attribute effects to arginine methylation.

Increases in PRMT1 during T helper lymphocyte activation observed by Mowen et al. (Mowen et al., 2004) were also seen in the present study (Figure 3.6 and 3.8, C). PRMT1 is the predominant arginine methyltransferase in terms of activity and can occur in up to 7 different splice forms named v1-v7. In unstimulated cells the main PRMT1 band detected likely corresponds to v1 (Figure 3.6, A), the band below this could be v3-v7 as they are very similar in molecular weight (Goulet et al., 2007). The v2 form appears after stimulation and all three bands corresponding to splice forms of PRMT1 begin to increase in intensity 20 hours after stimulation. In addition, higher molecular weight forms of PRMT1 appear. These higher bands are not likely to be due to non-specific binding of the PRMT1 antibody used in western blotting as another antibody raised in mouse against a different portion of PRMT1 also reveals these forms (Figure 3.6, B). There is a possibility that other PRMTs are being detected. PRMT4 and PRMT5 levels were assayed, ruling these out and only PRMTs 2, 3, 7 and 9 could be large enough to account for the higher molecular weight forms. It is unlikely both PRMT1 antibodies are detecting all these PRMTs, and they are too large to be other splice forms of PRMT1. It is more likely these higher molecular weight forms represent modified PRMT1, possibly PRMT1 conjugated to ubiquitin or an ubiquitin like modifier. If the v1 and v2 splice forms are assumed to become modified, the spacing between the unmodified, single and double modified forms corresponds to 9kDa, the molecular weight of ubiquitin or an ubiquitin like modifier.

Ubiquitin can be conjugated via its carboxy terminus to lysine side chains in proteins and target them for degradation by the proteasome (Kessler,

2006). Ubiquitin itself has several lysine chains which can be linked to other ubiquitin molecules forming polyubiquitin chains. Polyubiquitin chains formed by linkage of lysine 48 in ubiquitin tend to result in proteasomal degradation of the protein. Polyubiquitination via lysine 63 (K63) or other lysine residues in ubiquitin can result in a non-proteolytic outcome (Jiang and Chen, 2012). For example, ubiquitination can lead to kinase activation as in the case of RIP1 which becomes K63 polyubiquitinated upon TNFα stimulation, leading to recruitment of kinase complexes containing ubiquitin binding domains (Li et al., 2006; Ea et al., 2006). The recruited kinases then activate downstream signalling pathways, polyubiquitin is therefore acting as a scaffold to recruit other proteins. This could also be the case for PRMT1, where polyubiquitination may lead to recruitment of other proteins which determine substrate specificity and/or localisation. The modification may not be ubiquitin, however general protein ubiquitination is upregulated after T-lymphocyte activation making it a likely candidate (Figure 3.7).

PRMT1 is known to be regulated by the proteins BTG1 and BTG2 which can bind and augment activity (Lin et al., 1996). PRMT1 has also been shown to be phosphorylated on tyrosine 299, but the functional effect of this modification has not been demonstrated (Rikova et al., 2007). The various splice forms of PRMT1 have altered substrate preference at least *in vitro*, and they also show altered subcellular localisation. PRMT1v2 has an N-terminal nuclear export signal and is predominantly cytoplasmically localised (Goulet et al., 2007), it is also this form which is highly upregulated upon T-lymphocyte activation (Figure 3.6 and 3.8, C). This form of PRMT1 may methylate proteins involved in translation of mRNA and could

also co-translationally methylate proteins. Ubiquitination may represent an important way in which the activity and localisation of the PRMTs are regulated.

Here, increases in arginine methylation and PRMT1 have been shown in CD4+ T-lymphocytes, however it is likely that CD8+ T-lymphocytes will also show a similar up-regulation. CD4+ T-lymphocytes are a mixture of naïve cells (those that have not encountered antigen and become activated) and memory cells. Memory cells make up 50-60 % of the CD4+ population in a normal healthy adult (Nausch et al., 2012). Therefore it is not certain whether alterations in arginine methylation are observed in both naïve and memory T-lymphocytes, or if this is limited to one subset. In fact these changes may not be specific to lymphocytes but could be a general cell-cycle related phenomenon. Evidence for this comes from an early study by Lee et al. who observed increases (over two fold) in levels of an arginine methyltransferase from G_1 to G_2 phases of the cell cycle in HeLa cells (Lee et al., 1973). In a more recent study, arginine methylation was monitored through the cell cycle (also in synchronised HeLa cells) (Kim et al., 2010a). Three proteins of 80kDa, 68kDa and 64kDa showed increased reactivity to the Asym24 antibody during the transition from G_0 to $G_{1/S}$ in the cell cycle. The Sym10 antibody revealed a protein of 25kDa whose methylation state apparently decreased from G_0 to $G_{1/S}$. The changes observed in HeLa cells were relatively modest (20-50%) compared to those seen in primary T-lymphocytes perhaps reflecting the high basal levels of arginine methylation in the former. A protein of 65kDa showed a large increase in methylation using the Asym24 antibody in T-lymphocytes, which could correspond to

the 64kDa protein identified by Kim et al. as Cleavage stimulation factor subunit 2 (CSTF2). CSTF2 is a nuclear protein required for polyadenylation of pre-mRNAs and is known to be up-regulated in the G_0 to $G_{1/S}$ transition (Martincic et al., 1998). Kim et al. identified the 25kDa Sym10 reactive protein from a 2D gel as being triosephosphate isomerase (TPI), however this seems unlikely as the isoelectric point of TPI is 5 and the region of gel from which the protein was extracted corresponds to ph8 (Kim et al., 2010a). The protein is more likely to have been snRNP-B or snRNP-B' which are known to be recognised by the Sym10 antibody and are basic (pI 11) (Boisvert et al., 2003).

While activation of T-lymphocytes leads to increases in arginine methylation and levels of the main arginine methyltransferase, PRMT1, it is not clear whether pre-existing proteins become more methylated or synthesis of methylated proteins increases. In reality both situations may occur and determination of whether the proportion of methylated proteins increases will require a proteomic approach. Similarly, the identity of the methylated proteins is not easily revealed by western blotting, thus identifying arginine methylated proteins in T-lymphocytes will require a mass spectrometry based proteomics approach.

4. Discovering arginine methylated proteins in T-lymphocytes

4.1 Introduction

4.1.1 Techniques for the discovery of arginine methylation sites

Determining the repertoire of arginine methylated proteins in T-lymphocytes is crucial to enable an understanding of the role this modification plays in T-lymphocyte function. Techniques used for detecting arginine methylated proteins can be divided into two categories, low throughput and high throughput methods.

Low throughput methods involve the individual analysis of proteins to detect whether they are methylated on arginine. Peptides or individually expressed proteins are incubated with arginine methyltransferases, the most common method for detection of transferred methyl groups is based on radioactivity. Tritiated *S*-adenosyl methionine (^{3}H-SAM) is used as the methyl group donor during the *in vitro* methylation, the substrate protein is then separated by SDS-PAGE and any tritiated methyl groups are detected by fluorography (Sakamaki et al., 2011). As a method for discovering arginine methylated proteins, this approach suffers from many drawbacks. It is time

consuming to analyse individual proteins or peptides, not least because detection of tritium, a low energy β-emitter can take days of exposure or even weeks (Galichet et al., 2008). Alternative methods of detecting a methylation reaction are possible, antibodies can be used although these tend to only recognise methylarginine within certain sequence contexts. The antibodies currently available are not very sensitive and generally only recognise methylarginine within arginine-glycine rich sequences (Boisvert et al., 2003; Ong et al., 2004; Duan et al., 2007). The substrate protein can also be hydrolysed to its constituent amino acids and these can then be analysed either by thin layer chromatography (TLC) if (^{3}H-SAM) was present, or by mass spectrometry to detect production of methylated arginine. An advantage of detecting methylarginine in hydrolysates is the ability to distinguish between MMA, ADMA and SDMA.

Cheng et al. developed a small pool screening approach for arginine methyltransferase target proteins. This increased the throughput of the radioactive detection method as it involved the screening of pools containing 10 cDNA clones of possible target proteins. Corresponding proteins were produced in an *in vitro* transcription/translation system (IVTT) and incorporation of tritiated methyl groups was assayed (Cheng et al., 2007). However a problem with these assays on individual proteins is that the detection method for methylation usually does not allow localisation of the modification. Site directed mutagenesis must be performed to discover the exact residue which becomes methylated. Furthermore, methylation is performed *in vitro* often using purified arginine methyltransferases, any methylated substrates identified may not be true targets in a normal cellular context.

Use of high throughput methods is necessary for large scale proteome wide detection of arginine methylation sites. One way of rapidly generating many putative arginine methylation sites is by using computational prediction. Daily et al. carried out one of the first attempts to build an algorithm that could predict arginine and lysine methylation sites (Daily et al., 2005). Amino acid composition and physico-chemical properties of sequences surrounding known methylation sites was used to build a predictor through support vector machines (SVM), a technique for classifying datasets, in this case into non-methylated and methylated categories. Accuracy of prediction was estimated to be 77% for arginine methylation, however only a small dataset (116 sites) was used for training which is unlikely to be a broad enough sample. Chen et al. used a larger training set (250 sites from UniProt and published articles) to develop a predictor of arginine methylation sites (called MeMo) with an accuracy of 86% (Chen et al., 2006). A subsequent prediction tool named MASA was created by Shien et al. which used 246 arginine methylation sites in the training set but also incorporated information on the solvent accessible surface area of the amino acids surrounding a methylation site, achieving an accuracy of 85% (Shien et al., 2009). More recently, several additional methods for predicting arginine methylation sites have been published using various methods to extract sequence features surrounding methylarginine sites (Xu et al., 2012; Shao et al., 2009; Shi et al., 2012). While computational techniques are very rapid and convenient, building an accurate predictor is currently limited by the small number of experimentally verified methylation sites which can be used as a training set and any putative methylation sites would need to be experimentally validated.

Mass spectrometry based proteomics has emerged over the last decade as a powerful approach enabling the high throughput study of the proteome. Because it is rapid, sensitive and robust, mass spectrometry has become the method of choice for proteomics with applications including quantitative analysis of the proteome, characterisation of protein complexes, detection of post-translational modifications, structural analysis of proteins and single cell analysis of markers (Bodenmiller et al., 2012; Zhou et al., 2011).

4.1.2 Mass spectrometry based proteomics

Prior to the advent of mass spectrometry based proteomics, one of the most widely used techniques for determining the sequence of a protein was Edman degradation. This is a chemical method which uses successive reactions of phenylisothiocyanate with the amino group of N-terminal amino acids. After each reaction, the derivatised amino acid is cleaved off and can be analysed by chromatography (Edman, 1950). Although automated Edman sequencers are available and sensitivity has been improved (Tsugita, 1987), this method is slower and less sensitive compared to the use of mass spectrometry.

Mass spectrometry is an established field with origins almost a century ago when J.J Thomson created the first mass spectrograph, and Aston and Dempster measured masses and abundances of ions (Aston, 1919; Dempster, 1918). A mass spectrometer measures the mass-to-charge (m/z) ratio of ions based on their behaviour in an electromagnetic field. However in order to enter the instrument and be analysed molecules must enter the gas phase

(become ionized).

4.1.3 Ionization

Many different ionization techniques exist, early ionization methods used electron impact and were relatively harsh meaning that larger thermally labile biomolecules such as proteins and peptides were destroyed. Therefore mass spectrometry based proteomics only became feasible with the development of softer ionization techniques in the 1980s. Fast atom bombardment was introduced in 1981 and provided a means of ionizing larger biomolecules with good sensitivity. Karas and Hillenkamp together with Tanaka then introduced Matrix Assisted Laser Desorption Ionization (MALDI), where the molecules of interest are mixed with a small organic molecule (the matrix). A laser is fired causing desorption of the matrix and analyte. It is thought that the small organic molecule absorbs energy from the laser and becomes protonated, this proton is then transferred to the analyte so that it becomes charged. MALDI enables ionisation of large molecules (even intact proteins up to 70kDa) in a very sensitive manner. In 1989 Fenn described another technique, electrospray ionization (ESI), which could be used to ionize large biomolecules. In ESI, the analyte is dissolved in solvent and applied through a capillary with a small diameter tip. A high voltage (4000V) is applied to the tip creating a potential difference between the capillary tip and the mass spectrometer inlet. Although the exact mechanism behind the production of ions by this method is unknown, it is thought that droplets are created

which then shrink due to evaporation. The charge from the analyte in each droplet remains the same whilst the volume decreases leading to coulombic explosion and generation of an ion spray.

Today, the main ionization techniques used in proteomics are MALDI and ESI. The latter is more suited to high throughput analysis as it can be directly interfaced with chromatographic separation meaning that mixtures of proteins or peptides can be fractionated and complexity reduced prior to detection by mass spectrometry.

Proteins are frequently proteolytically cleaved to their constituent peptides for analysis by mass spectrometry (bottom up proteomics). The reason for this is that peptides are generally easier to ionise and fragment than intact proteins, increasing the sensitivity of detection. In addition peptides are mostly easier to handle, for example they are less likely to precipitate in the organic solvents and acids used in chromatography. Proteases with differing residue specificity may be used for the digestion of proteins. A common choice is trypsin as it is a robust, reliable protease cleaving C-terminal to lysine and arginine residues, generating peptides with at least two positive charges (one at the N-terminus and one at the C-terminal lysine/arginine) which can be readily ionized (Olsen et al., 2004).

If a very complex mixture of peptides is ionized and analysed by mass spectrometry without pre-fractionation, sensitivity and coverage will be limited. This is due to the fact that ion suppression can occur in ESI and also analysers in mass spectrometers have a limited dynamic range. Ion suppression occurs when a compound reduces the ionization of another compound when the two are present together during droplet formation (Annesley, 2003).

The mechanism behind ion suppression is not clear, but it is thought that molecules compete for limited space and charge at the surface of droplets. More abundant and also less polar molecules will out-compete less abundant polar molecules leading to ionization of the former to the detriment of the latter (Bruins et al., 1999). Fractionation of the sample will generate simpler mixtures at the point of droplet formation reducing the severity of ion suppression.

4.1.4 Mass analysers

The dynamic range of mass analysers in mass spectrometers is also a limiting factor for complex samples. The mass analyser is responsible for separating ions according to their m/z so that each type of ion can be detected. Various different mass analysers exist each with their own properties and applications, these include the time-of-flight (TOF), quadrupole, quadrupole ion trap, linear quadrupole ion trap, fourier transform ion cyclotron (FT-ICR) and Orbitrap analysers (Figure 4.1). The TOF analyser measures how long an ion takes to reach a detector which is used to calculate the m/z. The quadrupole analyser is made up of four parallel hyperbolic rods, a radio frequency voltage generates an alternating electric field that can be tuned to allow ions with certain m/z values to pass through. Based on similar principles, the quadrupole ion trap restricts ions within a three dimensional trap causing them to orbit. By altering the electric field ions with a particular m/z can be ejected and detected. The linear quadrupole ion trap generates

a two dimensional quadrupole field to trap ions and has higher ion storage capacity and rapid cycle times (time taken from injecting ions into the trap to detection). The TOF and quadrupole mass analysers are said to separate ions "in space" because ions are separated according to their m/z as soon as they enter the analyser. Ion trap analysers such as the linear quadrupole ion trap by contrast are said to separate ions "in time" because all ions that enter are held to begin with and those with a certain m/z are subsequently ejected by applying the appropriate radio frequency voltage (Michalski et al., 2011). FT-ICR and Orbitrap mass analysers belong to neither of the aforementioned categories as although ions are trapped and stored they measure all ions simultaneously for a period of time. Another difference is that in FT-ICR and Orbitrap analysers ions are detected without being consumed in a collision with a detector.

In an FT-ICR instrument ions rotate in a circular motion in a strong magnetic field. A radio frequency sweep is carried out, any ion with a cyclotron frequency matching a particular radio frequency is excited and follows a larger orbit radius which induces an alternating image current in the detection electrodes that is measured (Figure 4.1). All ions are measured at the same time requiring fourier transform to de-convolute the spectrum into individual frequency measurements which are converted to the m/z of the ions. FT-ICR instruments are exquisitely sensitive and have high resolution because frequencies can be measured very accurately but require costly superconducting magnets.

An Orbitrap analyser consists of a central electrode around which ions orbit in an electrostatic field. Similar to FT-ICR, the oscillations of all ions

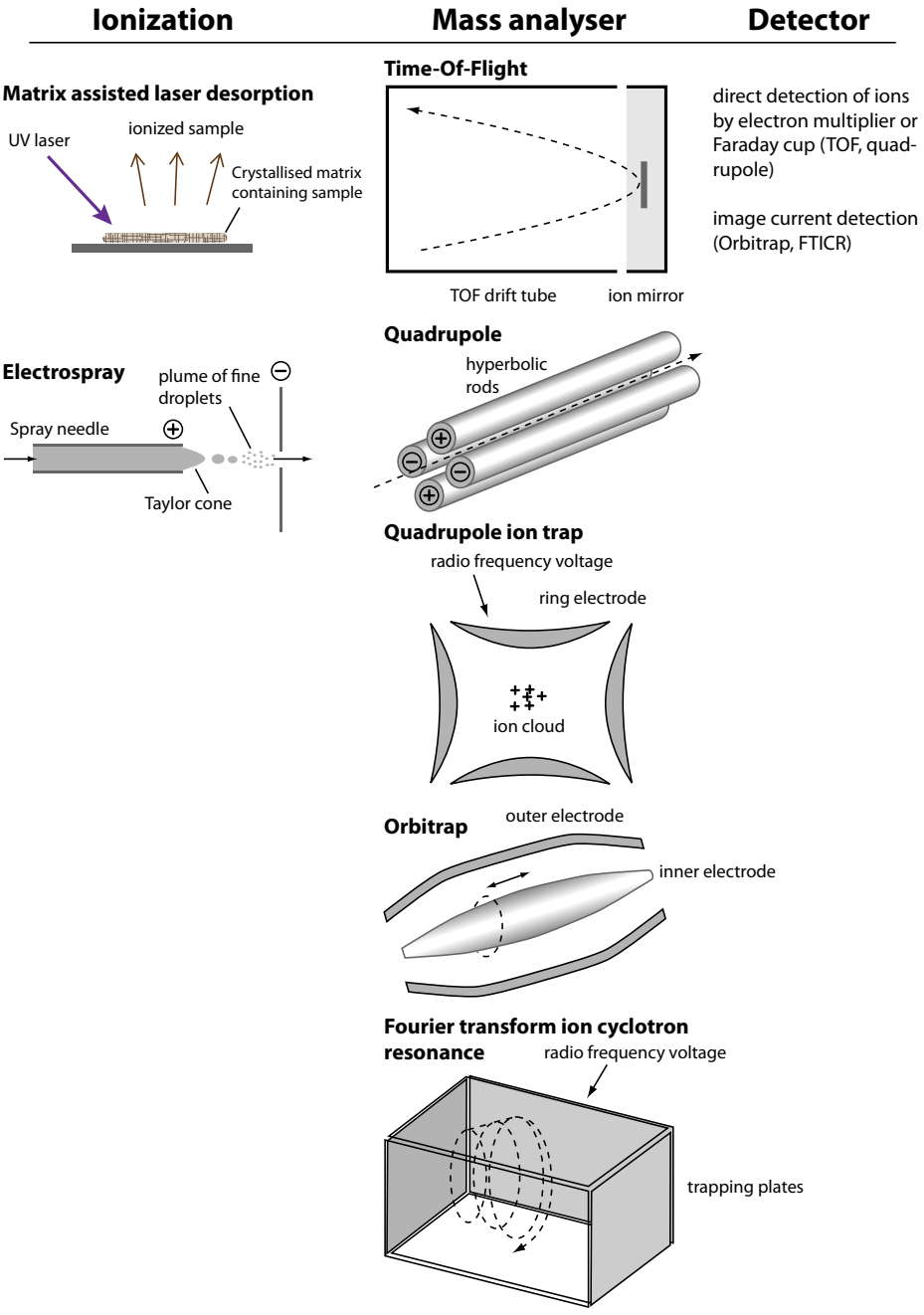


Figure 4.1: Modules used for mass spectrometry based proteomics. Peptides can be ionized by matrix assisted laser desorption ionization (MALDI) of electrospray ionization (ESI). Mass of peptides or peptide fragments can be measured using many different types of mass analyser, the most common types are shown here. Typically, analysers are assembled in a hybrid configuration (e.g. quadrupole-Orbitrap). Ions are either directly detected after collision with a detector or induce an image current.

around the electrode are detected as simultaneous frequency measurements which must be de-convoluted by fourier transform to extract the individual axial frequencies of the ions. Oscillation frequency is related to the m/z of the ion. Again because frequencies can be measured accurately and precisely, Orbitrap analysers have high resolution, mass accuracy and they also have a high dynamic range (10^5). Because they are less costly to produce and maintain compared to FT-ICR mass spectrometers, Orbitrap analysers have become very popular.

On their own each of these analysers would have limited usefulness in mass spectrometry based proteomics since simply measuring the mass of an intact peptide does not usually provide enough information for its unambiguous identification (an approach called peptide mass fingerprinting), many peptides of differing composition may match to a particular m/z value. Peptides (precursor ions) are therefore fragmented into smaller product ions which provide information on the composition of the peptide. Thus, a typical mass spectrometer used for proteomics consists of several mass analysers, sometimes of different types (hybrid mass spectrometer), linked together. One analyser allows selection of a particular peptide which is then fragmented and the product ions are measured by another mass analyser. For instance, an LTQ-Orbitrap mass spectrometer is composed of a linear trap quadrupole (LTQ) mass analyser and an Orbitrap mass analyser. The LTQ mass analyser is rapid, sensitive and has high capacity for storing ions but has a lower resolution. The Orbitrap has a high resolution but is less sensitive so takes more time to analyse ions. Speed is an important factor in mass spectrometry as peptides are continually eluting from the column used for fractionation,

so rapid analysis allows more peptides to be fragmented and identified. The LTQ-Orbitrap uses the strengths of both types of analyser, peptides are measured in the Orbitrap in a quick preview scan and this information is then used by the LTQ to select peptides to accumulate and fragment, product ions are then rapidly analysed also by the LTQ (to generate an MS/MS fragmentation spectrum) while the Orbitrap completes a longer high mass accuracy analysis of all the precursor ions (generating an MS spectrum) (Jedrychowski et al., 2011).

4.1.5 Fragmentation of peptides

Inside a mass spectrometer, peptides can be fragmented by various methods including collision induced dissociation (CID), higher energy collisional dissociation (HCD) and electron transfer dissociation (ETD). CID involves accelerating peptides to a higher kinetic energy and collision with an inert gas such as helium or nitrogen. Kinetic energy is thereby converted to internal vibrational energy leading to bond cleavage (Jones and Cooper, 2011). A disadvantage of CID is that fragmentation often follows the pathway of lowest energy and if a labile group (such as phosphate) is present this will be preferentially lost before any fragmentation, leading to uninformative product ion spectra. In CID peptides are fragmented after multiple low energy collisions with gas molecules. HCD by contrast, is a high energy collision dissociation method with fragmentation taking place after fewer higher energy collisions. This allows peptides to follow more fragmentation pathways

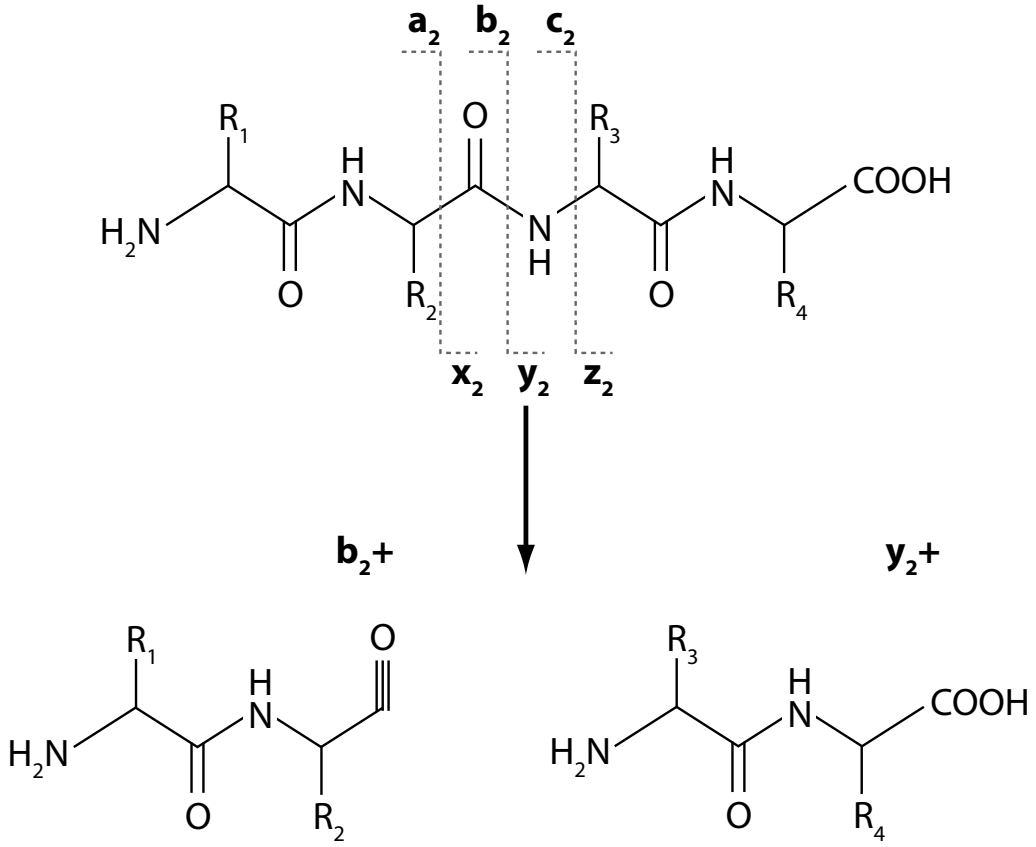


Figure 4.2: Fragmentation of peptides according to Roepstorff and Fohlman nomenclature (Roepstorff and Fohlman, 1984). Fragments labelled a, b, c contain the N-terminal portion of the peptide. N-terminal fragments are numbered starting from the N-terminus. Fragments labelled x, y, z contain the C-terminal portion and are numbered starting from the C-terminus. Peptides can fragment via different pathways depending on the dissociation method used. Fragmentation pathway occurring in HCD and CID is shown here. The b and y ions may fragment further (e.g. via loss of side chain) to form other products.

rather than just the low energy one, leading to information rich MS/MS spectra (Olsen et al., 2007; McAlister et al., 2011). ETD fragmentation involves collision of positively charged peptides with negatively charged anions, an electron is transferred to the peptide which therefore become unstable and fragments (Syka et al., 2004). ETD preserves labile peptide modifications

making it suitable for the study of certain post-translational modifications. However in its current form it is not very sensitive (only a portion of the peptides are fragmented) and reaction times are slower than in HCD and CID.

Various bonds can be broken within peptides during fragmentation, if an amide bond breaks, a b-ion (containing the N-terminal portion of the peptide) and a y-ion (containing the C-terminal portion of the peptide) are produced (Figure 4.2) (Roepstorff and Fohlman, 1984). This fragmentation pathway is observed during CID and HCD and is thought to be charge directed, in other words a proton needs to be present at the amide bond for cleavage to take place (mobile proton model) (Paizs and Suhai, 2005). This becomes an issue for larger peptides with multiple basic residues as they sequester the proton and inhibit fragmentation (Wysocki et al., 2000). ETD on the other hand leads to breakage of peptides at the N-C_α bond, producing a c-ion (containing the N-terminal portion of the peptide) and a z-ion (containing the C-terminal portion of the peptide). It can be used to fragment very large peptides and even whole proteins in top down proteomics.

4.1.6 Interpretation of mass spectra

In a typical proteomics experiment using mass spectrometry, many thousands of mass spectra are generated. The challenge is to assign each one to a peptide sequence. Manual interpretation of all MS/MS spectra to derive the peptide sequences is not feasible, therefore computational approaches have

been developed to carry out this task. The interpretation of MS/MS spectra is not straightforward as there is no way to distinguish between b- and y-ions (or c- and z-ions) and fragmentation is often incomplete or multiple redundant ion series are observed. A solution to this problem was to use sequenced genomes to generate a list of all protein sequences. The protein sequences are then digested *in silico* according to the appropriate enzyme specificity and the fragment ions theoretically observable for each peptide are calculated. Fragmentation spectra are then searched against these theoretical spectra to find the best match (Peptide Spectrum Match - PSM) and therefore the identity of the peptide. Only the pattern of product ions is considered during the matching process and not the relative intensities (Figure 4.3). The degree of similarity between the acquired spectrum and the theoretical spectrum is represented as a probability. This database matching approach is rapid and accurate and is the most popular method for identifying peptides from fragmentation spectra (Steen and Mann, 2004). When the genome of an organism has not been sequenced, fragmentation spectra can be interpreted in a *de novo* manner without relying on a sequence database. Several algorithms exist for performing this task (Horn et al., 2000; Zhong and Li, 2005; Searle et al., 2005). *De novo* interpretation of fragmentation spectra can be assisted by altering experimental methods. If Lys-N is used to digest proteins in place of trypsin, peptides will be produced with N-terminal lysines, when these peptides are fragmented by ETD the predominant product ions are c-ions making identification of the peptide sequence simpler (Taouatas et al., 2008).

Fragmentation spectra are often complex, containing numerous product

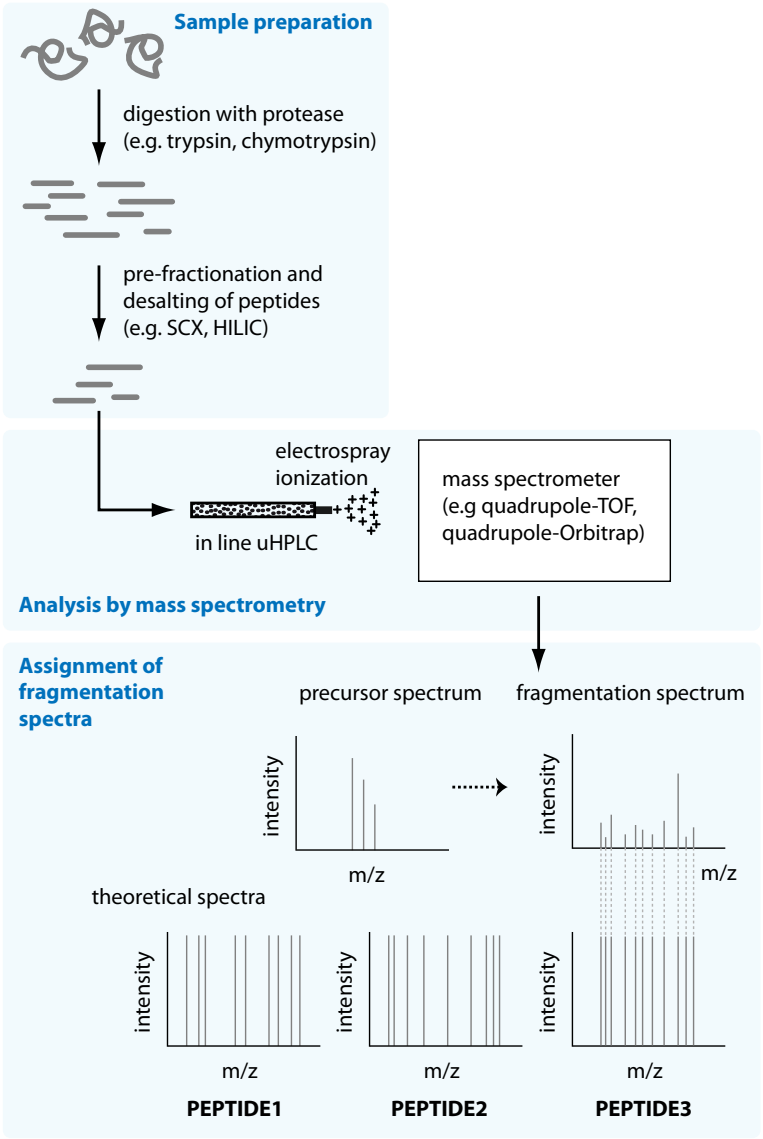


Figure 4.3: A typical mass spectrometry based proteomics workflow. During the sample preparation phase, proteins are extracted from cells and digested using a protease. Peptides are fractionated, possibly to enrich for those carrying a certain post-translational modification. For analysis by mass spectrometry, peptides are desalted, passed through a small reverse phase column and ionized. Ions enter the mass spectrometer and masses of intact peptides are first measured (precursor spectrum). Selected peptides are then fragmented and masses of fragmentation products are measured (fragmentation spectrum). In the data analysis phase (after acquisition of spectra), fragmentation spectra are matched to theoretical spectra to determine peptide sequences (and possibly modifications).

ion series. This raises the possibility that a spectrum may be incorrectly matched to a theoretical spectrum generating a false peptide match. One way of minimising the number of false matches is to only accept those with a high probability score. The problem with this approach is that every database search algorithm has a different method of scoring peptide matches making it difficult to implement a consistent confidence threshold across the proteomics community. A solution to this problem is to generate a false set of peptide sequences and corresponding theoretical spectra of equal size to the correct set (Moore et al., 2002). A simple way of generating a false set is to simply take the correct sequence database and reverse it to generate a decoy database. Any fragmentation spectrum which is matched to this decoy database is clearly a false positive identification. The decoy database is usually concatenated with the true database and fragmentation spectra are searched against both at the same time. To estimate the False Discovery Rate (FDR) the number of matches to the decoy database is doubled and divided by the total number of matches. The reason for doubling the false matches is that an equal number of hidden incorrect matches to the real database will also occur (Elias and Gygi, 2007). This is currently the most popular strategy for estimating the FDR and acceptable FDR values are typically 1-5%.

A problem with the current method of generating theoretical spectra is that it relies on prior knowledge of how peptides fragment and intensities of product ions are not taken into account. In addition, while the decoy database strategy can estimate an overall FDR value for a large group of fragmentation spectra, it does not give a false positive rate for each individ-

ual spectrum. Kim et al. have developed a novel database search algorithm which provides rigorous p-values for matches, assumes no prior knowledge of fragmentation and takes intensities into account. This uses the generating function approach and is able to calculate statistical significance (false positive rate) for each individual spectrum (Kim et al., 2008, 2010b).

4.1.7 Detecting post-translational modifications by mass spectrometry

Since post-translational modifications (PTMs) affect the behaviour of many proteins, their identification by mass spectrometry is a valuable endeavour. However, detecting modified peptides by mass spectrometry is not straightforward. A major problem is the limited dynamic range of mass spectrometers and sample complexity. Generally, modified peptides are present in substoichiometric amounts, unmodified peptides therefore will be ionised and detected at the expense of modified forms . Even if modified peptides are present in detectable amounts, the complexity of the sample may mean that the current generation of mass spectrometers are simply not rapid enough to obtain a good coverage of the peptides and they will be missed. Thus, modified peptides must first be enriched to reduce sample complexity. Fragmentation of a peptide containing a PTM can also be problematic. Some PTMs are labile meaning that they are preferentially lost before the peptide can fragment properly. Finally, searching fragmentation spectra for modified peptides leads to increased false positive matches due to the increased num-

ber of theoretical spectra to search against (increased search space) (Mann and Jensen, 2003).

Extensive efforts have been made to map PTMs such as phosphorylation, acetylation and ubiquitination, with thousands of identified sites (Henriksen et al., 2012; Wagner et al., 2011; Phanstiel et al., 2011). Reagents and techniques are available to selectively enrich such modified peptides so that they are observable by mass spectrometry. Phosphorylated peptides can be enriched with titanium oxide (TiO_2) while antibodies can be used to enrich for acetylated or ubiquitinated peptides.

4.1.8 Identification of arginine methylation sites by mass spectrometry

Despite being discovered more that 40 years ago, there has been no published large scale study to identify arginine methylation sites, consequently the number of known sites is rather low compared to more intensively studied modifications. The reason for this is that there has been a lack of available reagents and techniques for detecting arginine methylation by mass spectrometry.

Detecting arginine methylation sites by mass spectrometry represents a distinct challenge. Firstly, 1% of arginine residues exist in the methylated form (this work), thus enrichment is required for detection as with many other PTMs because many arginine methylation sites are likely to be substoichiometric (Ostareck-Lederer et al., 2006). Secondly, trypsin cleaves

less efficiently C-terminal to a methylated arginine (or lysine) (Baldwin and Carnegie, 1971) and arginine methylation tends to occur in arginine-glycine rich regions (Najbauer et al., 1967). Thus, tryptic arginine methylated peptides are longer and typically have multiple charged residues. This reduces fragmentation efficiency as the charged residues sequester the mobile proton which directs bond breakage (Syka et al., 2004; Snijders et al., 2010). Thirdly, due to the trypsin mis-cleavages after methylation, the number of theoretical spectra that need to be generated increases because every peptide containing an arginine may now occur in a mis-cleaved form. Every peptide containing an arginine must also be considered as potentially occurring in a monomethylated, dimethylated form or a combination. Finally, several amino acid substitutions can have the same mass difference as a methylation, for example glutamate for aspartate or alanine for glycine. All these factors lead to a large combinatorial increase in the search space, increasing the rate of spectrum matches by chance and giving a higher FDR for detection of arginine methylation.

The first attempt at a proteomic scale discovery of arginine methylation used four different antibodies against methylarginine to immunoprecipitate proteins prior to detection by mass spectrometry (Boisvert et al., 2003). If identified proteins contained glycine-arginine rich regions and were not detected in the control, this was taken as evidence of arginine methylation. Thus, while no actual methylation sites were detected, over 200 proteins were deemed as being highly likely to be methylated. One such protein, Mre11, was later confirmed as being methylated on arginine (Boisvert et al., 2005). This approach was useful as a preliminary screen, however the antibodies

used were all raised against methylated arginine-glycine rich peptides making the study biased towards this particular motif.

Ong et al. introduced heavy methyl SILAC as a method for increasing the confidence in identification of methylation sites by mass spectrometry (Ong et al., 2004) (Figure 4.4). In this technique, cells are cultured in the presence of either L-Methionine-(methyl-$^{12}CH_3$)(light) or L-Methionine-(methyl-$^{13}CD_3$)(heavy). Methionine is metabolically converted to the methyl donor *S*-Adenosyl Methionine (SAM), thus all methylation reactions in light labelled cells will use a light methyl group and those in heavy labelled cells will use a heavy methyl group. When proteins from both types of cell are mixed together in a 1:1 ratio and digested, all methylated peptides will occur in 1:1 pairs separated by the mass difference between light and heavy methyl groups. This 1:1 methyl SILAC pair can be used to corroborate any putative methylation sites identified by database search algorithms as well as providing relative quantitation between two samples. Example spectra from a heavy methyl SILAC labelled peptide are shown in Figure 4.5.

To demonstrate the utility of their method, Ong et al. used an antibody against methylarginine (7E6) and methyllysine to enrich for methylated proteins, leading to identification of 59 unique methylation sites in 33 proteins, of which two sites in two proteins were methylated lysine. These results significantly increased the number of known arginine methylation sites, however the 7E6 antibody is biased towards arginine-glycine rich methylation sites and this was not a large scale attempt at mapping methylarginine sites. The only other published attempt to identify arginine methylation sites on a large scale was by Pang et al. who searched previously generated pep-

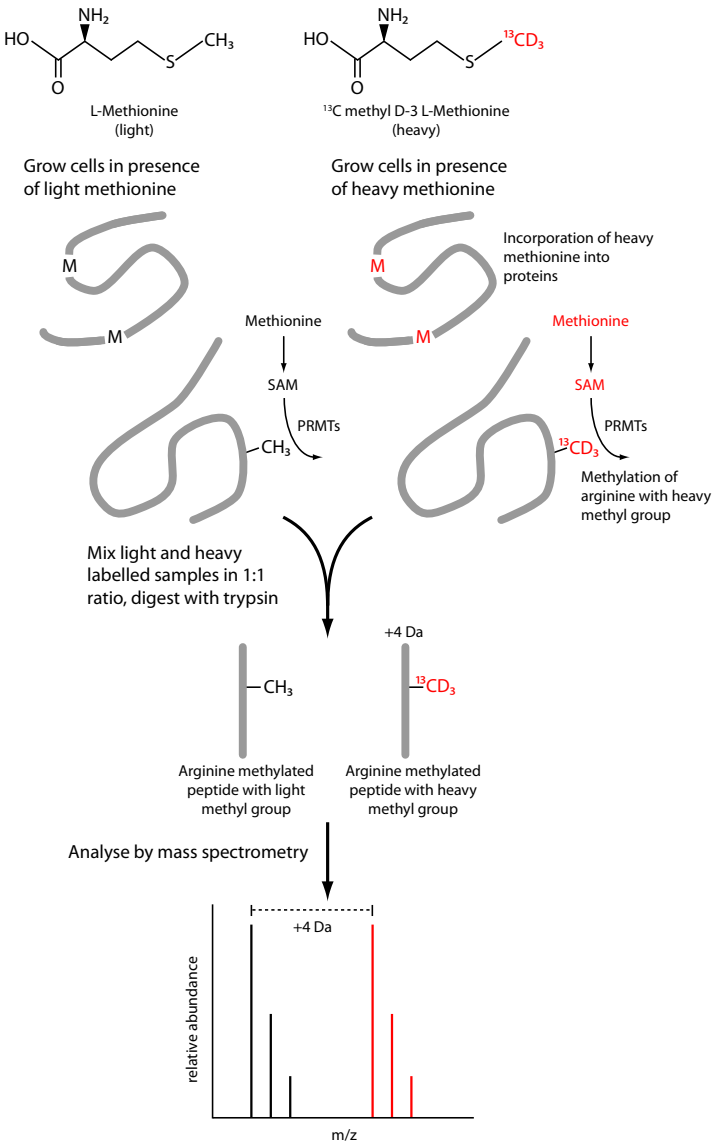


Figure 4.4: Heavy methyl SILAC. Cells are cultured with unlabelled methionine (light) or stable isotope labelled methionine (heavy). Methionine is incorporated into proteins but also metabolically converted to *S*-adenosyl methionine. Either a light or a heavy methyl group will be transferred during methylation reactions. Proteins are mixed 1:1 and digested. Every methylated peptide will occur in a 1:1 pair. These methyl SILAC pairs provide corroboration of any methylation site identified during database searches.

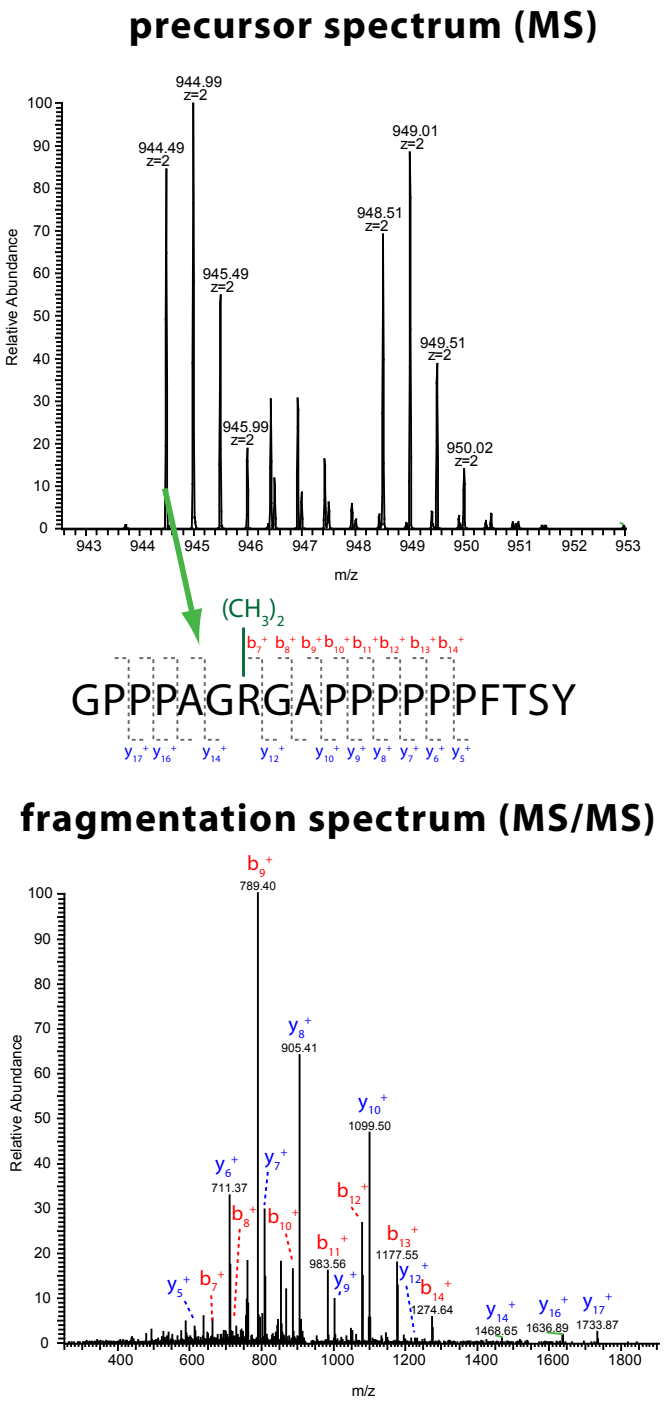


Figure 4.5: Example precursor (upper panel) and fragmentation (lower panel) spectra of an arginine methylated heavy methyl SILAC labelled chymotryptic peptide from DAZ-associated protein 1. The precursor spectrum shows the methyl SILAC pair. The light peptide was selected for fragmentation by CID. All b and y ions observed are labelled.

tide MALDI-TOF data from *S. cerevisiae* for evidence of lysine or arginine methylated peptides (Pang et al., 2010). Peptide spectra from purified digested proteins were queried with masses of theoretically observable peptides from those proteins. Where a mass difference was observed which could correspond to a methylation, this was taken as a putative methylation site. Potential hits were then filtered and methylation sites were required to be present in at least two different peptides to increase confidence. This approach led to identification of 38 arginine methylation sites in 31 proteins, however the false positive rate was estimated to be 11% and only a portion (40%) of the yeast proteome was studied. Because this relied on peptide mass fingerprinting, the sequence of peptides was inferred from their intact mass alone meaning that this study can only be seen as a preliminary screen.

4.1.9 Using mass spectrometry to identify arginine methylated proteins in T-lymphocytes

In order to study arginine methylation in T-lymphocytes by mass spectrometry, cells should be heavy methyl SILAC labelled to ensure high confidence identification of sites. This is straightforward for T-lymphocyte cell lines such as Jurkat but is not routine for primary T-lymphocytes. Primary cells cannot normally be efficiently labelled by SILAC as they do not divide and cannot be cultured for long enough to get sufficient incorporation of label. However, T-lymphocytes isolated from human blood can proliferate and be cultured for several weeks if they are stimulated in an appropriate manner. This al-

lows heavy methyl SILAC labelling and mass spectrometric identification of methylation sites in primary cells. In fact primary human T-lymphocytes have been successfully labelled by SILAC during differentiation into T_H1 or T_H2 (Loyet et al., 2005). However, all proteomic studies to date have used cell lines to identify arginine methylation sites and it is natural to ask whether the methylated proteome reflects the true situation in non-transformed cells.

Arginine methylated proteins or peptides from T-lymphocytes need to be enriched to be detected. Compared to the phosphate group in phosphorylated peptides which selectively binds to TiO_2 (Pinkse et al., 2004), the methyl group is a rather subtle modification which does not have distinct enough properties to allow its selective enrichment. Several antibodies against methylarginine exist, but they are biased towards arginine-glycine rich motifs and lead to identification of only a limited number of methylation sites. While the methyl group may not have distinct properties, the peptides containing methylarginine might. Due to mis-cleavage of methylarginine residues by trypsin, arginine methylated peptides will contain at least two basic residues. In addition, arginine methylation occurs frequently in arginine-glycine rich regions. Consequently, arginine methylated peptides should be highly positively charged, basic and hydrophilic. Such distinct properties may provide a means to enrich for these peptides, enabling identification of arginine methylated proteins in T-lymphocytes.

4.1.10 Aims

Only 3 proteins have been shown to be arginine methylated in T-lymphocytes, TARPP, NIP45 and Vav1. No mass spectrometry based screen for arginine methylated proteins in T-lymphocytes or primary cells has been reported, despite evidence that arginine methylation may play an important role in T-lymphocyte function (Chapter 3). The aim of the work in this chapter was to identify as many arginine methylated proteins as possible from T-lymphocytes in order to gain an understanding of the role this modification plays in T-lymphocyte activation. To accomplish this, primary T-lymphocytes had to be efficiently labelled by methyl SILAC and novel methods developed for enriching arginine methylated peptides prior to detection by mass spectrometry.

4.2 Results

4.2.1 Applying SILAC to primary T-lymphocytes

Labelling cells by heavy methyl SILAC is a prerequisite for high confidence identification of methylation sites by mass spectrometry. T-lymphocytes isolated from human blood proliferate in response to stimulation and can be cultured for several weeks. Under these conditions it is possible to efficiently label primary cells with isotope labelled amino acids. To test the efficiency of labelling primary T-lymphocytes by SILAC, CD4+ T-lymphocytes were isolated from human blood by separation on a Ficoll cushion followed by negative selection. The cells were then stimulated in CD3 antibody-coated flasks with the addition of CD28 antibody and IL-2. The RPMI 1640 medium contained L-Arginine-($^{13}C_6$, $^{15}N_4$) (Arg10) which is 10.0083 Da heavier than the unlabelled arginine it replaces. T-lymphocytes were cultured in this manner for 7 days before being transferred to a flask with no stimulating antibodies where they were rested for 3 days. Cells were then re-stimulated for a further 7 days to maximise label incorporation and harvested. Culturing the cells under these conditions led to a high labelling efficiency of 98.7%, with a mean heavy/light ratio for Arg10 containing peptides of 80.2 (Figure 4.6). As a control, data from an unlabelled sample was searched for Arg10 containing peptides, giving a mean heavy/light ratio of 2.23.

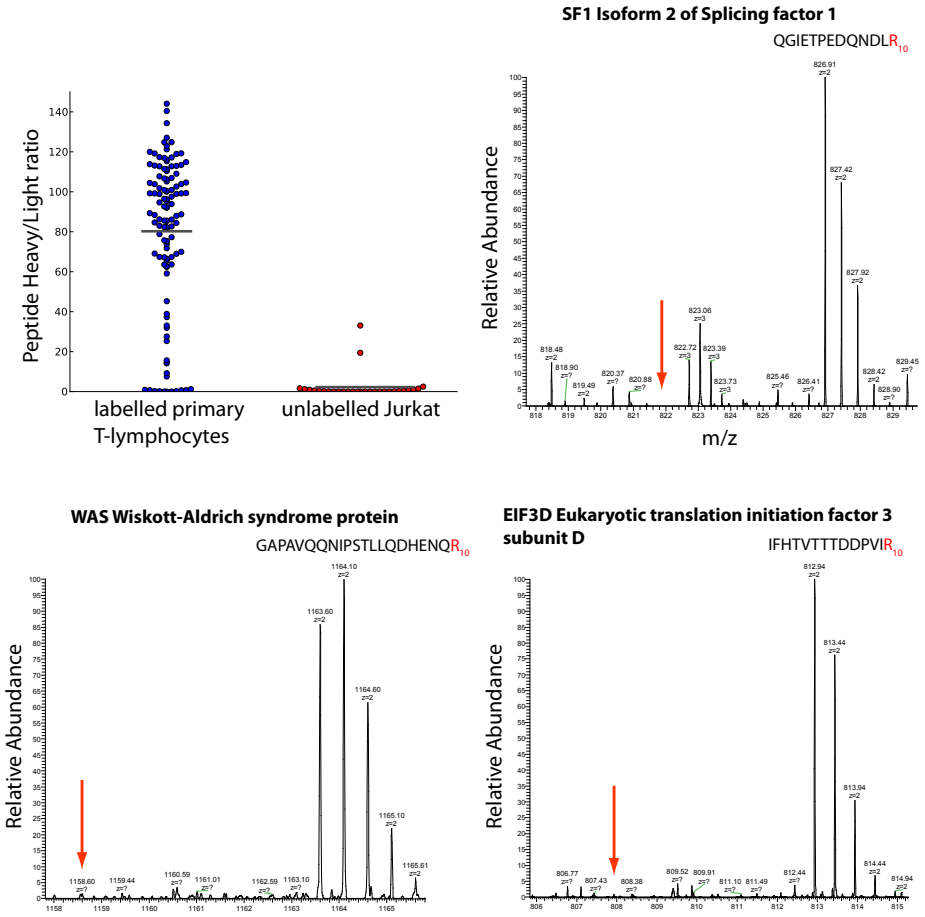


Figure 4.6: Primary human CD4+ lymphocytes can be efficiently labelled by SILAC. Human CD4+ T-lymphocytes were stimulated and cultured for 17 days in the presence of L-Arginine-($^{13}C_6$, $^{15}N_4$). The heavy/light ratios of 117 Arg10 containing peptides identified and quantified are plotted (Top left). For comparison, mass spectrometry data from an unlabelled Jurkat T-lymphocyte sample was searched for Arg10 containing peptides and the corresponding heavy/light ratios of these 29 peptides are plotted. Mean heavy/light ratios for each dataset are shown as grey lines. Example spectra are shown for identified Arg10 containing peptides, the red arrow indicates where the unlabelled version would occur in the event of incomplete labelling.

4.2.2 Using an anti-methylarginine antibody (7E6) to enrich for arginine methylated peptides

Since it has been shown that primary human CD4+ T-lymphocytes can be efficiently labelled, primary T-lymphocyte were labelled by heavy methyl SILAC. Cells were lysed and arginine methylated proteins immunoprecipitated with the 7E6 anti-methylarginine antibody, the only available antibody demonstrated to enrich for arginine methylated proteins prior to mass spectrometry (Ong et al., 2004). To verify the success of the immunoprecipitation,

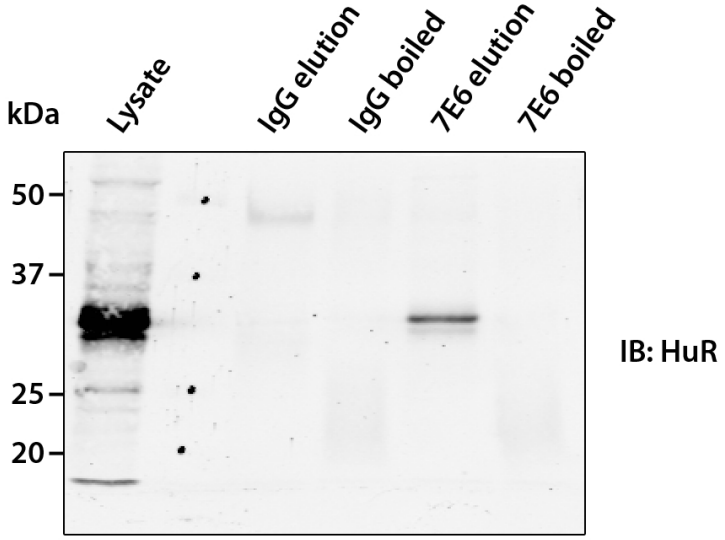


Figure 4.7: Immunoprecipitation of arginine methylated proteins prior to mass spectrometry. Primary human CD4+ T-lymphocytes were labelled by heavy methyl SILAC, lysed and proteins immunoprecipitated using the anti-methylarginine antibody 7E6 covalently linked to beads. An isotype control immunoprecipitation was also carried out. Proteins were eluted first by alkaline pH followed by boiling. Immunoblotting for the RNA binding protein HuR acts as a positive control as the methylated form of this protein is known to be specifically recognised by the 7E6 antibody.

an immunoblot for HuR was carried out as the 7E6 antibody is known to

Table 4.1: Arginine methylation sites identified in primary human CD4+ T-lymphocytes using 7E6 anti-methylarginine antibody.

Accession number	Protein name, *gene name*	Peptide	Methylation site	Methylation	InterProphet probability
P17844	Probable ATP-dependent RNA helicase DDX5, *DDX5*	**R**GGFNTFR	R502	DMA	0.984
P09651	hnRNP A1, *HNRNPA1* (heterogeneous nuclear ribonucleoprotein A1)	QEMASASSSQ**R**GR	R291	DMA	0.426
Q86V81	Tho4, *THOC4* (THO complex subunit 4)	RPAQSVN**R**GGMTR	R197	DMA	0.107
		N**R**GAGGFGGGGGTR	R204	DMA	0.498

specifically recognise the arginine methylated form of this protein (Li et al., 2002). HuR was eluted from 7E6 conjugated beads by alkaline pH, but was not seen on isotype control beads (Figure 4.7). Using the 7E6 antibody and heavy methyl SILAC labelling only 4 arginine methylation sites in 3 proteins were found in primary T-lymphocytes, none of these were novel (Table 4.1).

4.2.3 Alternative strategies for enriching arginine methylation sites prior to mass spectrometry

Use of the 7E6 antibody only resulted in identification of a small number of arginine methylation sites. Existing antibodies against methylarginine are biased towards arginine-glycine rich sequence contexts, therefore the development of antibody independent methods for enriching arginine methylation sites could lead to novel methylation sites and discovery of new sequence contexts. Trypsin cleaves less efficiently C-terminal to a methylated arginine or lysine (Baldwin and Carnegie, 1971), tryptic methylated peptides will therefore contain at least two basic residues and will have a charge of $\geq+3$ at

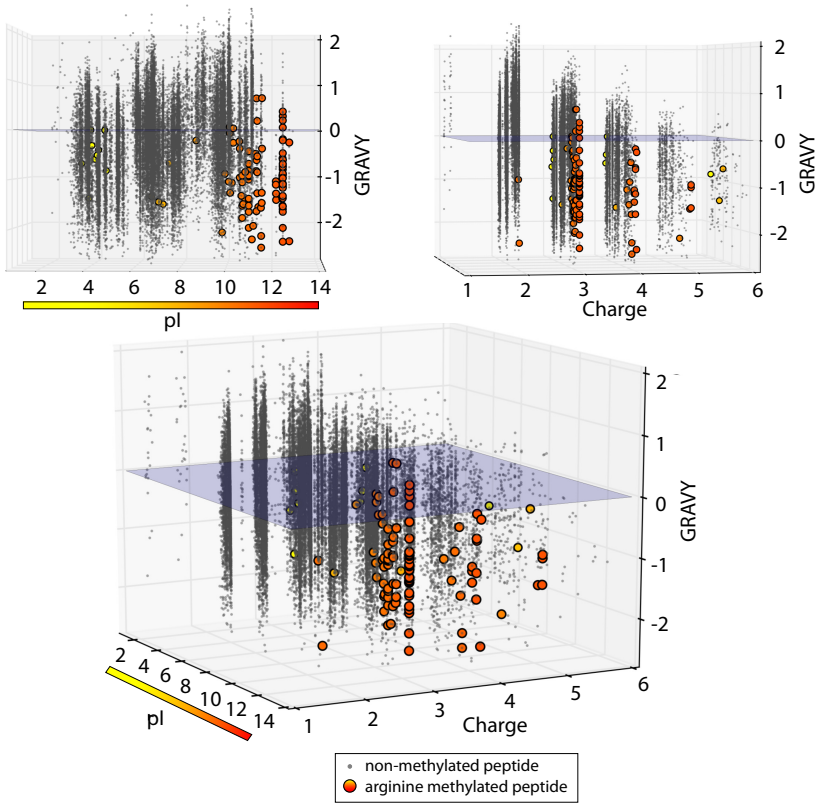


Figure 4.8: Physico-chemical properties of known tryptic arginine methylated peptides. Charge at pH2.7, pI and GRAVY scores of known arginine methylated peptides (coloured points) and non-methylated peptides (grey points) are plotted.

acidic pH. Arginine methylation also commonly occurs in glycine-arginine rich regions (Najbauer et al., 1967), thus arginine methylated tryptic peptides might also be expected to be basic (have a high isoelectric point) and hydrophilic. To verify this, all experimentally verified arginine methylation sites in the UniProtKB database (www.uniprot.org) were extracted and corresponding proteins were digested *in silico* according to tryptic specificity. For each peptide charge at pH 2.7, isoelectric point (pI) and grand average of hydropathy (GRAVY) score (Kyte et al., 1982) was plotted (Figure 4.8 coloured points). These properties were also plotted for 33,000 tryptic peptides identified from 12 fractions separated using a pH 3-11 immobilised pH gradient (IPG) (Figure 4.8 grey points). Tryptic arginine methylated proteins are indeed observed to be highly charged (98.3% have a charge of $\geq$+3 at pH2.7), basic (mean pI 10.6) and hydrophilic (mean GRAVY score -0.99) compared to non-methylated peptides.

4.2.4 Enrichment by strong cation exchange chromatography

The distinct physico-chemical properties of arginine methylated peptides should enable their enrichment prior to detection my mass spectrometry. Strong cation exchange chromatography (SCX) is an established method to separate differentially charged peptides. Peptides are bound to a negatively charged column, usually at a low pH, and a salt gradient is then used for elution. Peptides with lower charge states will bind more weakly and elute

earlier in the salt gradient, while more highly positively charged peptides

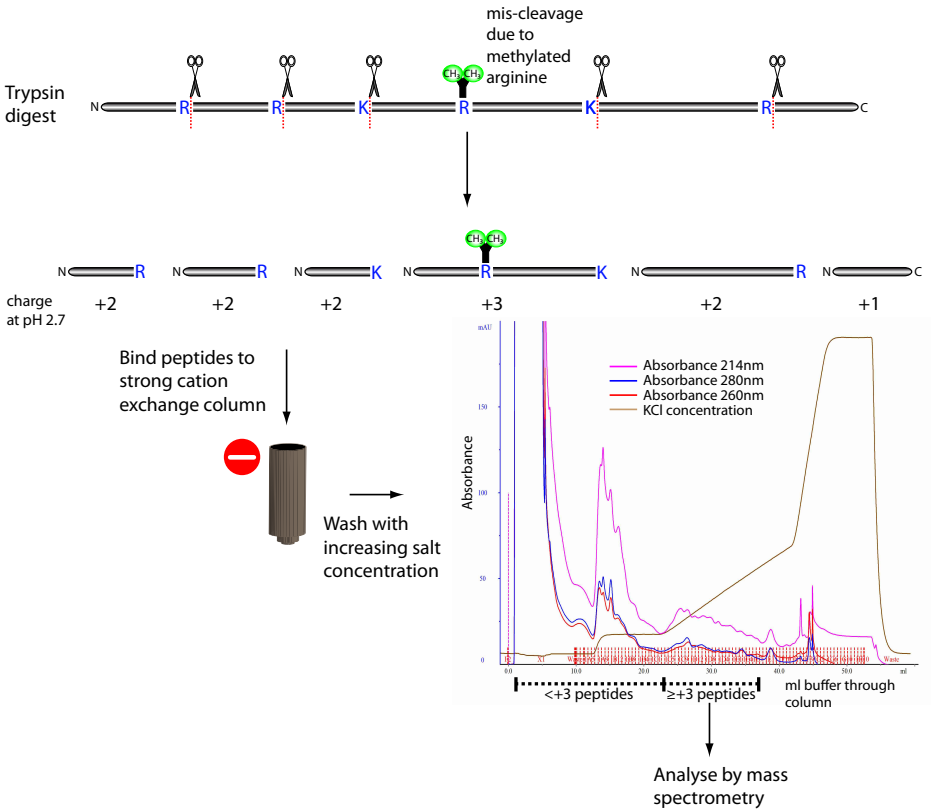


Figure 4.9: Strong Cation Exchange chromatography as a method to enrich for methylated peptides. Proteins were digested with trypsin and peptides were applied to a negatively charged column at pH2.7. An initial step of KCl eluted most of the +1 and +2 peptides and a further increase in KCl concentration eluted the higher charged peptides including arginine methylated peptides.

will only elute in the presence of higher salt concentrations. In a tryptic digest of proteins from whole cell lysate, approximately 70% of peptides will have a charge of +2 at pH2.7 (Beausoleil et al., 2004). Due to mis-cleavage, methylated peptides will have higher charges and can therefore be separated from the majority of other peptides (Figure 4.9).

Enrichment by SCX chromatography led to the identification of 39 argi-

nine methylation sites in heavy methyl SILAC labelled primary T-lymphocytes and Jurkat cells (work by Thomas Uhlmann). While this is less than the number found using the 7E6 antibody (Ong et al., 2004), 25 of these sites were novel (Table 4.5). In addition one lysine methylation site was found.

4.2.5 Enrichment by isoelectric focusing

Because arginine methylated peptides have a high pI, fractionation according to pI should result in enrichment of methylarginine containing peptides in the basic fractions. A popular method of pI based fractionation of peptides prior to mass spectrometric analysis is OFFGEL isoelectric focusing (IEF) (Michel et al., 2003). Here, an IPG strip lies beneath wells containing solution phase peptides to be focused. A voltage is applied and peptides enter the gel, migrate to their isoelectric point and equilibrate with the solution, making recovery easy. The most basic IPG strips commercially available contain a pH range of 7-11, these were therefore used to separate peptides from a tryptic digest of methyl SILAC labelled primary T-lymphocytes or Jurkat cells (work by Thomas Uhlmann) (Figure 4.10).

Using OFFGEL IEF, 66 arginine methylation sites were identified in primary T-lymphocytes and Jurkat cells (work by Thomas Uhlmann) of which 34 had not previously been described (Table 4.5). 3 lysine methylation sites were also found. Many arginine methylated peptides were so basic that they migrated out of the cathode end of the IPG strip and could be recovered from the electrode pads.

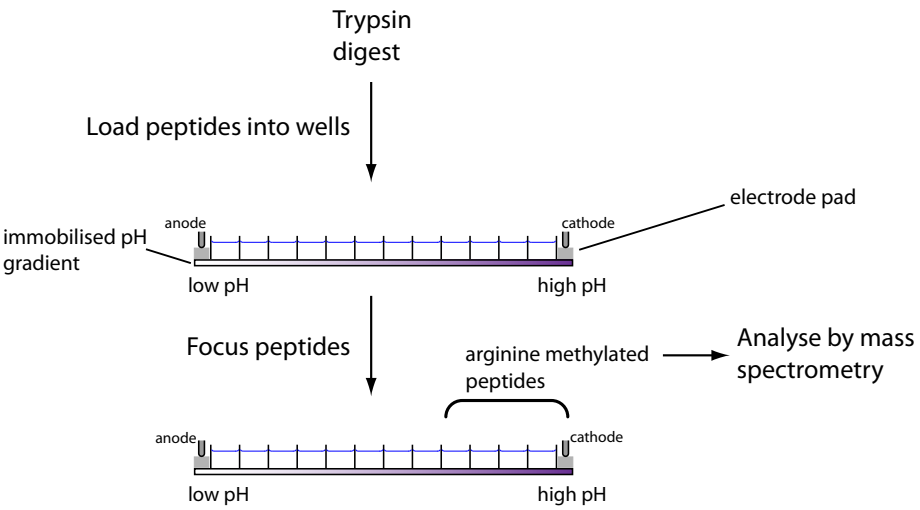


Figure 4.10: OFFGEL isoelectric focusing to enrich for arginine methylated peptides. Proteins were digested with trypsin and peptides were focused according to pI using a pH7-11 IPG strip in an OFFGEL apparatus. Arginine methylated peptides were identified in the most basic fractions.

4.2.6 Enrichment by weak cation exchange chromatography

Arginine methylated peptides were often more basic than the maximum pH of available IPG strips and exited the strip where they might be lost (Figure 4.8). Therefore an alternative method of selecting basic peptides was developed. Peptides bind to ion exchange columns when the pH of the solution is such that the peptide has a net charge opposite to that of the column. In chromatofocusing, peptides are bound to a cation exchange column in low pH. A pH gradient is then applied and peptides elute at their isoelectric point (Kang and Frey, 2004). Using the same principles, basic peptides could be

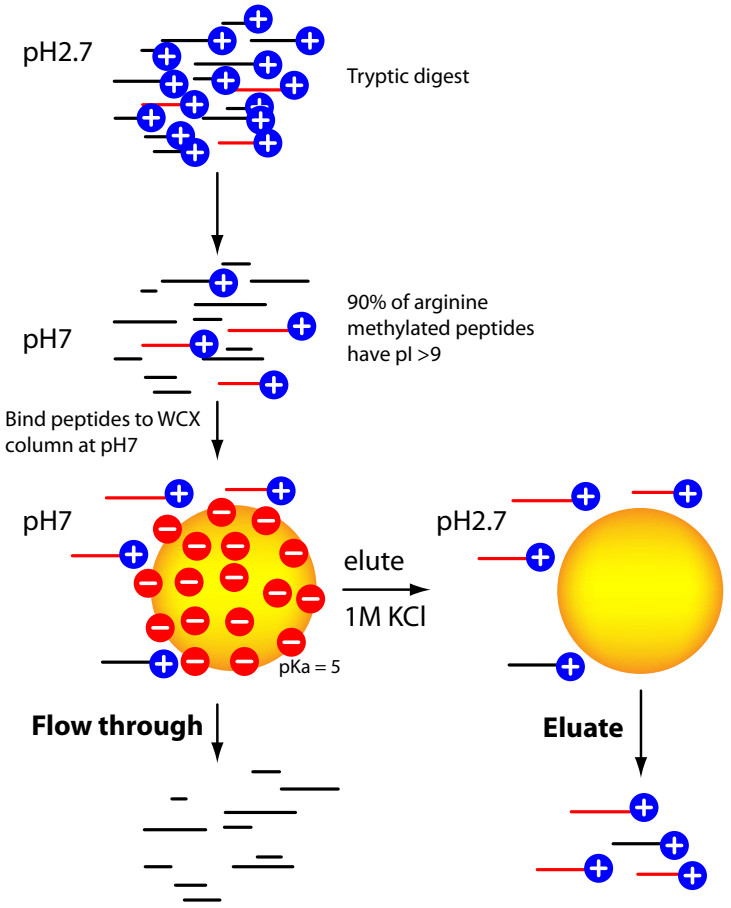


Figure 4.11: Weak Cation Exchange based separation of basic peptides. At acidic pH, most peptides from a tryptic digest will be positively charged. At pH7 the majority of peptides will be either uncharged or negatively charged. The majority of arginine methylated peptides (red lines) will remain positively charged as they have basic isoelectric points. Peptides are applied to the column at pH7 and only the basic peptides bind. Peptides with lower isoelectric points are found in the flow through. The functional groups of the weak cation exchange column can be protonated by acidic pH. Together with high salt this causes elution of basic peptides including those that are arginine methylated.

retained on a cation exchange column by carrying out the binding at higher pH (7 or above). Since 90% of arginine methylated peptides have a pI >9 (Figure 4.8), if the binding of peptides takes place at pH7 the majority of

the methylated peptides will be positively charged and will bind to the negatively charged column. The majority of non-methylated peptides will either be uncharged or negatively charged due to their lower isoelectric points and will therefore not bind to the column. If a weak cation exchange (WCX) column is used, the pH can be lowered to protonate the carboxylic acid groups on the matrix, together with high salt this should be sufficient to elute the bound basic peptides (Figure 4.11).

To test the ability of WCX to separate basic peptides, a tryptic digest of Jurkat whole cell lysate was applied to weak cation exchange columns at pH 7, 8, 9, 10, 11 and 12. The flow through and eluate fractions were collected and analysed by mass spectrometry and isoelectric points were calculated for each identified peptide. For comparison, a sample of unfractionated peptides (whole cell lysate sample) was also analysed. In the unfractionated sample, the majority of peptides identified have isoelectric points below 7 confirming the initial hypothesis (Figure 4.12, top). These peptides are more abundant than basic peptides and are therefore selected for fragmentation. By binding peptides at pH 7, 8, 9, 11 or 12 it was possible to separate basic peptides into the eluate fraction. At these pH values, separation of peptides was good with little overlap between peptides with isoelectric points below and above the binding pH (Figure 4.12). At pH 7, 8 and 9 the difference between the frequency of basic peptides (pI>7) and acidic peptides (pI<7) was less than in the whole cell lysate sample. This suggests that removing abundant acidic peptides enables more extensive fragmentation and identification of basic peptides. Binding at pH 10 did not result in good separation of peptides, possibly due to the buffer reacting with the column matrix. Arginine

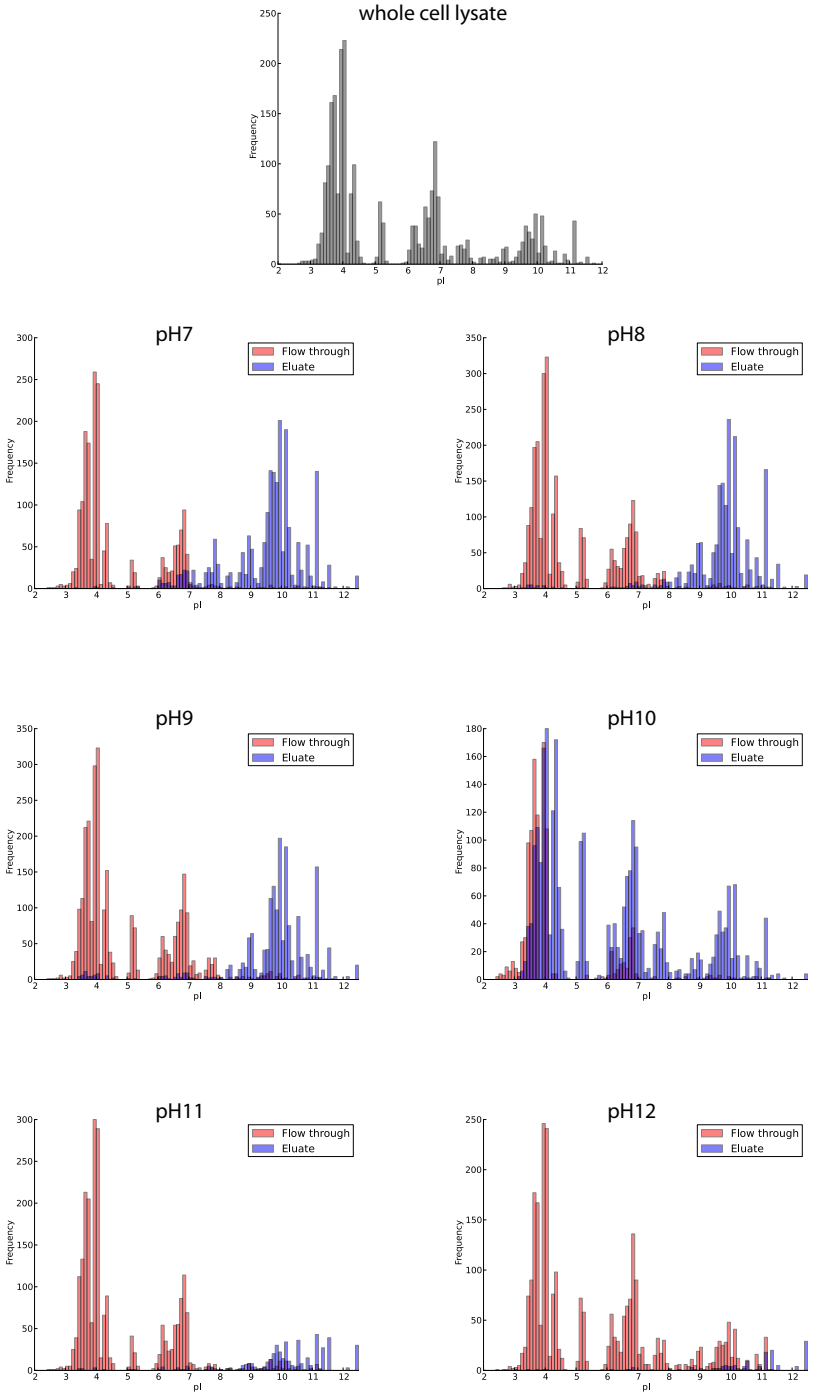


Figure 4.12: Separation of peptides above and below an isoelectric point cut-off using WCX. Tryptic peptides were applied to a WCX column in buffers of differing pH. Peptides in flow through and eluate fractions were identified by mass spectrometry and the frequency distribution of isoelectric points is plotted. For comparison, unfractionated peptides (whole cell lysate) were also analysed.

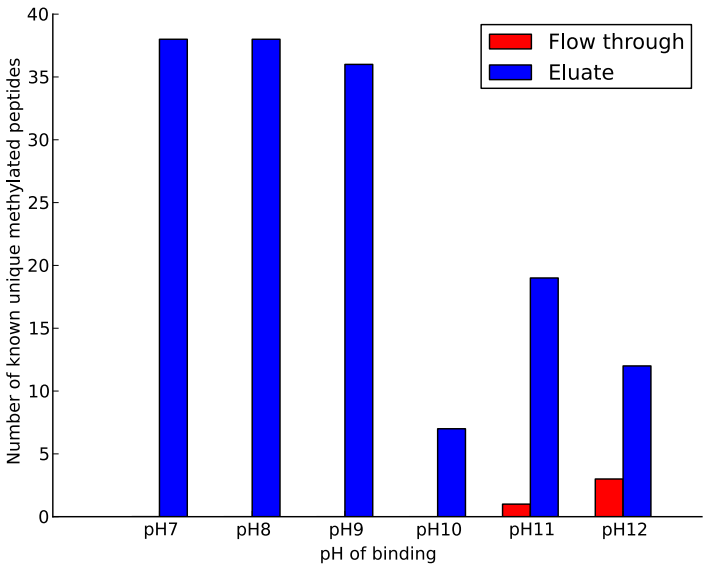


Figure 4.13: Enrichment of arginine methylated peptides using WCX at varying binding pH. A tryptic digest from unlabelled Jurkat cells was fractionated by WCX using buffers of differing pH. Since methyl SILAC material was not used, to assess the enrichment for arginine methylation the number of unique arginine methylated peptides known to be arginine methylated from previous methyl SILAC experiments are plotted for the flow through and eluate at each binding pH. Work carried out with Hannes Eichner.

methylated peptides should be enriched in the WCX eluate fraction as this contains basic peptides. In order to determine whether this was the case two approaches were used. Firstly, mass spectrometry data from flow through and eluate fractions at each pH were searched for arginine methylated peptides. Since material was not heavy methyl SILAC labelled, only the presence of known methylation sites could be confirmed (Figure 4.13). This analysis showed that no known arginine methylation sites were detected in the flow through fraction when binding took place <pH11. In the eluate fraction at pH7, 8 and 9, a significant number of arginine methylated peptides were found (38, 38 and 36 respectively). As the binding pH increases, more basic

peptides are found in the flow through and consequently arginine methylated peptides are detected in the flow through at pH11 and pH12. Even at pH12, 12 arginine methylated peptides were found in the eluate, highlighting the basic nature of these peptides.

In the second approach, peptides from flow through and eluate fractions at a binding pH of 7 or 8 were subject to amino acid analysis by Alta Biosciences (Birmingham, UK). Binding pHs of 7 or 8 were chosen as these gave the highest number of identified arginine methylated peptides. This involved hydrolysis of peptides to amino acids by boiling overnight in 6M HCl, followed by derivatization with ninhydrin and separation of amino acids by SCX chromatography. A standard of monomethylated, asymmetrically dimethylated and symmetrically dimethylated arginine was run to determine elution times for these amino acids. Being basic, arginine elutes later in the salt gradient than all other amino acids, therefore only this portion is shown (Figure 4.14). In the whole cell lysate, pH 7 flow through and pH 8 flow through samples, no peak corresponding to methylated arginine can be seen. In the eluate samples on the other hand, a small peak corresponding to methylated arginine is observed. The peak is too small to distinguish between the different forms of methylated arginine. In the pH 7 and pH 8 eluate samples the proportion of arginine existing in the methylated form is 1.5%.

Both analyses (Figures 4.13 and 4.14) indicated that WCX could be used to enrich for arginine methylated peptides. Therefore, Jurkat cells were heavy methyl SILAC labelled, cells were lysed, proteins digested and tryptic peptides were separated by WCX at pH7 (work carried out together with Hannes Eichner). Flow through and eluate fractions were analysed by mass

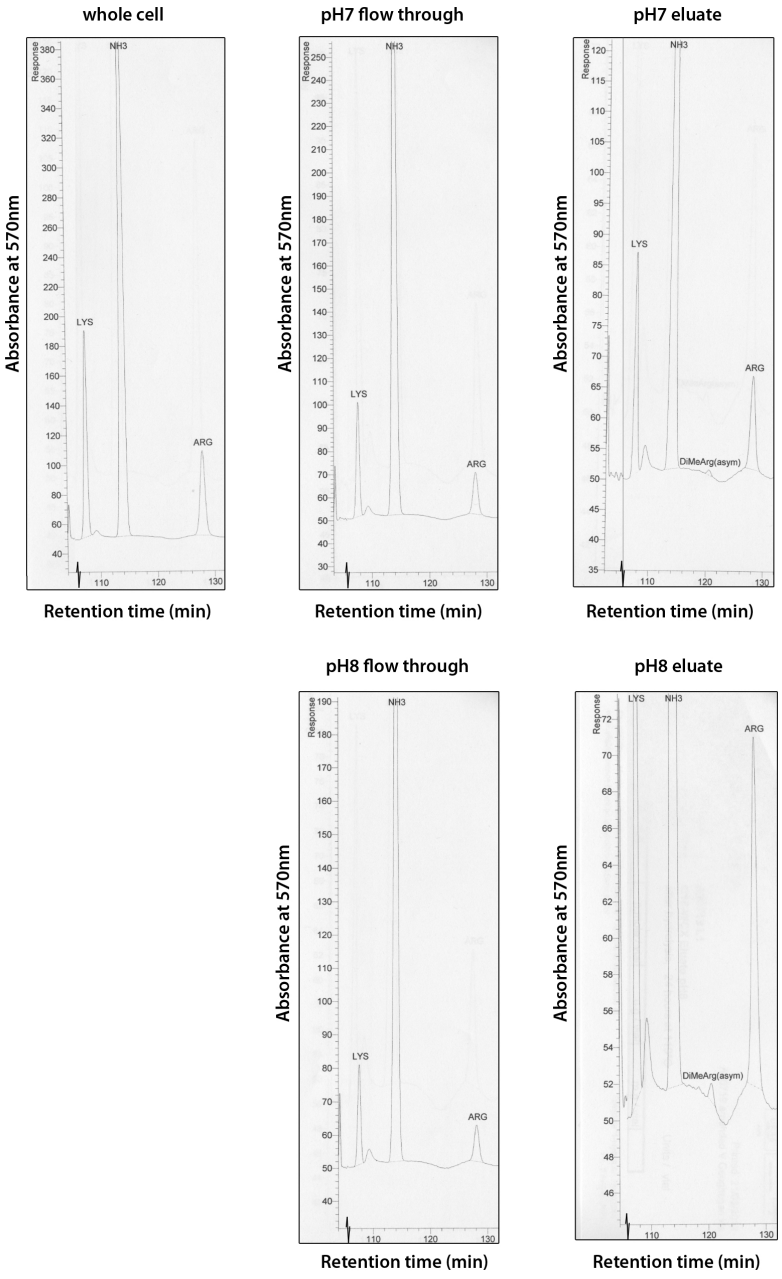


Figure 4.14: Amino acid analysis of peptides from flow through and eluate of WCX separation carried out at pH7 or pH8. Amino acid analysis of peptides from whole cell lysate is shown for comparison. Peaks corresponding to lysine (LYS), arginine (ARG) and methylated arginine are labelled. Work carried out with Hannes Eichner.

spectrometry. A total of 52 arginine methylation sites in 33 proteins were found, 26 of these sites were novel (Table 4.5). Two lysine methylation sites in one protein were also discovered. Only two of the arginine methylation sites were detected in the flow through fraction.

4.2.7 Enrichment by hydrophilic interaction liquid chromatography

Hydrophilic Interaction Liquid Chromatography (HILIC) fractionates peptides according to their hydrophilicity. A polar stationary phase (e.g. silica) and a polar mobile phase are typically used. Analytes are bound in the presence of a water miscible solvent such as acetonitrile and low amounts of water (typically 10%). The polar stationary phase adsorbs water, leading to a layer of water forming. Analytes partition between the organic solvent and the water layer according to how hydrophilic they are (Jian et al., 2010). Analytes are eluted from the column by increasing the water content of the mobile phase, hydrophobic analytes will elute in low amounts of water, whereas hydrophilic molecules require a higher water content to be eluted. Arginine methylated peptides are generally hydrophilic so they should be enriched in the hydrophilic fractions of a tryptic digest separated by HILIC (Figure 4.15). To discover arginine methylation sites, heavy methyl SILAC labelled human primary CD4+ T-lymphocytes or Jurkat cells (Thomas Uhlmann) were lysed and proteins digested with trypsin. Peptides were bound to a zwitterionic HILIC column in 90% acetonitrile 10% water and gradually eluted

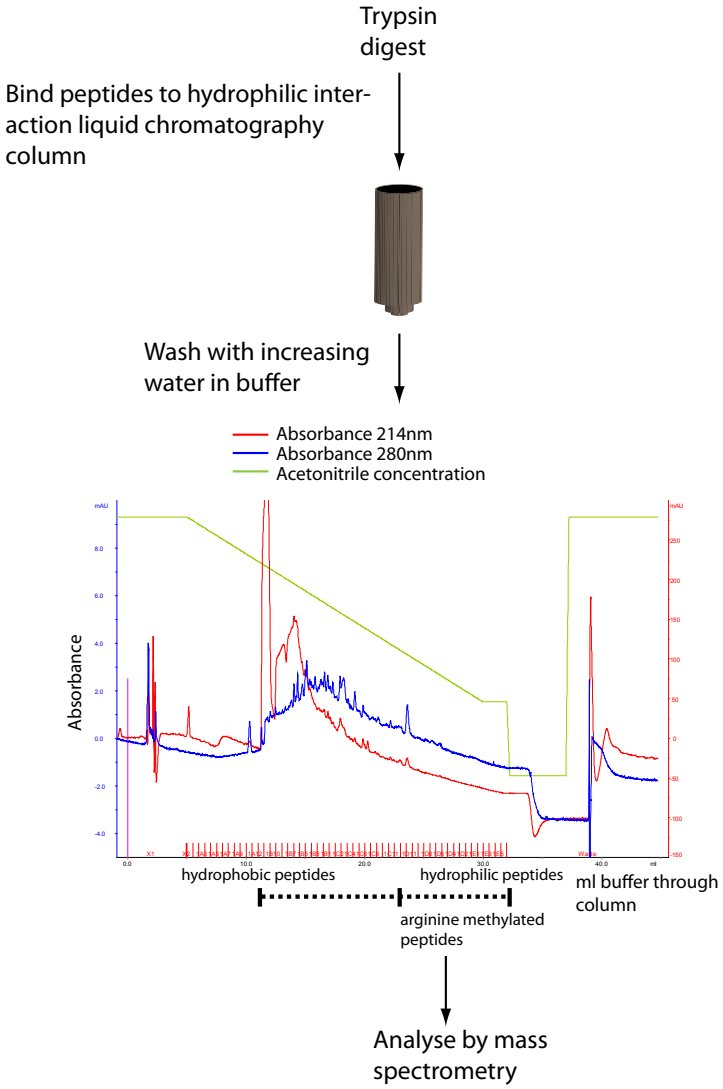


Figure 4.15: Hydrophilic Interaction Liquid Chromatography for enrichment of arginine methylated peptides. Peptides are applied to a HILIC column in acetonitrile containing 10% water. A gradient of increasing water content (and decreasing acetonitrile content) is applied to elute peptides according to hydrophilicity. Arginine methylated peptides should be found in the hydrophilic fractions.

by increasing the water content of the mobile phase. Analysis of the most hydrophilic fractions led to identification of 178 arginine methylation sites from primary T-lymphocytes and Jurkat cells (Table 4.5). These results indicate that HILIC is the most powerful separation method for enriching arginine methylated peptides prior to mass spectrometry. The majority of bottom-up proteomic studies use trypsin to cleave proteins into peptides. It can be useful to use proteases with differing cleavage specificity as this generates a different mixture of peptides and therefore results in different peptides being detected by mass spectrometry (Baldwin, 2004). In PTM discovery studies, reliance on one protease may mean that some sites are missed as after digestion modifications may be present in peptides which are not readily detectable (they may not ionise or may be too small or too large). Because arginine methylation often occurs in arginine rich regions, some methylated tryptic peptides may be too small to be detected. Additional sites may be discovered by using a protease with different cleavage specificity. Chymotrypsin cleaves C-terminal to amino acids with aromatic side chains such as tyrosine, phenylalanine and tryptophan and could lead to generation of different peptides compared to trypsin, and therefore identification of additional arginine methylation sites. A heavy methyl SILAC labelled lysate from primary human T-lymphocytes was digested with chymotrypsin and peptides were separated by HILIC as this was demonstrated to be the most effective method to enrich for methylated peptides. Arginine methylated chymotryptic peptides were found in the hydrophilic fractions and 37 additional arginine methylation sites were discovered which had not been observed by using trypsin (Table 4.5). Chymotryptic arginine methy-

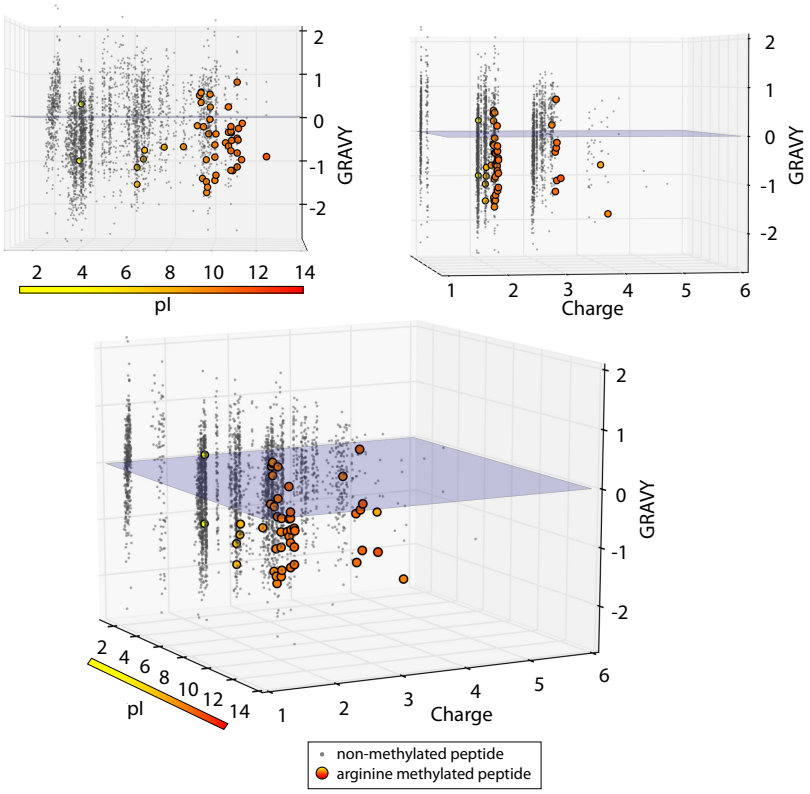


Figure 4.16: Physico-chemical properties of chymotryptic arginine methylated (coloured points) and non-methylated (grey points) peptides. Proteins from heavy methyl SILAC labelled primary T-lymphocytes were digested with chymotrypsin and peptides separated by HILIC. Fractions were analysed by mass spectrometry to search for arginine methylation sites.

Table 4.2: Physico-chemical properties of tryptic and chymotryptic non-methylated and arginine methylated peptides. Charge, pI and GRAVY scores were calculated for all arginine methylated peptides identified by 7E6 immunoprecipitation, SCX, IEF, WCX and HILIC. Properties of non-methylated peptides were calculated for peptides identified after IEF (trypsin) or HILIC (chymotrypsin).

		Proportion of peptides with charge >= +3	Mean pI	Mean GRAVY score
Trypsin	Non-methylated	55%	7.10	-0.22
	Arginine methylated	96%	10.64	-0.96
Chymotrypsin	Non-methylated	32%	5.47	-0.23
	Arginine methylated	21%	9.73	-0.54

lated peptides are basic and hydrophilic (Figure 4.16 and Table 2) although slightly less so than tryptic methylated peptides. However they are not more highly charged than non-methylated chymotryptic peptides as cleavage pattern is unaffected by methylation.

4.2.8 Using a combination of separation methods

Linking separation methods in series may allow greater enrichment and identification of more arginine methylation sites. The greatest enrichment should be achieved by combining orthogonal separation methods. If two methods are combined in series and they separate analytes based on the same physico-chemical properties, the gain will be negligible. Tryptic arginine methylated peptides are highly charged, have high pIs and low GRAVY scores, to maximise enrichment potential a series of separation methods should separate peptides according to all three parameters.

HILIC was shown to be the most effective enrichment technique and therefore was chosen for enrichment in series. Performed with low pH buffers, HILIC separates peptides according to charge as well as hydrophilicity. To select peptides with high pI (pI>9), WCX was chosen as it has a high capacity for basic peptides with no upper limit on basicity (unlike IPG strips). WCX was performed prior to HILIC separation as WCX generates just one fraction for further separation (the eluate) and HILIC requires that peptides are desalted before application to the column. If WCX was performed after HILIC, many fractions would have to be separated by WCX making it cumbersome, and two desalting steps would have to be carried out, likely leading to losses. Heavy methyl SILAC labelled Jurkat cells were lysed, proteins digested with trypsin and peptides with pI>9 were selected using WCX at pH7. These peptides were then desalted and separated by HILIC. Hydrophilic fractions were analysed by mass spectrometry to identify arginine methylated peptides. However, only 5 additional sites (all novel) in 5 proteins were discovered (Table 4.5).

4.2.9 Analysis of arginine methylation sites discovered

In total, the use of 7E6, SCX, IEF, WCX, HILIC and WCX+HILIC enrichment led to the identification of 265 unique arginine methylation sites in 141 proteins in primary human CD4+ T-lymphocytes and Jurkat cells (Table 4.5). 204 (77%) of the methylation sites were novel and 103 (73%) of the proteins had not previously been described as arginine methylated. In ad-

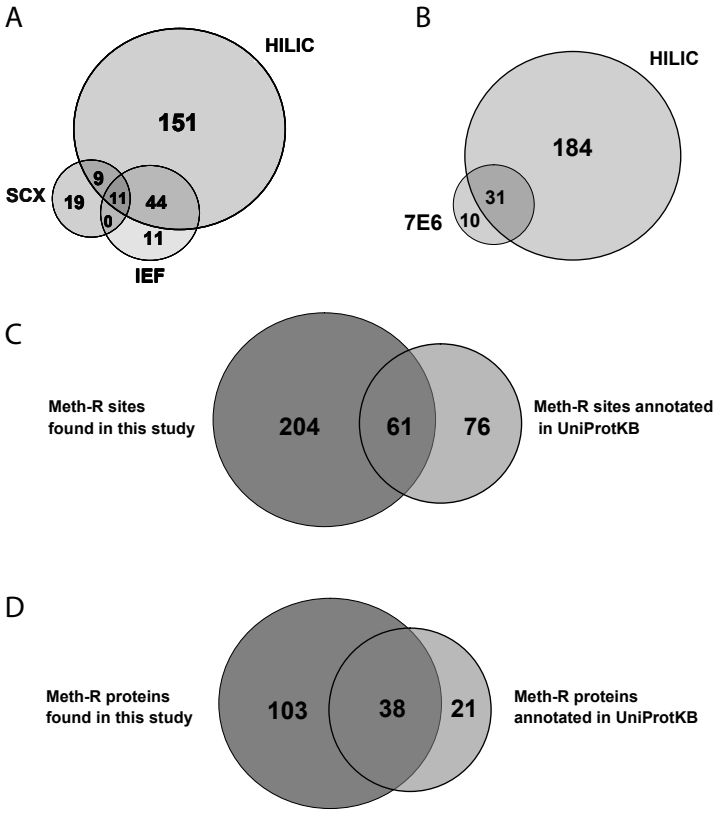


Figure 4.17: Analysis of arginine methylation sites identified in this study. **(A)** Comparison of the three most effective methods for enrichment of arginine methylated peptides. **(B)** Enrichment by HILIC compared to immunoprecipitation with the anti-methylarginine antibody 7E6. **(C)** Overlap between arginine methylation sites found here and those in UniProtKB. **(D)** Overlap between arginine methylated proteins found here and those in UniProtKB.

dition, 13 lysine methylation sites in 11 proteins were identified (Table 4.5). HILIC was the most effective method for identification of methylation sites,

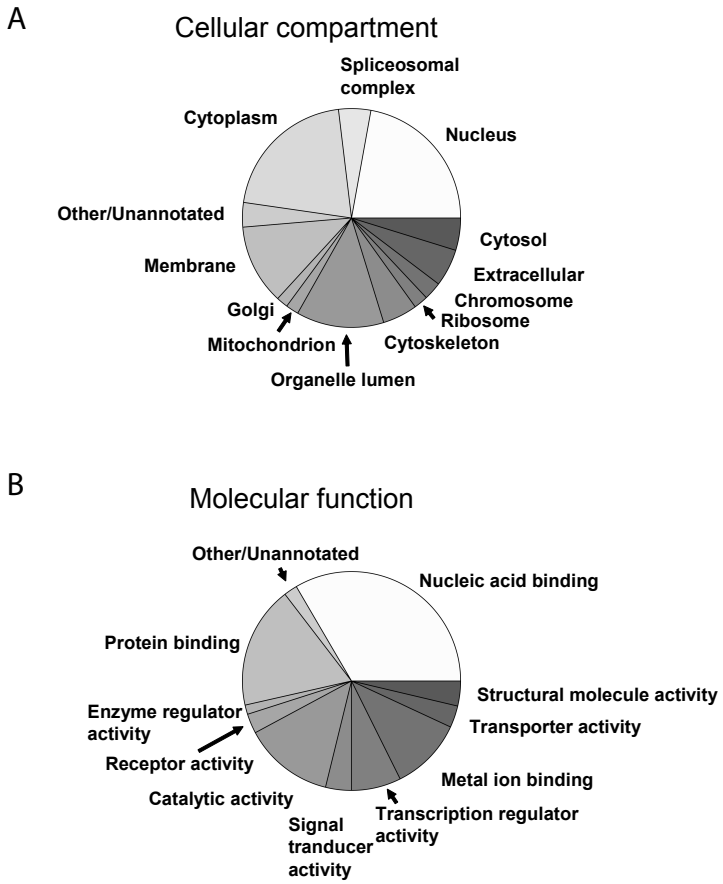


Figure 4.18: Gene ontology analysis of arginine methylated proteins identified. Proportion of arginine methylated proteins by annotated sub-cellular compartment **(A)**, or by annotated molecular function **(B)**.

significantly out-performing all other methods, followed by IEF, WCX, SCX and use of the 7E6 antibody (Figure 4.17, A and B). However some sites were uniquely identified by other methods indicating that HILIC could not provide complete coverage.

These results substantially increase the number of arginine methylation

sites and methylated proteins already known (137 sites in 59 proteins in UniProtKB) (Figure 4.17, C and D). Arginine methylated proteins are apparently equally present in the nucleus and cytoplasm (Figure 4.18, A). This is surprising as the majority of arginine methylated proteins are involved in transcription and RNA processing (Figure 4.18, B and Figure 4.19) and might be expected to mostly reside in the nucleus. Arginine methylated proteins were found to be involved in endosomal trafficking and re-arrangement of the actin cytoskeleton (Figure 4.19), processes which had not previously been suspected of involving methylation on arginine. A significant proportion of methylated proteins had no known annotated function as the proteins are poorly characterised.

Susceptibility of an amino acid to enzymatic modification is partly determined by its sequence context. To determine which amino acids are over-represented near methylarginine to aid prediction of methylation sites and gain an understanding of the structural context, all arginine methylation sites were submitted to WebLogo (Crooks et al., 2004) for generation of sequence logos (Schneider and Stephens, 1990). Sequence logos are a method of visualising representation of amino acids of nucleic acids in binding or modification sites. Glycine was commonly seen in the sequence surrounding a methylated arginine, in particular it was very often C-terminal to the arginine (Figure 4.20, top). Arginine is relatively frequent -2 and +2 residues from the methylation site. Proline residues are also commonly found surrounding a methylarginine (Figure 4.20, bottom). Glycine, arginine and proline were observed to be some of the most common amino acids flanking arginine methylation sites. To determine their distribution around methylation

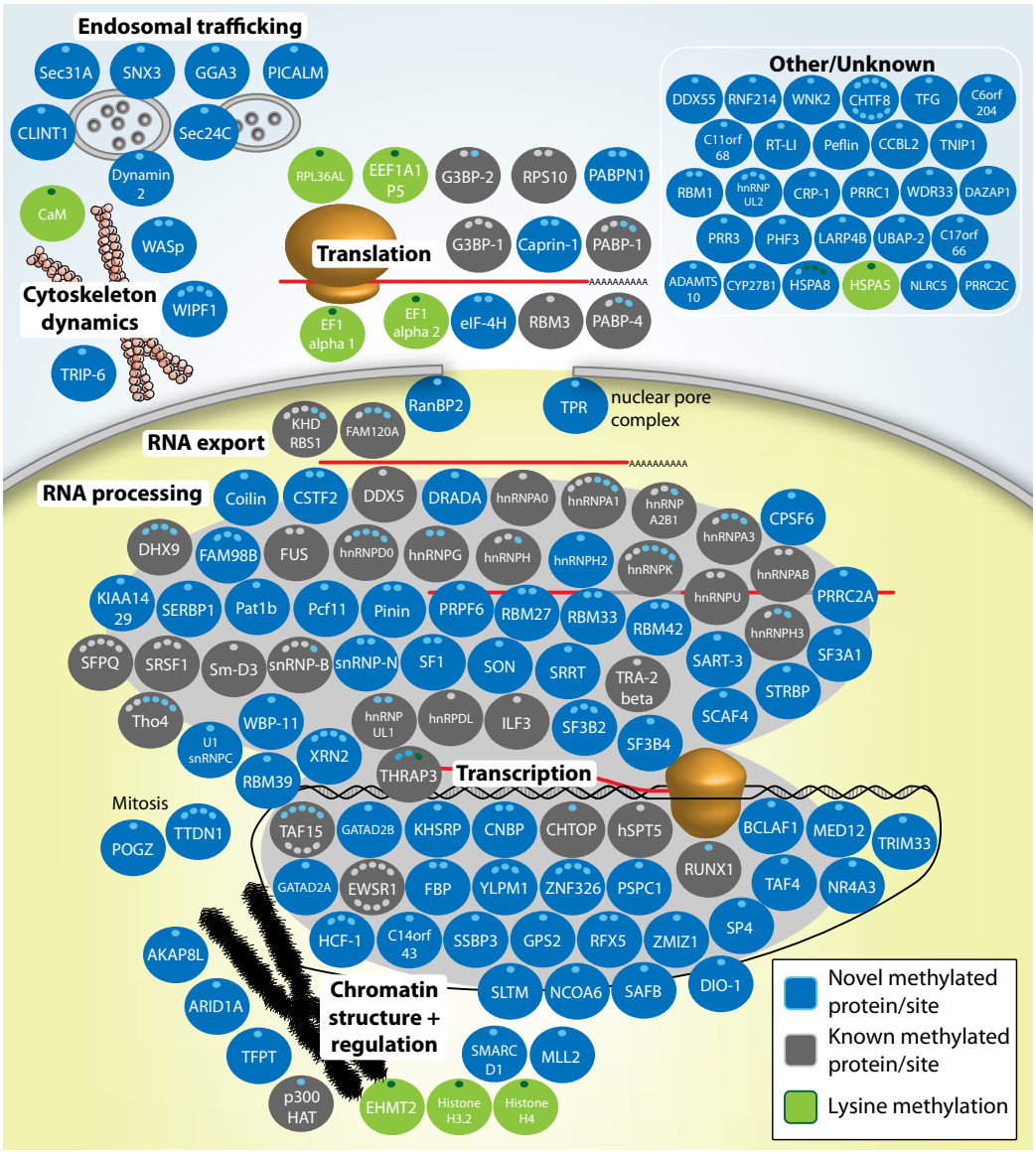


Figure 4.19: Overview of all methylated proteins identified. Proteins are placed in a cellular context based on annotation in UniProtKB. Grey represents a known and blue a novel arginine methylated protein. Individual sites are shown as points, with grey denoting a known site and blue a novel site. Lysine methylated proteins are shown in green.

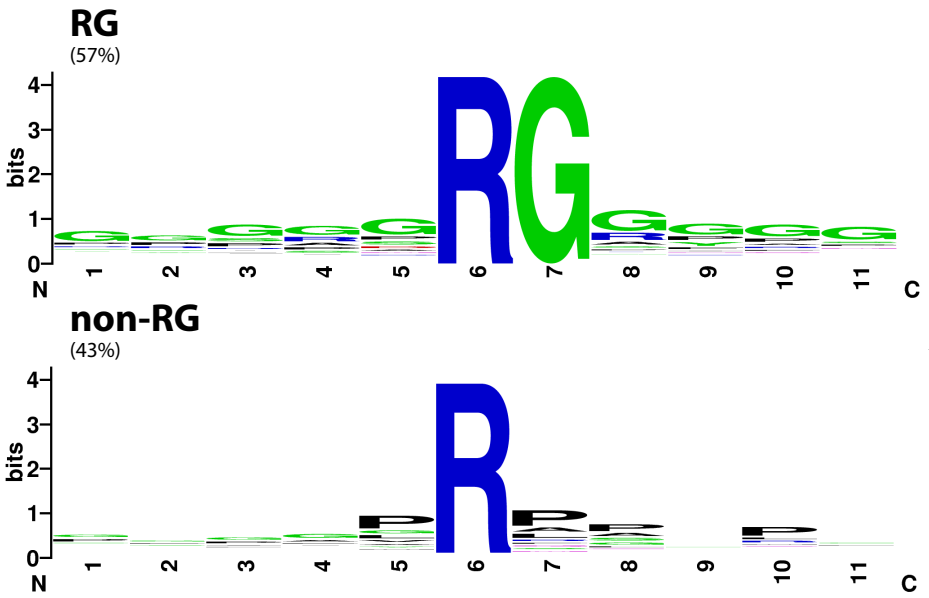


Figure 4.20: Sequence logo representation of all arginine methylation sites discovered. Sequences containing the methylated arginine and 5 flanking amino acids either side were separated into two groups, one group where glycine was present C-terminal to the methylated arginine and the other group were the remainder of the sequences. These were uploaded to WebLogo for generation of sequence logos.

sites, frequency of occurrence for each of these amino acids at each position either side of methylarginine was plotted (Figure 4.21). Glycine was found to occur frequently at all positions, but was especially frequent adjacent to the methylated arginine on the C-terminal side (Figure 4.21, A). Arginine was observed to be favoured at positions -2 and +2 either side of the methylarginine (Figure 4.21, B). Proline was found to be significantly favoured at positions -1 and +4 relative to methylarginine (Figure 4.21, C). Arginine methylation sites were also searched using motif-x, a tool for extracting common motifs, originally conceived for determining phosphorylation consensus motifs (Schwartz and Gygi, 2005). Motif-x takes two datasets, one consisting of modification sites (in this case arginine methylation sites) and the other is

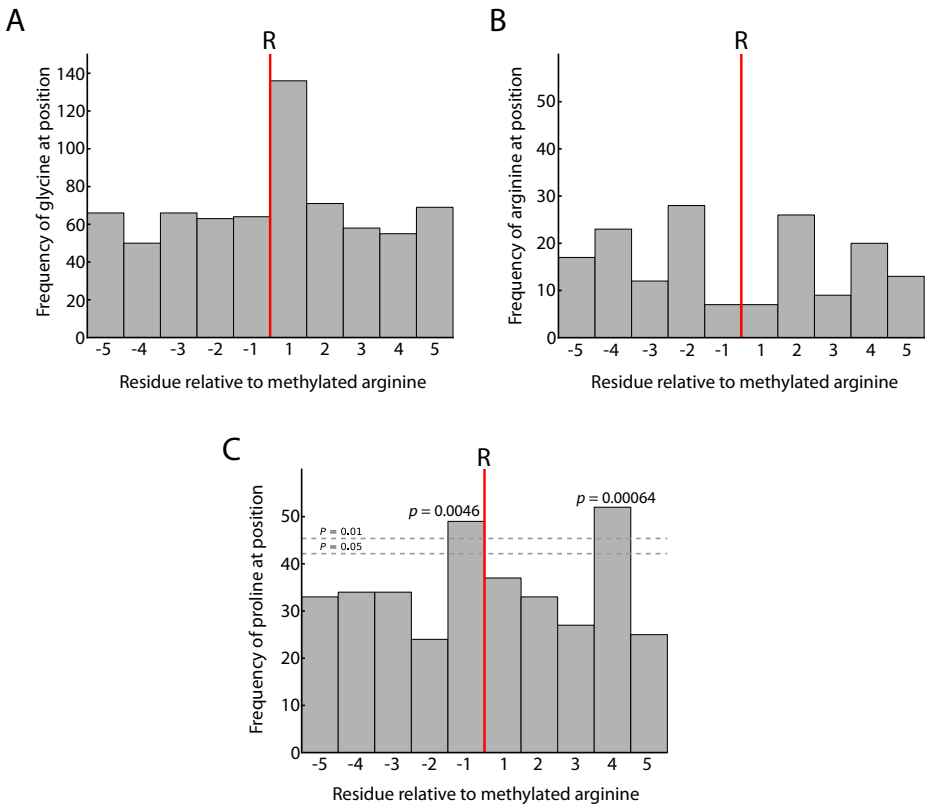


Figure 4.21: Frequency of glycine **(A)**, arginine **(B)** and proline **(C)** at each position relative to methylarginine. Probabilities were calculated by bootstrapping.

a background dataset of all sequences in the organism of interest. An occurrence probability of each amino acid in the modification dataset is calculated and compared with the background probability of occurrence in all sequences from the organism. Using this search algorithm confirmed the glycine rich nature of methylation sites, RG and RGG motifs scored highly(Table 4.3). Two proline motifs were also highlighted, a proline C-terminal to methylarginine is common and the importance of proline at +4 was confirmed. The motif-x algorithm attempts to separate out distinct motifs, thus when the RG motif is recognised, all sequences containing this are removed from fur-

Table 4.3: Significant arginine methylation motifs found by motif-x. Arginine methylation sites were aligned on arginine and five residues either side were taken for searching by motif-x (foreground dataset). The IPI human proteome was used as the background dataset.

Motif	Motif score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
....GRG....	32.00	55	243	6125	756422	27.95
.....RGG...	31.65	35	188	4042	750297	34.56
.....RG....	16.00	50	153	45183	746255	5.40
.....RP....	16.00	37	103	49413	701072	5.10
.....R...P.	10.07	22	66	42398	651659	5.12

ther motif searches. This means that when the proline motifs are discovered by the algorithm, the number of sequences which match these will be lower, even though some glycine-arginine motifs may also contain prolines.

Methylation on arginine was observed to occur frequently in proline rich sequences, proline was favoured at positions -1 and +4 giving rise to the consensus motif (P)Rxxx(P). Many domains such as SH3 domains which bind proline rich sequences contain hydrophobic grooves at the binding interface, lined by aromatic amino acids . Interacting prolines within proline rich sequences are therefore often preceded by a hydrophobic amino acid (Li, 2005). To test whether this was the case for (P)Rxxx(P) arginine methylated sequences, amino acids were assigned a hydropathy score (Kyte et al., 1982) and average hydropathy values were plotted for each amino acid position (Figure 4.22). This revealed that the +4 proline is often preceded by a hydrophobic amino acid, suggesting that these proline containing arginine methylation sites interact with SH3 or SH3-like domains.

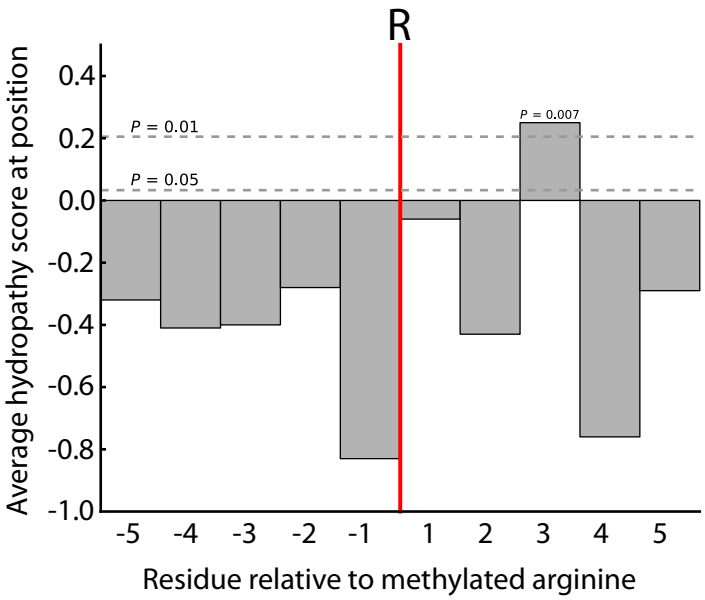


Figure 4.22: Average hydropathy scores for amino acids in (P)Rxxx(P) type arginine methylation sites. Amino acids were assigned a hydropathy score according to Kyte and Doolittle (Kyte et al., 1982), these were averaged for each amino acid position in arginine methylation sites containing a proline at -1 and/or +4. A positive hydropathy score indicates hydrophobicity and a negative score hydrophilicity. Probabilities were calculated by bootstrapping.

4.2.10 Reliability of identified methylation sites

In mass spectrometry based proteomics, a fragmentation spectrum can be falsely assigned to a peptide sequence. This is especially problematic when attempting to detect methylation sites, heavy methyl SILAC labelling was introduced to provide increased confidence (Ong et al., 2004) (Figure 4.4). However while this approach clearly does provide increased confidence there was no quantitative measure of this. To determine the false discovery rate (FDR) for arginine methylation with and without using heavy methyl SILAC, a mass spectrometry dataset from heavy methyl SILAC labelled peptides

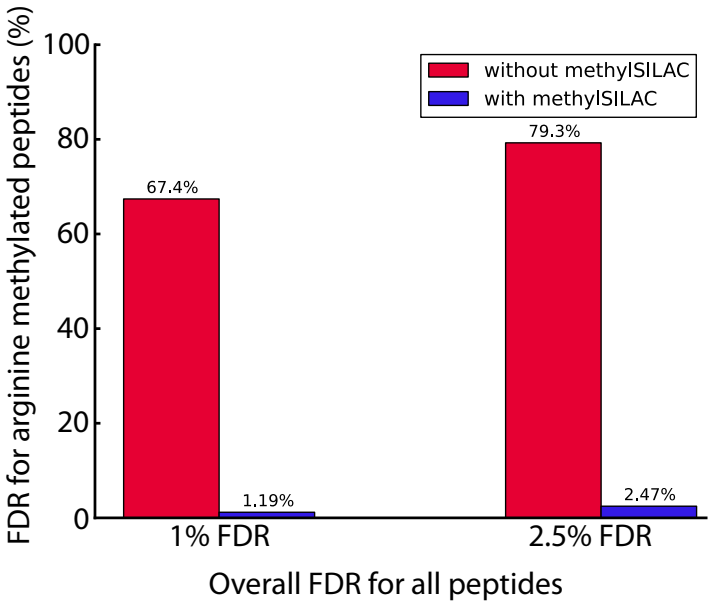


Figure 4.23: Effect of requiring presence of methyl SILAC pairs on false discovery rate for detection of arginine methylation. A dataset of heavy methyl SILAC labelled peptides separated by HILIC was searched for arginine methylation sites. At an overall FDR of 1% or 2.5%, putative arginine methylated peptides were extracted and the proportion of decoy hits was counted, this gave the FDR when presence of methyl SILAC pairs was not required for accepting the site as valid. To calculate the false positive rate when identifying sites with methyl SILAC, all methylated decoy hits were searched for confirming methyl SILAC pairs, the total number found was used to calculate the FDR.

separated by HILIC was searched for arginine methylation sites. The overall FDR for methylated and non-methylated peptides was 1%, however when all putative arginine methylated peptides were extracted, the proportion of these corresponding to decoy hits was 67% (Figure 4.23). This represents the FDR in the absence of heavy methyl SILAC as decoy hits were taken as being false positives without any screening for presence of methyl SILAC pairs. To determine the rate of false positives when identifying arginine methylation sites by heavy methyl SILAC, all decoy hits at 1% FDR were inspected for

the presence of a confirming methyl SILAC pair, the majority did not contain methyl SILAC pairs and were therefore not true false positives. This gave a FDR with methyl SILAC of 1.2% showing that heavy methylSILAC gives a 60 fold decrease in the number of false positives. This was repeated for the same dataset at an overall FDR of 2.5%, again methyl SILAC decreased the number of false positives (30 fold).

4.2.11 Improving heavy methyl SILAC

The introduction of labelling by heavy methyl SILAC greatly increased identification confidence for methylation (Ong et al., 2004) (Figure 4.23). However one flaw of labelling in this way is that isotope labelled methionine will also be incorporated into proteins. After digestion with proteases and analysis by mass spectrometry, every methionine containing peptide will occur in pairs indistinguishable in mass difference from methylated SILAC pairs (Figure 4.24, A). This complicates the peptide mixture, thus reducing the likelihood that a methylated peptide will be sequenced. In addition, confidence is decreased, as a putative methylated hit may be corroborated by the presence of a SILAC pair, but this pair may come from a methionine containing peptide. To solve this problem, a new labelling scheme was devised. Light methionine was substituted for L-methionine-($^{13}C_4$) in the heavy methyl SILAC labelling method. This meant that both isotope labelled methionines were $\sim$4 Da heavier than light methionine, so the method is called isomethionine methyl SILAC (Figure 4.24, B). However the isotopes are dis-

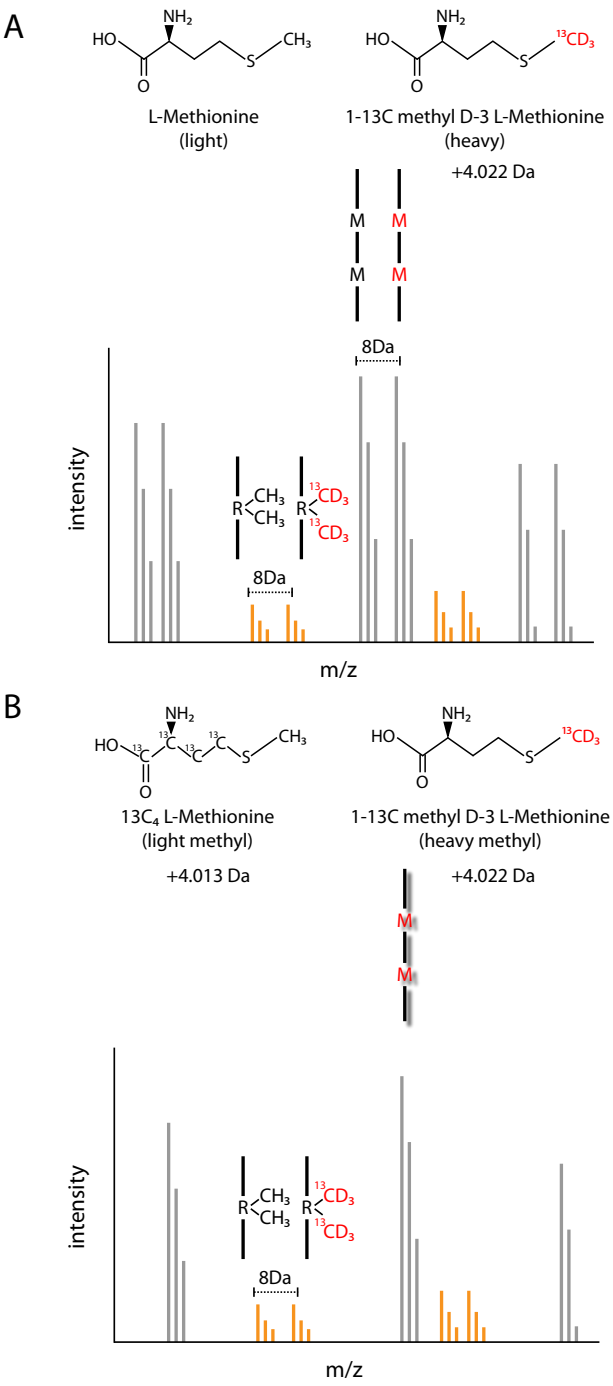


Figure 4.24: Labelling with heavy methyl SILAC **(A)** and isomethionine methyl SILAC **(B)**. In heavy methyl SILAC, methylated peptides (orange) have the same mass shift as methionine containing peptides (grey). In isomethionine methyl SILAC, only methylated peptides occur in pairs.

tributed differently in the two methionines. In L-methionine-($^{13}C_4$) the isotopes are not in the methyl group transferred during methylation reactions. Thus light and heavy methylated peptides will occur in methyl SILAC pairs, whilst light and heavy methionine containing peptides will not.

To verify that isomethionine methyl SILAC eliminates pairs due to methionine containing peptides, Jurkat cells were labelled by heavy methyl SILAC and isomethionine methyl SILAC. Proteins were digested with trypsin and arginine methylated peptides were enriched by WCX at pH7 and analysed by mass spectrometry. SILAC pairs due to methionine containing peptides were indeed eliminated by using isomethionine methyl SILAC, whilst methylated pairs were still present (Figure 4.25).

4.2.12 Occupancy of arginine methylation sites

Measuring the fraction of a site existing in a modified form can provide additional information to understand the dynamics and biological role of PTMs. However, determining occupancy by mass spectrometry is not straightforward for many PTMs as the modification alters the charge of the peptide. A phosphorylated peptide will typically ionise less efficiently than a non-phosphorylated peptide due to the negative charge, making it difficult to directly compare abundances by mass spectrometry. Methods to measure phosphorylation occupancy on a large scale have been developed which involve using alkaline phosphatase to dephosphorylate one half of the tryptic digest and measuring the resultant increase in intensity of the non-phosphorylated

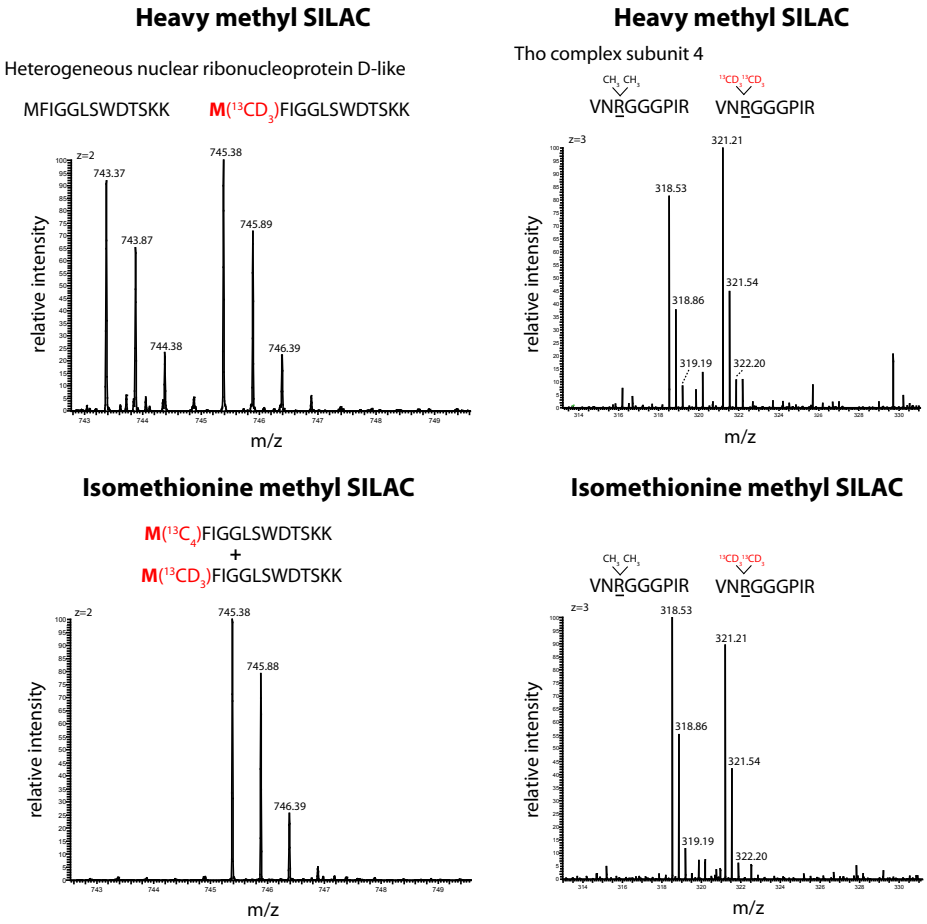


Figure 4.25: A sample of peptide mass spectra from Jurkat cells labelled by heavy methyl SILAC (top) or isomethionine methyl SILAC (bottom).

form (Wu et al., 2011). Methylation on arginine does not appreciably affect the charge or ionization efficiency of a peptide (Pahlich et al., 2005) making it possible to compare intensities of non-methylated and methylated forms to estimate the proportion of the site occupied by methylation. However, because arginine methylation causes trypsin to mis-cleave, comparing intensities is only possible if the peptide contains a methylation site where cleavage cannot occur in the absence of methylation (e.g. due to a C-terminal

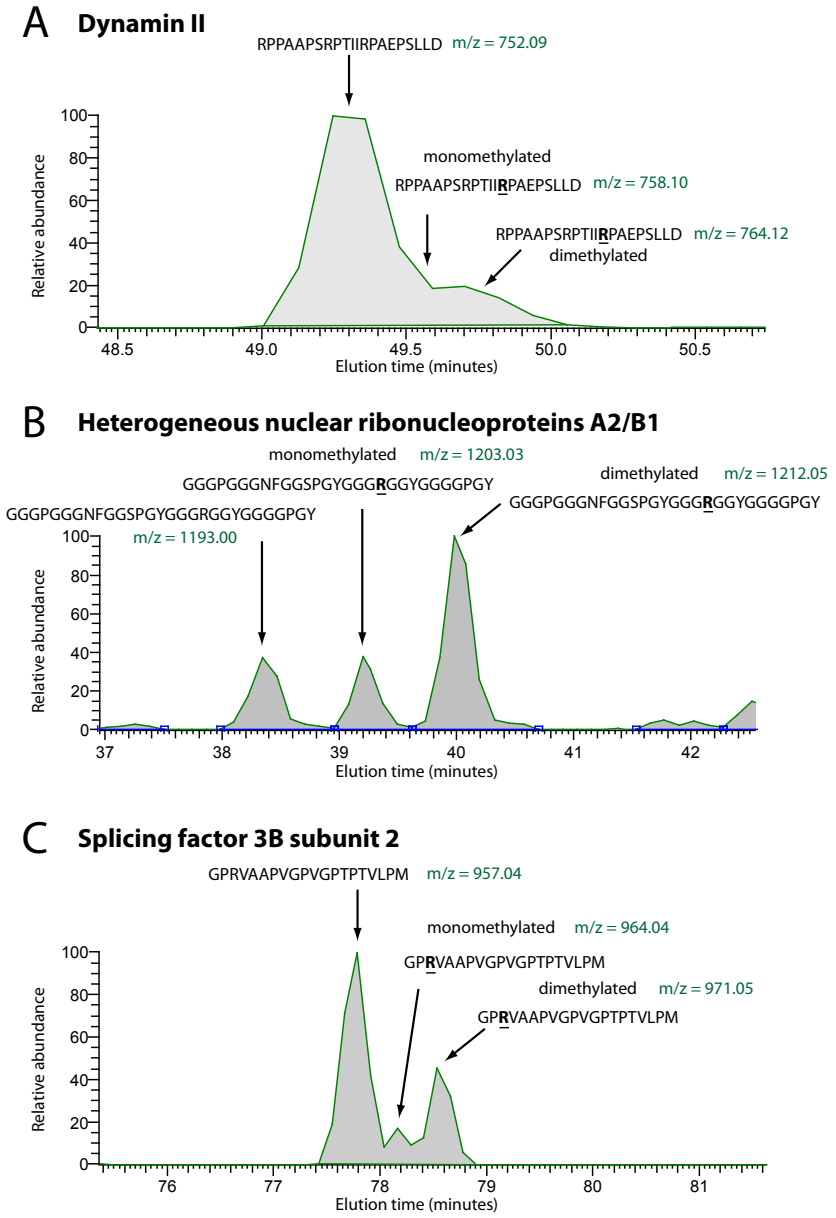


Figure 4.26: Comparison of intensities for non-methylated and arginine methylated peptides. Intensities of peptides as they eluted from the in-line column were measured using Thermo Xcalibur software and resulting peak areas were calculated to determine total abundance. Where the monomethylated form was not resolved its contribution to the total area of the ions could be determined by subtracting the area due to the ion of the corresponding mass. Only methylation sites where the cleavage pattern is unaffected by methylation were analysed. Peptides in B and C are chymotryptic.

proline). This means non-methylated and methylated peptides will have the same length and will differ only in their methylation status. Mass spectrometry data from HILIC enrichment was searched for non-methylated and methylated forms of peptides which matched this criterion. In data from tryptic peptides, only one set of peptides was identified which allowed direct comparison (Figure 4.26, top). This revealed that approximately 17% of Dynamin II is methylated on arginine, with monomethylation accounting for 5% of this. Chymotrypsin cleavage is unaffected by methylation on arginine, making it possible to compare non-methylated and methylated forms of any peptide, providing both have been sequenced and identified. Two examples of methylated peptides were found which allowed this comparison (Figure 4.26, middle and bottom). For Heterogeneous nuclear ribonucleoproteins A2/B1, non-methylated, monomethylated and dimethylated forms were clearly resolved. Peak areas showed that 76% of this site exists in the methylated form, monomethylation accounting for 18% of the methylated peptide. In Splicing factor 3B subunit 2, the occupancy of the methylation site identified was calculated to be 50%, with monomethylation comprising 27% of the total methylated form. These values are only approximate as it is possible that arginine methylation affects ionisation efficiency by ESI.

4.2.13 Comparing arginine methylation sites in unstimulated and stimulated T-lymphocytes

Activation of primary human T-lymphocytes causes increases in arginine methylation and levels of PRMT1 and PRMT5 (Chapter 3). Determining the identity of the proteins responsible for the observed increases in arginine methylation would provide some insight into the role of arginine methylation in T-lymphocyte activation. The most rapid and accurate way of comparing methylation sites in unstimulated and stimulated T-lymphocytes is by using mass spectrometry. However, identifying arginine methylation with high confidence requires that cells are labelled by heavy methyl SILAC (Figure 4.4). This is not possible to achieve in unstimulated T-lymphocytes as they do not proliferate. Therefore a new labelling scheme was devised combining Arg10 with heavy methyl SILAC (Figure 4.27). Combining Arg10 labelling with heavy methyl SILAC to compare unstimulated and stimulated cells provides two benefits. Firstly, arginine methylated peptides can still be corroborated by 1:1 methyl SILAC pairs. Secondly, the Arg10 label permits changes in abundance at the protein level to be measured. This means that when a methylated peptide is identified the effect of stimulation on methylation status can be measured, but also the effect on the total level of that protein can be measured. The change in the proportion of the protein that is methylated can thus be calculated.

Primary human CD4+ T-lymphocytes were isolated from blood and labelled according to Figure 4.27. Proteins were digested with trypsin, and arginine methylated peptides were enriched by IEF as this requires low

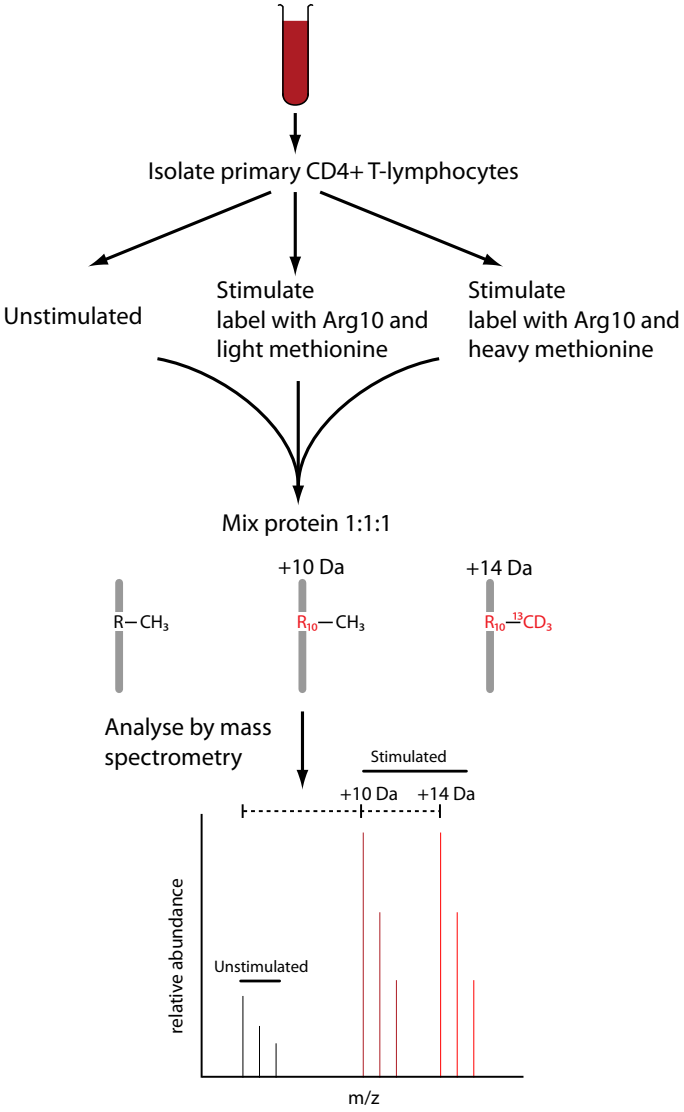
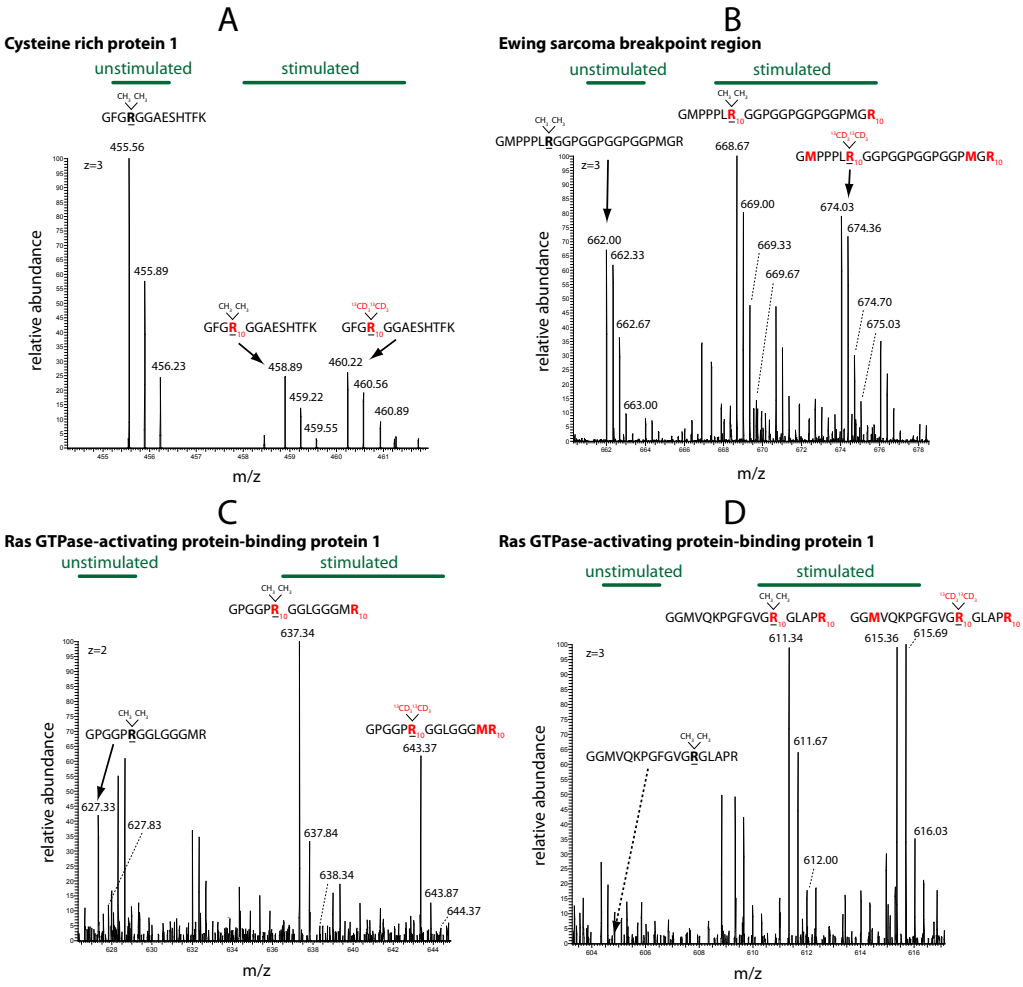
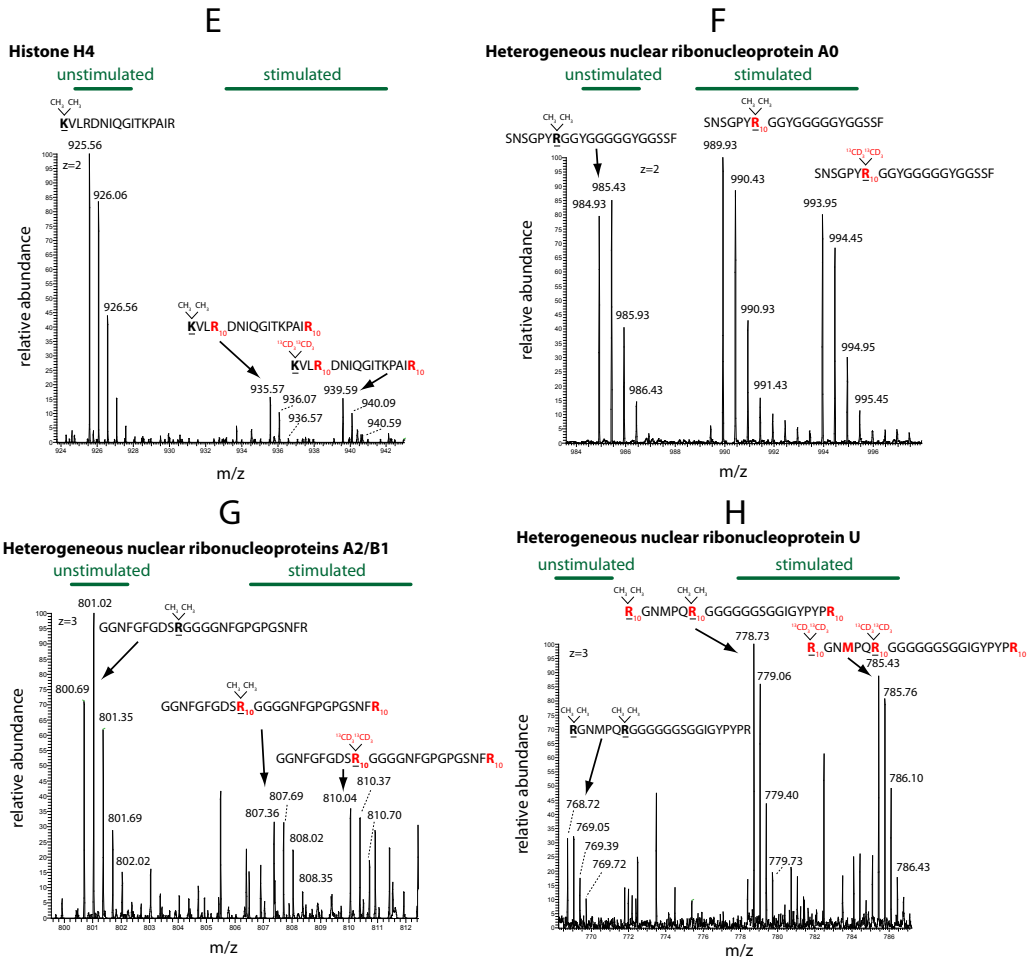
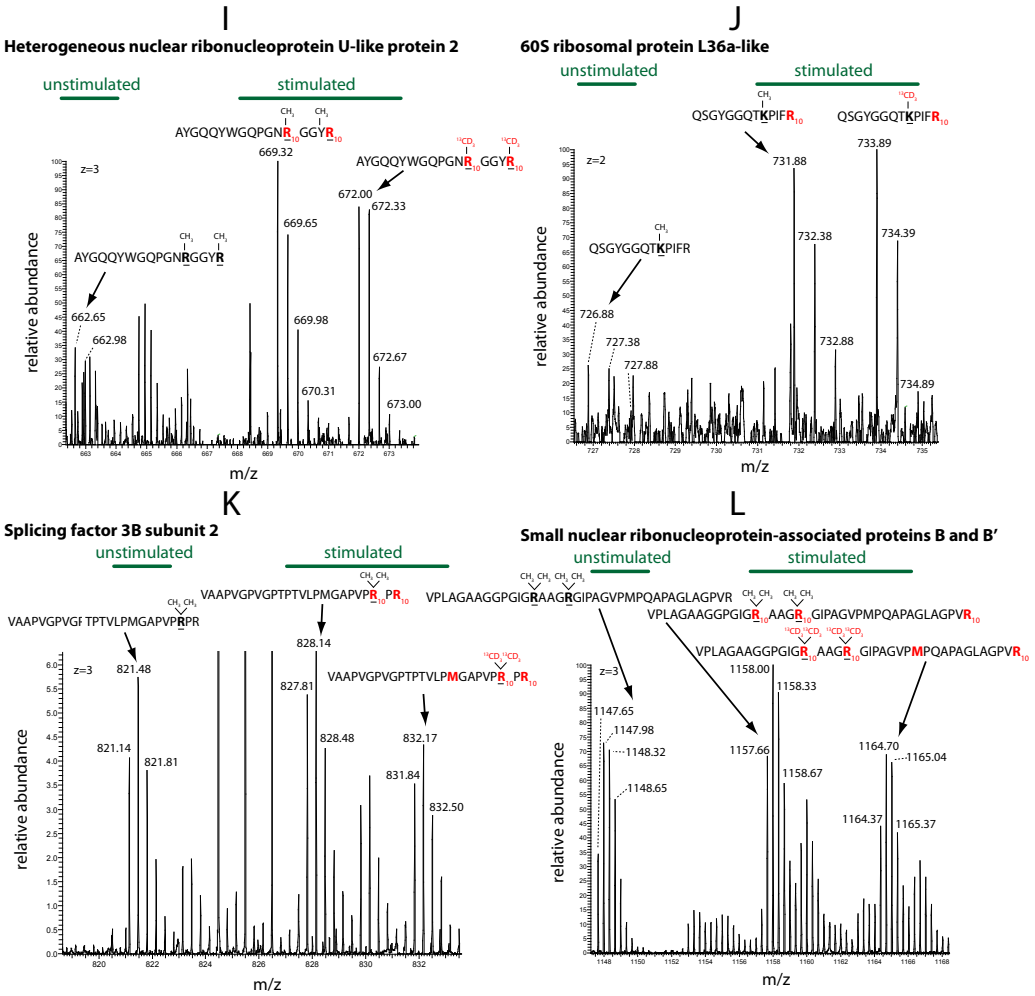
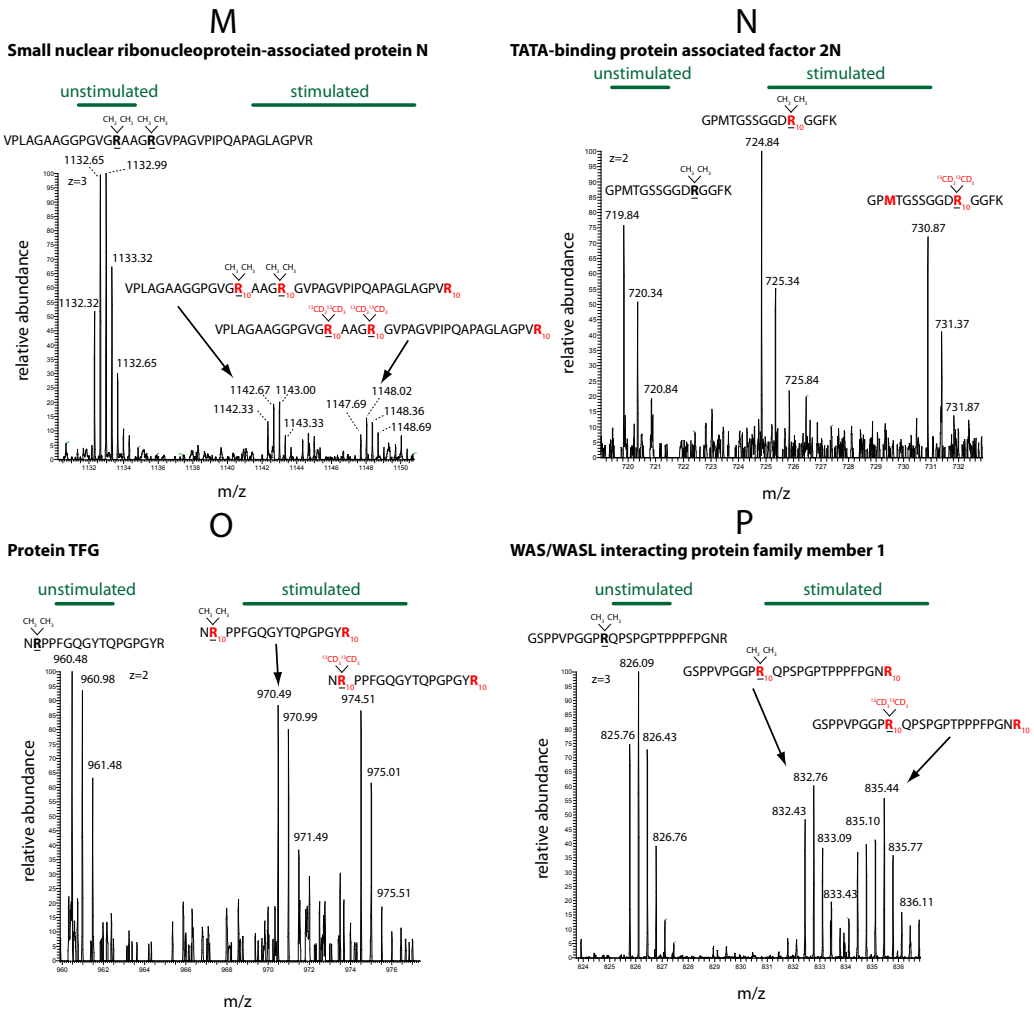


Figure 4.27: Stable isotope labelling method to compare levels of arginine methylation in unstimulated and stimulated primary T-lymphocytes. Primary human CD4+ T-lymphocytes are isolated from blood and a portion snap frozen (unstimulated cells). The remainder are divided into two cultures, stimulated and labelled with Arg10 and either light or heavy methionine. After being stimulated for 7 days, rested for 3 days and re-stimulated for 7 days, cells are harvested. Unstimulated, stimulated (light methionine labelled) and stimulated (heavy methionine labelled) cells are mixed in a 1:1:1 protein ratio and relative intensities of arginine methylation are measured.


A
Cysteine rich protein 1
unstimulated
stimulated
GFGRGGAESHTFK
455.56
z=3
455.89
456.23
GFGR10GGAESHTFK
458.89
459.22
459.55
GFGR10GGAESHTFK
460.22
460.56
460.89
relative abundance
m/z
B
Ewing sarcoma breakpoint region
unstimulated
stimulated
GMPPPLR10GGPGGPGGPGGPMGR10
GMPPPLRGGPGGPGGPGGPMGR
668.67
z=3
GMPPPLR10GGPGGPGGPGGPMGR10
674.03
669.00
674.36
662.00
669.33
662.33
669.67
674.70
662.67
675.03
663.00
relative abundance
m/z
C
Ras GTPase-activating protein-binding protein 1
unstimulated
stimulated
GPGGPR10GGLGGGMR10
637.34
z=2
GPGGPRGGLGGGMR
627.33
627.83
GPGGPR10GGLGGGMR10
643.37
637.84
638.34
643.87
644.37
relative abundance
m/z
D
Ras GTPase-activating protein-binding protein 1
unstimulated
stimulated
GGMVQKPGFGVGR10GLAPR10
GGMVQKPGFGVGR10GLAPR10
611.34
615.36
615.69
z=3
611.67
GGMVQKPGFGVGRGLAPR
616.03
612.00
relative abundance
m/z


E
Histone H4
unstimulated
stimulated
KVLRDNIQGITKPAIR
925.56
z=2
926.06
926.56
KVLR10DNIQGITKPAIR10
KVLR10DNIQGITKPAIR10
935.57
936.07
936.57
939.59
940.09
940.59
relative abundance
m/z
F
Heterogeneous nuclear ribonucleoprotein A0
unstimulated
stimulated
SNSGPYRGGYGGGGGYGGSSF
SNSGPYR10GGYGGGGGYGGSSF
SNSGPYR10GGYGGGGGYGGSSF
z=2
985.43
984.93
985.93
986.43
989.93
990.43
990.93
991.43
993.95
994.45
994.95
995.45
relative abundance
m/z
G
Heterogeneous nuclear ribonucleoproteins A2/B1
unstimulated
stimulated
GGNFGFGDSRGGGGNFGPGPGSNFR
801.02
z=3
800.69
801.35
801.69
802.02
GGNFGFGDSR10GGGGNFGPGPGSNFR10
GGNFGFGDSR10GGGGNFGPGPGSNFR10
807.36
807.69
808.02
808.35
810.04
810.37
810.70
relative abundance
m/z
H
Heterogeneous nuclear ribonucleoprotein U
unstimulated
stimulated
R10GNMPQR10GGGGGGSGGIGYPYPR10
R10GNMPQR10GGGGGGSGGIGYPYPR10
z=3
778.73
779.06
779.40
779.73
RGNMPQRGGGGGGSGGIGYPYPR
768.72
769.05
769.39
769.72
785.43
785.76
786.10
786.43
relative abundance
m/z


I
Heterogeneous nuclear ribonucleoprotein U-like protein 2
unstimulated
stimulated
AYGQQYWGQPGNR10GGYR10
AYGQQYWGQPGNR10GGYR10
AYGQQYWGQPGNRGGYR
z=3
669.32
672.00
672.33
669.65
662.65
662.98
669.98
672.67
670.31
673.00
relative abundance
m/z
J
60S ribosomal protein L36a-like
unstimulated
stimulated
QSGYGGQTKPIFR10
QSGYGGQTKPIFR10
QSGYGGQTKPIFR
z=2
731.88
733.89
732.38
734.39
726.88
727.38
727.88
732.88
734.89
relative abundance
m/z
K
Splicing factor 3B subunit 2
unstimulated
stimulated
VAAPVGPVGPTPTVLPMGAPVPR10PR10
VAAPVGPVGPTPTVLPMGAPVPRPR
VAAPVGPVGPTPTVLPMGAPVPR10PR10
z=3
821.48
828.14
827.81
821.14
821.81
828.48
832.17
831.84
832.50
relative abundance
m/z
L
Small nuclear ribonucleoprotein-associated proteins B and B'
unstimulated
stimulated
VPLAGAAGGPGIGRAAGRGIPAGVPMPQAPAGLAGPVR
VPLAGAAGGPGIGR10AAGR10GIPAGVPMPQAPAGLAGPVR10
VPLAGAAGGPGIGR10AAGR10GIPAGVPMPQAPAGLAGPVR10
z=3
1158.00
1158.33
1147.65
1147.98
1148.32
1157.66
1158.67
1164.70
1165.04
1148.65
1164.37
1165.37
relative abundance
m/z


M
Small nuclear ribonucleoprotein-associated protein N
unstimulated
stimulated
VPLAGAAGGPGVGRAAGRGVPAGVPIPQAPAGLAGPVR
1132.65
1132.99
z=3
1133.32
1132.32
1132.65
VPLAGAAGGPGVGR10AAGR10GVPAGVPIPQAPAGLAGPVR10
13CD3 13CD3 13CD3 13CD3
VPLAGAAGGPGVGR10AAGR10GVPAGVPIPQAPAGLAGPVR10
1142.67
1143.00
1142.33
1143.33
1147.69
1148.02
1148.36
1148.69
relative abundance
m/z
N
TATA-binding protein associated factor 2N
unstimulated
stimulated
GPMTGSSGGDR10GGFK
724.84
z=2
GPMTGSSGGDRGGFK
719.84
720.34
720.84
725.34
725.84
GPMTGSSGGDR10GGFK
730.87
731.37
731.87
relative abundance
m/z
O
Protein TFG
unstimulated
stimulated
NRPPFGQGYTQPGPGYR
960.48
960.98
z=2
961.48
NR10PPFGQGYTQPGPGYR10
970.49
970.99
974.51
975.01
971.49
975.51
NR10PPFGQGYTQPGPGYR10
relative abundance
m/z
P
WAS/WASL interacting protein family member 1
unstimulated
stimulated
GSPPVPGGPRQPSPGPTPPPFPGNR
826.09
z=3
825.76
826.43
826.76
GSPPVPGGPR10QPSPGPTPPPFPGNR10
832.76
832.43
833.09
833.43
835.44
835.10
835.77
836.11
GSPPVPGGPR10QPSPGPTPPPFPGNR10
relative abundance
m/z

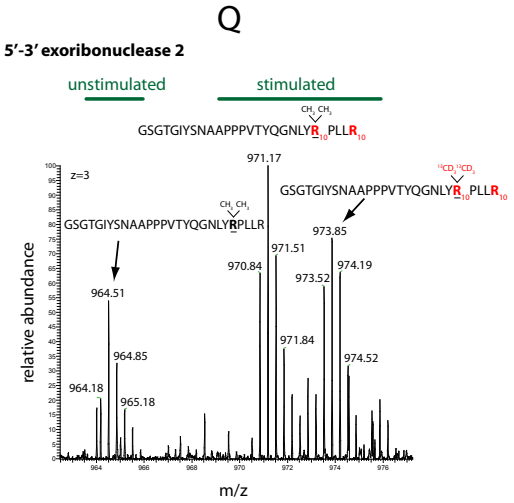


Figure 4.28: Panels A-Q, changes in arginine and lysine methylation upon activation of primary human CD4+ T-lymphocytes measured by mass spectrometry. CD4+ T-lymphocytes were isolated from whole blood and a portion were harvested as the unstimulated, unlabelled sample. The remainder were stimulated and labelled with Arg10 and either light or heavy methionine according to the labelling method described in Figure 4.27. Proteins were digested with trypsin and methylated peptides enriched by IEF. Each panel shows a protein methylation site from unstimulated and stimulated cells and the corresponding peptide. Precursor spectra were manually verified and intensity of each peptide was summed over the elution time for relative quantitation.

amounts of material and is the most reliable way to identify methylation in WIPF1, a protein important to T-lymphocyte activation (Le Bras et al., 2009). This permitted the relative quantification of 21 methylation sites in 20 proteins, including 2 lysine methylation sites (Figure 4.28). Peptide intensities were summed across elution profiles for methylated peptides from unstimulated and stimulated cells. The ratio of these intensities gave the stimulated/unstimulated peptide ratio (Table 4.4). Where a non-methylated peptide from the same protein was identified, the stimulated/unstimulated protein ratio was similarly calculated. The change in proportion of protein methylated was then calculated (dividing the change in methylation by the change in protein level) (Table 4.4, final column).

Large (>2 fold) changes in methylation were observed for 8 of the 21 methylation sites quantified. For 7 of the methylation sites the change in proportion of methylated protein could be calculated, of these, 5 proteins showed large increases in the fraction existing in the methylated form. In almost all spectra, methylation was observed to be present in unstimulated cells. The only exception was for a methylation site in G3BP1 which was not detected in unstimulated cells (Figure 4.28, D). Non-methylated, arginine containing peptides were identified from G3BP1 which allowed the relative total protein levels in unstimulated and stimulated cells to be measured. This showed that the protein was present in unstimulated cells and its abundance only increased slightly after stimulation. Consequently, any large changes in methylation also represent large changes in the proportion of G3BP1 methylated on arginine. By contrast, measuring levels of lysine methylation on Histone H4 pointed to a ∼7-fold decrease upon stimulation. However, mea-

suring relative Histone H4 protein levels revealed a similar decrease indicating that the proportion of Histone H4 methylated at this site does not change dramatically (data not shown). Arginine methylation of WIPF1 decreased approximately 2-fold after activation of T-lymphocytes. Total WIPF1 protein increased by almost 50%, thus the fold change in proportion of WIPF1 methylated at this site was 0.34.

Accession no.	Protein name, *gene name*	Peptide *(methylation)*	Stimulated/unstimulated methylated peptide ratio	Change in proportion of methylated protein upon stimulation
P50238	CRP-1, *CRIP1* (Cysteine-rich protein 1)	GFG**R**GGAESHTFK *(MMA)*	0.25	unknown
Q01844	RNA-binding protein EWS, *EWSR1*	GMPPPL**R**GGPGGPGGPGGPMGR *(DMA)*	1.34	unknown
Q13283	G3BP-1, *G3BP1* (Ras GTPase-activating protein-binding protein 1)	GPGGP**R**GGLGGGMR *(DMA)*	1.90	1.64
		GGMVQKPGFGVG**R**GLAPR *(DMA)*	50	43.1
P62805	Histone H4, *HIST1H4E*	**K**VLRDNIQGITKPAIR *(DML)*	0.16	0.58
Q13151	hnRNP A0, *HNRNPA0* (Heterogeneous nuclear ribonucleoprotein A0)	SNSGPY**R**GGYGGGGGYGGSSF *(DMA)*	1.05	unknown
P22626	hnRNP A2/B1, *HNRNPA2B1* (Heterogeneous nuclear ribonucleoproteins A2/B1)	GGNFGFGDS**R**GGGGNFGPGPGSNFR *(DMA)*	0.35	unknown
Q00839	hnRNP U, *HNRNPU* (Heterogeneous nuclear ribonucleoprotein U)	**R**GNMPQ**R**GGGGGGSGGIGYPYPR *(DMA,DMA)*	2.97	2.22
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2, *HNRNPUL2*	AYGQQYWGQPGN**R**GGY**R** *(MMA,MMA)*	2.63	2.68
Q969Q0	60S ribosomal protein L36a-like, *RPL36AL*	QSGYGGQT**K**PIFR *(MML)*	3.75	5.62
Q13435	Splicing factor 3B subunit 2, *SF3B2*	VAAPVGPVGPTPTVLPMGAPVP**R**PR *(DMA)*	0.85	unknown
P14678	Small nuclear ribonucleoprotein-associated proteins B and B', *SNRPB*	VPLAGAAGGPGIG**R**AAG**R**GIPAGVPMPQAPAGLAGPVR *(DMA,DMA)*	1.16	unknown
P63162	snRNP-N, *SNRPN* (Small nuclear ribonucleoprotein-associated protein N)	VPLAGAAGGPGVG**R**AAG**R**GVPAGVPIPQAPAGLAGPVR *(DMA,DMA)*	0.17	unknown
Q92804	TATA-binding protein-associated factor 2N, *TAF15*	GPMTGSSGGD**R**GGFK *(DMA)*	1.13	unknown
Q92734	Protein TFG, *TFG*	N**R**PPFGQGYTQPGPGYR *(DMA)*	0.87	unknown
O43516	WAS\WASL-interacting protein family member 1, *WIPF1*	GSPPVPGGP**R**QPSPGPTPPPFPGNR *(DMA)*	0.53	0.34
Q9H0D6	5'-3' exoribonuclease 2, *XRN2*	GSGTGIYSNAAPPPVTYQGNLY**R**PLLR *(DMA)*	1.67	unknown

Table 4.4: Summary of changes in arginine and lysine methylation measured upon activation of primary human CD4+ T-lymphocytes.

Methylation sites identified

Accession no.	Protein name, *gene name*	Peptide	Methylation site	Methylation	Method	Novel?
Chromatin structure and regulation						
Q9ULX6	AKAP8-like protein, *AKAP8L* (A-kinase anchor protein 8-like)	LPSLFSQNIIPEYGMFQGM**R**GGGAFPGGSR	R237	DMA	HILIC	Yes
O14497	ARID domain-containing protein 1A, *ARID1A* (AT-rich interactive domain-containing protein 1A)	AAGPGLGNVAMGP**R**QHYPYGGPYDR	R1276	DMA	HILIC	Yes
P62805	Histone H4, *HIST1H4E*	**K**VLRDNIQGITKPAIR	K20	DML	HILIC, IEF, SCX	
P84228	Histone H3.2, *HIST2H3C*	**K**SAPATGGVK	K27	DML	HILIC	
Q96KQ7	Histone-lysine N-methyltransferase EHMT2, *EHMT2*	ILLGHAT**K**SFPSSPSK	K114	DML	HILIC	Yes
O14686	Histone-lysine N-methyltransferase MLL2, *MLL2*	NPQL**R**HLSPQQQQQLQALLMQR *(TU)*	R4212	DMA	HILIC	Yes
Q7Z3K3	Pogo transposable element with ZNF domain, *POGZ*	NVQQGQTV**R**PITLVPAPGTQF	R183	DMA	HILIC	Yes
Q09472	p300 HAT, *EP300* (Histone acetyltransferase p300)	TL**R**SPSSPLQQQQVLSILHANPQLLAAFIK	R2059	DMA	HILIC	Yes
Q96GM5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1, *SMARCD1*	MTPQGPSMGPPGYGGNPSV**R**PGLAQSGMDQSR	R88	DMA	HILIC	Yes
P0C1Z6	TCF3 fusion partner, *TFPT*	ERDEEEEAA**R**GR *(TU)*	R68	DMA	HILIC	Yes
Q8TAP9	TTD non-photosensitive 1 protein, *TTDN1*	**R**PPTPPYPGPGGGGWGSGSSF	R7	DMA	HILIC	Yes
		GTPGGGGP**R**PPSPR	R37	DMA	HILIC	Yes
		HGGSFPGG**R**FGSPSPGGYPGSYSR	R77	DMA	HILIC	Yes
		SPAGSQQQFGYSPGQQQTHPQGSP**R**TSTPFGSGR	R117	DMA	HILIC, WCX	Yes
Cytoskeleton dynamics						

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

P62158	CaM, *CALM1;CALM2;CALM3* (Calmodulin)	HVMTNLGE**K**LTDEEVDEMIR	K116	TML	HILIC	
P50570	Dynamin-2, *DNM2*	RPPAAPSRPTII**R**PAEPSLLD *(TU)*	R862	DMA	HILIC	Yes
		II**R**PAEPSLLD	R862	MMA	HILIC	Yes
Q15654	TRIP-6, *TRIP6* (Thyroid receptor-interacting protein 6)	**R**GASQASGPLPGPHFPLPGR *(TU)*	R186	DMA	HILIC	Yes
P42768	WASp, *WAS* (WAS Wiskott-Aldrich syndrome protein)	G**R**SGPLPPVPLGIAPPPPTPR	R338	MMA	HILIC	Yes
		SGPLPPVPLGIAPPPPTP**R**GPPPPGR	R357	DMA	HILIC	Yes
O43516	WAS\WASL-interacting protein family member 1, *WIPF1*	TEQAG**R**NALLSDISK	R33	DMA	HILIC	Yes
		STSAKPFSPPSGPG**R**FPVPSPGHR	R149	DMA	HILIC, IEF, SCX	Yes
		GSPPVPGGP**R**QPSPGPTPPPFPGN**R**GTALGGGSIR	R211, R226	DMA, MMA	HILIC, IEF	Yes
Endosomal trafficking						
Q14677	Clathrin interactor 1, *CLINT1*	SSPSNMLPV**R**PQ	R538	DMA	HILIC	Yes
Q9NZ52	ADP-ribosylation factor-binding protein GGA3, *GGA3*	SGLVKPTTSPLIPTTTPA**R**PLLPF	R548	DMA	HILIC	Yes
Q13492	Phosphatidylinositol-binding clathrin assembly protein, *PICALM*	SQPVM**R**PPNPFGPVSGAQIQFM	R636	DMA	HILIC	Yes
P53992	Protein transport protein Sec24C, *SEC24C*	TPPASGGP**R**LPSMTGPLLPGQSF	R222	DMA	HILIC	Yes
O94979	Protein transport protein Sec31A, *SEC31A*	**R**PQQPVAPPTSNAYPNTPY	R894	DMA	HILIC	Yes
O60493	Sorting nexin-3, *SNX3*	LITKPQNLNDAYGPPSNFLEIDVSNPQTVGVG**R**GR	R43	DMA	HILIC	Yes
RNA processing						
P38432	Coilin, *COIL*	TVGLFAG**R***G**R***PGPGLSSQTAGAAGWR	R350 or R352	DMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

Q16630	Cleavage and polyadenylation specificity factor subunit 6, *CPSF6*	FPGAVPGGD**R**FPGPAGPGGPPPPFPGNLIK	R216	DMA	WCX	Yes
P33240	Cleavage stimulation factor subunit 2, *CSTF2*	GMDT**R**GPVPGP**R**GPIPSGMQGPSPINMGAVVPQGSR	R468 + R475	MMA, DMA	HILIC	Yes
		GPVPGP**R**GPIPSGMQGPSPINMGAVVPQGSR	R475	DMA	HILIC, WCX	Yes
P17844	Probable ATP-dependent RNA helicase DDX5, *DDX5*	**R**GGFNTFR	R502	DMA	DMA-IP, HILIC, IEF, SCX	
		RGGFNTFR	R502	MMA	HILIC, SCX	
Q08211	ATP-dependent RNA helicase A , *DHX9*	YGDGP**R**PPK *(TU)*	R1160	DMA	IEF, HILIC	Yes
		GSGGFQ**R**GGG**R**GAYGTGYFGQG**R**GGGGY *(TU)*	R1249, R1253, R1265	DMA, DMA, DMA	HILIC	Yes
P55265	DRADA, *ADAR* (Double-stranded RNA-specific adenosine deaminase, isoform 2)	QL**R**YQQPGPGSSPSSFLLK	R26	DMA	HILIC, IEF	Yes
A8MUW5	*FAM98B*	GGGGGGGGGWGGGGGGG**R**GGF	R384	DMA	HILIC	Yes
		GDYGG**R**GGYGG**R**GGYGGR *(TU)*	R396, R402	DMA, DMA	HILIC	Yes
P35637	RNA-binding protein FUS, *FUS*	GG**R**G**R**GGSGGGGGGGGGGYNR	R216, R218	DMA, DMA	DMA-IP, HILIC, SCX	
Q13151	hnRNP A0, *HNRNPA0* (Heterogeneous nuclear ribonucleoprotein A0)	SNSGPY**R**GGYGGGGGYGGSSF	R291	DMA	DMA-IP, HILIC, IEF, WCX	
P09651	hnRNP A1, *HNRNPA1* (Heterogeneous nuclear ribonucleoprotein A1)	ALSKQEMASASSSQ**R**GR	R194	DMA	DMA-IP, HILIC, IEF	
		SGSGNFGGG**R**GGGFGGNDNFG**R**GGNFSG**R**GGFGGSR	R206, R218, R225	DMA, DMA, DMA	DMA-IP, HILIC, SCX	Yes(218)
		GGS**R**GGGGYGGSGDGYNGF	R232	DMA	HILIC	Yes
		GGS**R**GGGGYGGSGDGYNGF	R232	MMA	HILIC	Yes
		NQGGYGGSSSSSSYGSG**R**	R370	DMA	HILIC	Yes
P22626	hnRNP A2/B1, *HNRNPA2B1* (Heterogeneous nuclear ribonucleoproteins A2/B1)	SG**R**GGNFGFGDS**R**GGGGNFGPGPGSNFR	R203, R213	DMA, MMA	HILIC, IEF, WCX	
		SG**R**GGNFGFGDS**R**GGGGNFGPGPGSNFR	R203, R213	DMA, DMA	HILIC, IEF	
		SG**R**GGNFGFGDS**R**GGGGNFGPGPGSNFR	R203, R213	MMA, DMA	HILIC, IEF	

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		GGG**R**GGYGGGGPGYGNQGGGY	R266	DMA	HILIC	Yes
P51991	hnRNP A3, *HNRNPA3* (Heterogeneous nuclear ribonucleoprotein A3)	MG**R**GGNFGGGGGNF	R226	DMA	HILIC	Yes
		GGGGGNFG**R**GGNFGG**R**GGY	R239, R246	DMA, DMA	HILIC	Yes(239)
		SS**R**GGYGGGGPGYGNQGGGYGGGGGYDGY	R286	DMA	HILIC	Yes
Q99729	hnRNP A/B, *HNRNPAB* (Heterogeneous nuclear ribonucleoprotein A/B, isoform 4)	EVYQQQQYGSGG**R**GNR	R245	DMA	DMA-IP, HILIC, WCX	
		RGGHQNNYKPY	R322	DMA	DMA-IP, IEF, WCX	
Q14103	hnRNP D0, *HNRNPD* (Heterogeneous nuclear ribonucleoprotein D0)	EQYQQQQQWGS**R**GGFAG**R**	R272, R278	MMA, DMA	HILIC, WCX	Yes
		EQYQQQQQWGS**R**GGFAG**R**A**R**	R272, R278, R280	DMA, DMA, DMA	HILIC	Yes
		G**R**GGDQQSGYGK	R282	DMA	HILIC	Yes
		G**R**GGDQQSGYGK	R282	MMA	HILIC	Yes
		RGGHQNSYKPY	R345	DMA	DMA-IP, IEF, WCX	
P38159	hnRNP G, *RBMX* (Heterogeneous nuclear ribonucleoprotein G)	GPPS**R**GGHMDDGGYSMNFNMSSSR	R125	DMA	HILIC	Yes
		SAPSGPV**R**SSSGMGGR	R172	MMA	HILIC	Yes
P31943	hnRNP H, *HNRNPH1* (Heterogeneous nuclear ribonucleoprotein H)	LMAMQRPGPYDRPGAG**R**GYNSIG**R**GAGFER	R217, R224	DMA, DMA	SCX	Yes
		RGAYGGGYGGYDDYNGYNDGYGFGSDR	R233	DMA	SCX	
P55795	hnRNP H2, *HNRNPH2* (Heterogeneous nuclear ribonucleoprotein H2)	**R**GAYGGGYGGYDDYGGYNDGYGFGSDRFGR *(TU)*	R233	DMA	SCX	Yes
P31942	hnRNP H3, *HNRNPH3* (Heterogeneous nuclear ribonucleoprotein H3)	LLGQRPGPYDRPIGG**R**GGYYGAGR *(TU)*	R121	DMA	HILIC	
		GAG**R**GSYGGFDDY	R129	DMA	HILIC	Yes
		MGMGNNYSGGYGTPDGLGGYG**R**GGGGSGGYYGQGGMSGGGWR	R323	DMA	HILIC	Yes
		MGMGNNYSGGYGTPDGLGGYG**R**GGGGSGGYYGQGGMSGGGWR	R323	MMA	HILIC	Yes
P61978	hnRNP K, *HNRNPK* (Heterogeneous nuclear ribonucleoprotein K)	GRPVGFPM**R**GR	R256	DMA	HILIC	Yes
		MPPG**R**GGRPMPPSR *(TU)*	R268	DMA	SCX	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		RGPPPPPPG**R**GG**R**GGSR *(TU)*	R296, R299	DMA, DMA	SCX	
		RGPPPPPPG**R**GG**R**GGS**R** *(TU)*	R296, R299, R303	DMA, MMA, MMA	SCX	Yes(303)
		A**R**NLPLPPPPPPR	R305	MMA	HILIC	Yes
Q00839	hnRNP U, *HNRNPU* (Heterogeneous nuclear ribonucleoprotein U)	**R**GNMPQ**R**GGGGGGSGGIGYPYPR	R733, R739	DMA, DMA	DMA-IP, HILIC, IEF, WCX	
Q9BUJ2	Heterogeneous nuclear ribonucleoprotein U-like protein 1, *HNRNPUL1*	GGGGF**R**G**R**GGGGGFQR *(TU)*	R618, R620	DMA, DMA	HILIC	Yes
O14979	hnRNP DL, *HNRPDL* (Heterogeneous nuclear ribonucleoprotein D-like)	AS**R**GGGNHQNNYQPY	R408	DMA	DMA-IP, HILIC, IEF, WCX	
Q69YN4	Protein virilizer homolog, *KIAA1429*	GNYNES**R**GGQSNFNR *(TU)*	R1741	DMA	HILIC	Yes
Q92615	La-related protein 4B	VEPGSLESSPGLG**R**GR *(TU)*	R493	DMA	HILIC	Yes
Q86TB9	Pat1b, *PATL1* (Protein PAT1 homolog 1)	MSPSQFA**R**VPGFVGSPLAAMNPK	R284	DMA	HILIC, IEF	Yes
O94913	Pre-mRNA cleavage complex 2 protein Pcf11, *PCF11*	IEGPLGQGGP**R**FEGCHALR	R1030	DMA	HILIC	Yes
Q9H307	Pinin, *PNN*	LLALSGPGGG**R**G**R**GSLLLR	R54, R56	DMA, MMA	HILIC	Yes
O94906	Pre-mRNA-processing factor 6, *PRPF6*	KKPFLGMPAPLGYVPGLG**R**GATGFTTR	R23	DMA	SCX	Yes
P48634	Protein PRRC2A, *PRRC2A*	VFTP**R**GVPSR	R1164	DMA	WCX+HILIC	Yes
Q9P2N5	RNA-binding protein 27, *RBM27*	DQPGTSAVPNLASVGT**R**LPPPLPQNLLYTVSER	R411	DMA	HILIC	Yes
		LGILPVG**R**GK	R943	DMA	HILIC	Yes
Q96EV2	RNA-binding protein 33, *RBM33*	SQPP**R**LPLQDQW	R442	DMA	HILIC	Yes
		DPFFLGVSGEP**R**FPSHLFLEQR	R470	DMA	HILIC	Yes
Q14498	RNA-binding protein 39, *RBM39*	FNSAI**R**GK *(TU)*	R109	DMA	HILIC	Yes
Q9BTD8	RNA-binding protein 42, *RBM42*	APIL**R**PAFVPHVLQR	R158	DMA	HILIC, IEF	Yes
		AL**R**PPHQALVGPPLPGPPGPPM	R186	DMA	HILIC	Yes
Q15020	SART-3, *SART3*	KAPGGPMLLPQTYGA**R**GK	R906	DMA	HILIC, IEF	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

	(Squamous cell carcinoma antigen recognized by T-cells 3)					
Q15459	Splicing factor 3A subunit 1, *SF3A1*	IGPSKPNEIPQQPPPPSSATNIPSSAPPITSVP**R**PPTMPPPVR	R578	DMA	WCX	Yes
P14678	Small nuclear ribonucleoprotein-associated proteins B and B', *SNRPB*	VPLAGAAGGPGIG**R**AAG**R**GIPAGVPMPQAPAGLAGPVR	R108, R112	DMA, DMA	HILIC, IEF, WCX	
		VPLAGAAGGPGIG**R**AAG**R**GIPAGVPMPQAPAGLAGPVR	R108, R112	MMA, DMA	HILIC, IEF	
		VPLAGAAGGPGIG**R**AAG**R**GIPAGVPMPQAPAGLAGPVR	R108, R112	DMA, MMA	HILIC, IEF	
		TPQG**R**GTVAAAAAAATASI	R147	DMA	HILIC	
		GPPPPGM**R**PPRP	R236	DMA	HILIC, IEF	Yes
P62318	Sm-D3, *SNRPD3* (SNRPD3 Small nuclear ribonucleoprotein Sm D3)	NQGSGAG**R**GK	R97	DMA	HILIC, WCX	
P63162	snRNP-N, *SNRPN* (Small nuclear ribonucleoprotein-associated protein N)	VPLAGAAGGPGVG**R**AAG**R**GVPAGVPIPQAPAGLAGPVR	R108, R112	DMA, DMA	HILIC, IEF, WCX	Yes
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein, *SERBP1*	GEGGEFSVDRPIIDRPI**R**GR *(TU)*	R163	DMA	HILIC	Yes
P18583	Protein SON, *SON*	LAP**R**PLMLASR	R989	DMA	HILIC	Yes
Q15637	Splicing factor 1, *SF1*	ASAP**R**PAAPANNPPPPSLM	R345	DMA	HILIC	Yes
		ASAP**R**PAAPANNPPPPSLM	R345	MMA	HILIC	Yes
		STTQS**R**PPWMNSGPSESRPY	R365	DMA	HILIC	Yes
Q13435	Splicing factor 3B subunit 2, *SF3B2*	TPLGP**R**VAAPVGPVGPTPTVLPMGAPVP**R**PR	R222, R245	DMA, DMA	HILIC, IEF, WCX	Yes
		NSVPVP**R**HWCFK	R502	DMA	HILIC, SCX	Yes
Q15427	Splicing factor 3B subunit 4, *SF3B4*	TGPP**R**PPPYGY	R390	DMA	HILIC	Yes
P23246	Splicing factor, proline- and glutamine-rich, *SFPQ*	F**R**S**R**GGGGGGFHR	R7, R9	DMA, DMA	DMA-IP, IEF	
		F**R**S**R**GGGGGGFHR	R7, R9	DMA, MMA	DMA-IP	
		RGGGGG**R**GGLHDFR	R19, R25	DMA, DMA	WCX	
		GQGGAGPVGGQGP**R**GMGPGTPAGY	R681	MMA	HILIC	
Q07955	Serine/arginine-rich splicing factor 1, *SRSF1*	SG**R**GTG**R**GGGGGGGGGAPR	R93, R97	DMA, DMA	HILIC	
		GTG**R**GGGGGGGGGAPR	R97	DMA	HILIC, IEF	

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		GTG**R**GGGGGGGGAPR	R97	MMA	HILIC, IEF	
		GTG**R**GGGGGGGGAP**R** *(TU)*	R97, R109	DMA, MMA	IEF	
O95104	Splicing factor, arginine/serine-rich 15, *SCAF4*	**R**GPPPGPGGF	R888	DMA	HILIC	Yes
Q9BXP5	Serrate RNA effector molecule homolog, *SRRT*	GNYDAF**R**GQGGYPGKPR	R840	DMA	HILIC, IEF, SCX	Yes
Q96SI9	Spermatid perinuclear RNA-binding protein, *STRBP*	GVVNTAVSAAVQAV**R**GR	R608	DMA	HILIC	Yes
Q86V81	Tho4, *THOC4* (THO complex subunit 4)	AGSQGG**R**GGGAQAAAR	R38	DMA	HILIC, IEF	Yes
		VN**R**GGGPIR	R50	DMA	DMA-IP, HILIC, WCX	
		N**R**PAIA**R**GAAGGGGR	R58, R63	MMA, DMA	DMA-IP, SCX	Yes
		N**R**PAIA**R**GAAGGGGR	R58, R63	DMA, DMA	SCX, WCX	Yes
		RPAQSVN**R**GGMTR	R197	DMA	DMA-IP	Yes
		N**R**GAGGFGGGGGTR	R204	DMA	DMA-IP, HILIC, IEF, WCX	
P62995	TRA-2 beta, *TRA2B* (Transformer-2 protein homolog beta)	SY**R**GGGGGGGGWR	R241	DMA	DMA-IP, HILIC, IEF, SCX, WCX	
P09234	U1 snRNP C, *SNRPC* (U1 small nuclear ribonucleoprotein C)	SAPPPAGAMIPPPPSLPGPP**R**PGMMPAPH	R88	DMA	HILIC	Yes
Q9Y2W2	WBP-11, *WBP11* (WW domain-binding protein 11)	**R**PPLVPPLGPAPPGLFPPAPLPNPGVL	R508	DMA	HILIC	Yes
Q9H0D6	5'-3' exoribonuclease 2, *XRN2*	TLGHVMP**R**GSGTGIYSNAAPPPVTYQGNLY**R**PLLR	R824, R847	DMA, DMA	HILIC, IEF	Yes
		GSGTGIYSNAAPPPVTYQGNLY**R**PLL**R**GQAQIPK	R847, R851	DMA, DMA	HILIC, IEF	Yes
		LMSNMRPQDSW**R**GPPPLFQQQR	R870	DMA	HILIC	Yes
RNA export						
Q07666	KH domain-containing, RNA-binding, signal transduction-associated protein 1, *KHDRBS1*	GVPV**R**G**R**GAAPPPPPVP**R** *(TU)*	R289, R291, R302	DMA, DMA, DMA	IEF	Yes(R288 + R302)

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		GATVT**R**GVPPPPTVR	R331	DMA	HILIC	
		GATVT**R**GVPPPPTV**R**GAPAPR	R331, R340	DMA, DMA	WCX+HILIC	
Q9NZB2	Constitutive coactivator of PPAR-gamma-like protein 1, *FAM120A*	HFGPVPPSQG**R**G**R**GFAGVCGFGGPYGETVATGPYR	R884, R886	DMA, MMA	SCX	Yes
		GPFPLQVVSVGGPA**R**GRPR	R982	DMA	HILIC	
		YYG**R**GYKNQAAIQGRPPYAASAEEVAK *(TU)*	R1006	DMA	SCX	Yes
P49792	E3 SUMO-protein ligase RanBP2, *RANBP2*	TNFVQPMPGEGL**R**PSLPTQAHTTQPTPFK	R1016	DMA	HILIC	Yes
P12270	Nucleoprotein TPR, *TPR*	QSVG**R**GLQLTPGIGGMQQHFFDDEDR	R2111	DMA	HILIC	Yes
Transcription						
Q9NYF8	Btf, *BCLAF1* (Bcl-2-associated transcription factor 1)	A**R**GVFAGTNTGPNNSNTTFQK *(TU)*	R809	DMA	HILIC	Yes
Q6PJG2	Uncharacterized protein C14orf43, *C14orf43*	VQLEVG**R**PQAPLNSFHAAK	R193	DMA	HILIC	Yes
Q9Y3Y2	Chromatin target of PRMT1 protein, *CHTOP*	LGRPIGALA**R**GAIGGR *(TU)*	R97	DMA	HILIC	Yes
P62633	CNBP, *CNBP* (Cellular nucleic acid-binding protein, isoform 2)	S**R**G**R**GFQFVSSSLPDICYR	R32, R34	DMA, MMA	HILIC, IEF, SCX	Yes
		S**R**G**R**GFQFVSSSLPDICYR	R32, R34	DMA, DMA	HILIC, IEF, SCX	Yes
Q9BTC0	DIO-1, *DIDO1* (Death-inducer obliterator 1)	FTNQ**R**APAPLQFGGLR	R1982	DMA	HILIC	Yes
Q01844	RNA-binding protein EWS, *EWSR1*	EG**R**GMPPPL**R**GGPGGPGGPGGPMGR	R410, R417	MMA, DMA	DMA-IP, HILIC, IEF, WCX	
		GGD**R**GGFPP**R**GP**R**	R494, R500, R503	DMA, MMA, MMA	DMA-IP	
		GS**R**GNPSGGGNVQHR	R506	DMA	SCX	
		D**R**GGPGGMF	R581	DMA	HILIC	
		RGGPGGPPGPLMEQMGG**R**	R615, R632	DMA, MMA	DMA-IP, HILIC, IEF	
Q96AE4	FBP, *FUBP1* (Far upstream element-binding protein 1)	G**R**G**R**GQGNWNMGPPGGLQEFNFIVPTGK	R361, R363	DMA, MMA	HILIC	Yes
Q92945	FUSE-binding protein 2,	G**R**GQGNWGPPGGEMTFSIPTHK	R415	MMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

	KHSRP (Far upstream element-binding protein 2)					
Q86YP4	Hp66alpha, *GATAD2A* (Transcriptional repressor p66-alpha)	QHSSTGPPPLLLAP**R**ASVPSVQIQGQR	R273	DMA	HILIC, WCX	Yes
Q8WXI9	Transcriptional repressor p66-beta, *GATAD2B*	GPPKPGLV**R**TTTPNMNPAINYQPQSSSSVPCQR	R286	DMA	HILIC	Yes
Q13227	GPS-2, *GPS2* (G protein pathway suppressor 2)	AATSQPGP**R**LPF	R312	DMA	HILIC	Yes
P51610	HCF-1, *HCFC1* (HCFC1 of Host cell factor 1)	VTGPQATTGTPLVTM**R**PASQAGK	R504	DMA	HILIC	Yes
		TVPMGGV**R**LVTPVTVSAVKPAVTTLVVK	R855	DMA	IEF	Yes
		V**R**LSSPSIK	R1219	DMA	HILIC	Yes
Q12906	Interleukin enhancer-binding factor 3, *ILF3*	RAPVPV**R**GGPK	R609	DMA	DMA-IP, IEF, WCX	
Q93074	Mediator of RNA polymerase II transcription subunit 12, *MED12*	TSVY**R**QQQPAVPQGQR *(TU)*	R1899	DMA	HILIC	Yes
Q14686	Nuclear receptor coactivator 6, *NCOA6*	SLQGGPS**R**VPTPLQQPHLTNK	R431	DMA	WCX+HILIC	Yes
Q92570	Nuclear receptor subfamily 4 group A member 3, *NR4A3*	VAGA**R**FPLF	R187	DMA	HILIC	Yes
Q8WXF1	Paraspeckle component 1, *PSPC1*	TGSETPQAPMSGVGPVSGGPGGFG**R**GSQGGNFEGPNKR	R507	MMA	HILIC	Yes
P48382	DNA-binding protein RFX5, *RFX5*	SVVESSAPGANNLQVNALVA**R**LPLLLPR	R328	DMA	HILIC	Yes
		SLIPPIPVSPPILAP**R**LSSGALK	R354	DMA	HILIC	Yes
Q01196	Runt-related transcription factor 1, *RUNX1*	SDPRQFPALPSISDP**R**MHYPGAF	R319	DMA	HILIC	Yes
IPI00646058	Scaffold attachment factor B, *SAFB*	GLPPPP**R**GR	R818	DMA	HILIC	Yes
O75182	Paired amphipathic helix protein Sin3b, *SIN3B*	**K***LS**R***MAPEDQEK *(TU)*	K514 or R517	DML or DMA	SCX	Yes
Q9NWH9	SAFB-like transcription modulator, *SLTM*	IVQISGNSMP**R**GSGSGFKPFK	R1017	DMA	WCX+HILIC	Yes
Q02446	Transcription factor Sp4, *SP4*	I**R**APTLTPSGQISW	R457	DMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing,
SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

Q9BWW4	Single-stranded DNA-binding protein 3, *SSBP3*	YAGGP**R**PPI**R**MGNQPPGGVPGTQPLLPNSMDPTR	R161, R165	DMA, DMA	HILIC	Yes
O00267	hSPT5, *SUPT5H* (Transcription elongation factor SPT5)	ISSPMHPSAGGQ**R**GGFGSPGGGSGGMSR *(TU)*	R681	DMA	HILIC	
O00268	Transcription initiation factor TFIID subunit 4, *TAF4*	AP**R**LPQPPQNPTNIQNF	R435	DMA	HILIC	Yes
Q92804	TATA-binding protein-associated factor 2N, *TAF15*	GPMTGSSGGD**R**GGFK	R206	DMA	DMA-IP, HILIC, IEF, SCX, WCX	
		SGGGYGGD**R**GGGYGGD**R**GGGYGGDR	R459, R467	MMA, DMA	HILIC	Yes
		GGGYGGD**R**GGGYGGD**R**GGYGGDR	R475, R483,	MMA, DMA	HILIC	Yes(R475)
		GGGYGGD**R**GGGYGGD**R**GGYGGDR	R475, R483,	DMA, DMA	HILIC	Yes(R475)
		S**R**GGYGGD**R**GGGSGYGGDR	R528, R535	DMA, DMA	DMA-IP, SCX, WCX	Yes
		SGGGYGGD**R**GGGYGGD**R**GGYGGK	R561, R570	DMA, DMA	DMA-IP, HILIC, IEF, SCX	Yes(R561)
		SGGGYGGD**R**GGGYGGD**R**GGYGGK	R561, R570	MMA, DMA	HILIC, IEF	Yes(R561)
Q9Y2W1	Thyroid hormone receptor-associated protein 3, *THRAP3*	N**R**GFYPWGQYN**R**GGYGNYR	R91, R101	DMA, DMA	HILIC, SCX	Yes
		DSRPSQAAGDNQGDEA**K**EQTFSGGTSQDTK	K202	DML	HILIC	Yes
Q9UPN9	E3 ubiquitin-protein ligase TRIM33, *TRIM33*	TPGQINLAQL**R**LQHMQQQVYAQK	R115	DMA	HILIC	Yes
P49750	YLP motif-containing protein 1, *YLPM1*	GP**R**FDGPR	R563	DMA	WCX+HILIC	Yes
		GPASQFYITPSTSLSP**R**QSGPQWK	R636	DMA	HILIC	Yes
		RGGFDRPSYDR	R1563	DMA	HILIC	Yes
		K*SD**R***PVYEGPSMFGGER	K1574 or R1577	DML or DMA	HILIC, IEF	Yes
Q9ULJ6	Zinc finger MIZ domain-containing protein 1, *ZMIZ1*	GPNSQFPTQPGQYPAPNPP**R**PL	R438	DMA	HILIC	Yes
Q5BKZ1	Zinc finger protein 326, *ZNF326*	G**R**GTPAYPESTFGSR	R175	DMA	HILIC	Yes
		NYDAFGGPSTG**R**G**R**GR *(TU)*	R200, R202	DMA, DMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		STNVTVAAA**R**GIK	R235	DMA	HILIC, IEF	Yes
Translation						
Q14444	Caprin-1, *CAPRIN1*	GPANGF**R**GGYDGYRPSFSNTPNSGYTQSQFSAPR	R633	DMA	HILIC, IEF, WCX	Yes
		RPSFSNTPNSGY	R640	DMA	HILIC	Yes
P68104	EF-1-alpha-1, *EEF1A1* (Elongation factor 1-alpha 1)	GITIDISLW**K**FETSK	K79	TML	HILIC	
Q05639	EF-1-alpha-2, *EEF1A2* (Elongation factor 1-alpha 2)	GSF**K**YAWVLDK	K55	DML	HILIC	
Q5VTE0	EF-1-alpha-like 3, *EEF1A1P5* (Similar to Elongation factor 1-alpha 1)	MDSTEPPYSQ**K**R	K165	MML	IEF	
		MDSTEPPYSQ**K**R	K165	DML	IEF	
Q15056	eIF-4H, *EIF4H* (Eukaryotic translation initiation factor 4H)	AYSSFGGG**R**GSR *(TU)*	R19	DMA	HILIC	Yes
		AYSSFGGG**R**GSR *(TU)*	R19	MMA	HILIC	Yes
		DDFNSGFRDDFLGG**R**GGSRPGDR *(TU)*	R166	DMA	SCX	Yes
Q13283	G3BP-1, *G3BP1* (Ras GTPase-activating protein-binding protein 1)	GPGGP**R**GGLGGGMR	R435	DMA	DMA-IP, HILIC	
		GPP**R**GGMVQKPGFGVG**R**GLAPR	R447, R460	DMA, DMA	HILIC, IEF, SCX	
		GPP**R**GGMVQKPGFGVG**R**GLAPR	R447, R460	DMA, MMA	IEF, SCX	
Q9UN86	G3BP-2, *G3BP2* (Ras GTPase-activating protein binding protein 2)	GPPP**R**GGMAQK	R424	DMA	HILIC	Yes
		LGSG**R**GTGQMEGR	R468	DMA	HILIC	
P11940	PABP-1, *PABPC1* (Polyadenylate-binding protein 1)	N**R**AAYYPPSQIAQL	R419	DMA	HILIC	Yes
		AAYYPPSQIAQL**R**PSPR	R432	DMA	IEF	Yes
		VMSTQ**R**VANTSTQTMGP**R**PAAAAAAATPAVR	R481, R493	MMA, DMA	DMA-IP, HILIC, IEF, WCX	
Q13310	PABP-4, *PABPC4* (Polyadenylate-binding protein 4)	VGVPTAVQNLAP**R**AAVAAAAP**R**AVAPYK	R496, R505	DMA, DMA	DMA-IP, HILIC, IEF	Yes(R496)
		YASSV**R**SPHPAIQPLQAPQPAVHVQGQEPLTASMLAAAPPQEQK	R530	DMA	WCX	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

B4DEH8	cDNA FLJ57714, highly similar to Poly(A)-binding protein 2, *PABPN1*	TSLALDESLF**R**GR	R110	MMA	HILIC	Yes
		A**R**TTNYNSSR	R141	DMA	HILIC	Yes
P98179	Putative RNA-binding protein 3, *RBM3*	SYS**R**GGGDQGYGSGR	R105	DMA	DMA-IP, HILIC, IEF, WCX	
P46783	40S ribosomal protein S10, *RPS10*	GGFG**R**G**R**GQPPQ *(TU)*	R158, R160	DMA, DMA	HILIC	
		GGFG**R**G**R**GQPPQ *(TU)*	R158, R160	DMA, MMA	HILIC	
Q969Q0	60S ribosomal protein L36a-like, *RPL36AL*	QSGYGGQT**K**PIFR	K53	MML	IEF	
Other/Unknown function						
Q59FE5	A disintegrin and metalloproteinase with thrombospondin motifs 10, *ADAMTS10*	**R**PFQLQCTPSCGPGLR	R753	MMA	WCX	Yes
Q5SZL2	Coiled-coil domain-containing protein C6orf204, *C6orf204*	TIGPSCLHDSMEML**R** *(TU)*	R212	DMA	HILIC	Yes
Q9H3H3	UPF0696 protein C11orf68	A**R**GWAGVER	R29	MMA	HILIC	Yes
A2RTY3	Isoform 1 of Uncharacterized protein C17orf66, *C17orf66*	KTPAFPPCCSKP**R**K	R544	MMA	WCX	Yes
P0CG12	Chromosome transmission fidelity protein 8 homolog isoform 2, *CHTF8*	ERPTPWT**R**APLPPR *(TU)*	R17	DMA	SCX	Yes
		DPSMASFP**R**GMNPTGTGAVSFPRPGGLLGPGPGPGPTLNPR	R100	DMA	HILIC	Yes
		TSGNGPPNP**R**PVGLGPGPNPNLRSGF	R201	DMA	HILIC	Yes
		SGFLGTNPAP**R**SGVFPGPGLGPNP**R**PSGLGPGPNLDAR	R225, R239	DMA, DMA	HILIC, WCX	Yes
		AGGLLGTGSGLNL**R**MAGPQGLDLAPILR	R266	DMA	HILIC, IEF	Yes
		AAGLLGANSASFSQASGNMGTSPSSMA**R**VPGPMGPNSGPSSR	R322	DMA	HILIC	Yes
		SAGLQGSNPTIFP**R**ASGPLGPNPANFPR	R415	DMA	WCX+HILIC, WCX	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

		VPFP**R**VGSLPGTNPAAFP**R**PGGPM	R494, R508	DMA	HILIC	Yes
Q6YP21	Kynurenine-oxoglutarate transaminase 3, *CCBL2*	GPPS**R**GGHMDDGGYSMNFNMSSSR *(TU)*	R125	DMA	HILIC	Yes
P50238	CRP-1, *CRIP1* (Cysteine-rich protein 1)	GFG**R**GGAESHTFK	R68	MMA	HILIC, IEF, WCX	Yes
O15528	25-hydroxyvitamin D-1 alpha hydroxylase, mitochondrial, *CYP27B1*	EGEEWQ**R**LR	R136	MMA	WCX	Yes
Q96EP5	DAZ-associated protein 1, *DAZAP1*	GPPPAG**R**GAPPPPPPFTSY	R253	DMA	HILIC	Yes
		GPPPAG**R**GAPPPPPPFTSY	R253	MMA	HILIC	Yes
Q8NHQ9	ATP-dependent RNA helicase DDX55, *DDX55*	AGL**R**NPVR	R218	MMA	HILIC	Yes
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2, *HNRNPUL2*	S**R**GQGYVGGQR	R656	DMA	HILIC, IEF	Yes
		S**R**GQGYVGGQR	R656	MMA	HILIC	Yes
		AYGQQYWGQPGN**R**GGY**R**	R684, R688	MMA, MMA	HILIC, IEF, WCX	Yes
P11021	GRP-78, *HSPA5* (78 kDa glucose-regulated protein)	LGG**K**LSSEDKETMEK	K586	TML	HILIC	Yes
Q96IS6	HSPA8 protein, *HSPA8*	AMT**K**DNNLLG**K**FELTGIPPAPR	K392, K399	MML, MML	WCX	Yes
		FELTGIPPAP**R**	R410	MMA	HILIC	Yes
		ATVEDEKLQG**K**INDEDK	K502	TML	HILIC	Yes
Q86WI3	Protein NLRC5 Isoform 2, *NLRC5*	**R**PHQSCGSSPRR	R122	MMA	WCX	Yes
Q9UBV8	Peflin, *PEF1*	**R**QGCPGAAGQAPGAPPGSY	R7	DMA	HILIC	Yes
Q92576	PHD finger protein 3, *PHF3*	TSTNFSPM**R**PQQPNLQHLK	R1799	DMA	HILIC	Yes
Q96M27	Protein PRRC1, *PRRC1*	STPQPPLPPV**R**PSAPLPF	R67	DMA	HILIC	Yes
Q9Y520	Protein PRRC2C, Isoform 6, *PRRC2C*	QFQSAPATV**R**MTQPFPTQFAPQAK	R2654	DMA	WCX	Yes
P79522	Proline-rich protein 3, *PRR3*	GWWGVNAEPPFPGPGHGGPT**R**GSFHK	R120	DMA	HILIC	Yes
Q2VIN3	RBM1	DSYSSSS**R**GAP**R**GAGPGGSR	R368, R372	DMA, DMA	HILIC	Yes
Q8ND24	RING finger protein 214, *RNF214*	I**R**ALFPAPLAQISTPMFLPSAQVSYPGR	R621	DMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing, SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

Q9BPU5	RT-LI	TPQG**R**GTVAAAAVAATA	R6	DMA	HILIC	Yes
Q92734	Protein TFG, *TFG*	N**R**PPFGQGYTQPGPGYR	R385	DMA	DMA-IP, HILIC, IEF, WCX	Yes
Q15025	TNFAIP3-interacting protein 1, *TNIP1*	LFHLPEYTW**R**LPCGGVR	R599	DMA	HILIC, SCX	Yes
Q5T6F2	UBAP-2, *UBAP2* (Ubiquitin-associated protein 2)	GEENGIDCNQVDKPSD**R**GK	R166	DMA	HILIC	Yes
Q9C0J8	WD repeat-containing protein 33, *WDR33*	SGSNWG**R**GSNMNSGPPR *(TU)*	R1315	DMA	HILIC	Yes
Q9Y3S1	Serine/threonine-protein kinase WNK2, *WNK2*	GDQP**R**SEVCGGDLALPPVPKEAVSGR	R1631	DMA	HILIC	Yes

MMA = mono-methylarginine, DMA = di-methylarginine, MML = mono-methyllysine, DML = di-methyllysine, TML = tri-methyllysine
DMA-IP = di-methylarginine immunoprecipitation, HILIC = Hydrophilic Interaction Liquid Chromatography, IEF = isoelectric focusing,
SCX = Strong Cation Exchange, WCX = Weak Cation Exchange, *(TU)* = site identified uniquely by Thomas Uhlmann
* = Localisation of methylation uncertain

Table 4.5: Methylation sites identified in both primary T-lymphocytes and the Jurkat T-lymphocyte cell line.

4.3 Discussion

Mass spectrometry based proteomics has advanced rapidly over the last decade, enabling ever more comprehensive studies of post-translational modifications such as acetylation, phosphorylation and ubiquitination (Henriksen et al., 2012; Wagner et al., 2011; Phanstiel et al., 2011). Extensive efforts have been made to develop techniques and reagents to enrich for these modifications so that it is now possible to routinely identify thousands of sites for each of these PTM types. By contrast, arginine methylation has remained an area largely unexplored by mass spectrometry based proteomics due to a lack of available reagents and techniques for enrichment. This has limited our understanding of the significance and biological role of arginine methylation.

Several studies point to an important role for arginine methylation in T-lymphocyte activation (Mowen et al., 2004; Richard et al., 2005; Blanchet et al., 2005; Lawson et al., 2007; Kim et al., 2004), in addition, this work has demonstrated an increase in arginine methylation and an increase in PRMT1 after long term stimulation of primary human T-lymphocyte (Chapter 3). Thus, from an immunological and cell biology point of view, primary human T-lymphocytes are interesting cells in which to study arginine methylation. In addition, unlike in cell lines such as Jurkat which have been cultured for over 30 years, primary cells are non-transformed and cellular processes are not dysregulated.

4.3.1 Labelling primary cells

The first step to enable identification of arginine methylation sites in primary T-lymphocytes was labelling by heavy methyl SILAC (Ong et al., 2004). Primary T-lymphocytes have been successfully labelled by SILAC in the past (Loyet et al., 2005), but no mass spectrometric study of arginine methylation in primary cell material has previously been published. Because CD4+ T-lymphocytes are more numerous in the blood compared to CD8+ T-lymphocytes, this sub-type was chosen for studying arginine methylation as more cells could be purified from whole blood. As an initial test, CD4+ T-lymphocytes were activated by plate bound anti-CD3 and soluble anti-CD28 antibodies. Interleukin-2 was added and cells proliferated for 7 days in labelling media, followed by a rest of 3 days and then 7 days of re-stimulation. This resulted in very high labelling efficiency (98.7%, Figure 4.6). In fact, labelling by heavy methyl SILAC for 7-9 days of stimulation and proliferation was sufficient to observe methyl SILAC pairs, therefore this procedure was followed for most experiments.

Primary cells are rarely used for quantitative SILAC-based proteomics (Zhang et al., 2011), however, these results show that primary T-lymphocytes are amenable to labelling by SILAC and thus could be used for a wide range of SILAC-based studies (e.g. quantitative phosphoproteomic studies, protein turnover measurements, perturbation analyses).

4.3.2 Enriching for arginine methylation

The next step after labelling was to identify an effective method for enriching arginine methylation sites so they could be identified by mass spectrometry. Ong et al. used an antibody against methylarginine (7E6) to enrich for arginine methylated proteins (Ong et al., 2004). The 7E6 antibody was raised against asymmetrically dimethylated arginine and is claimed to specifically recognise monomethyl, asymmetric dimethyl and symmetric dimethyl arginine both in the free and bound form. Unfortunately the performance of this antibody falls rather short of these claims and appears to be rather non-specific (Chen et al., 2004). Here, only 4 arginine methylation sites in 3 proteins were found in primary T-lymphocytes using this antibody (Figure 4.1). This is considerably less than the 57 arginine methylation sites in 31 proteins found by Ong et al., which could be due to the fact that only one experiment was performed with this antibody and primary cells were used here where levels of arginine methylation may be lower than in transformed cell lines. Almost all arginine methylation sites identified using the 7E6 antibody are arginine-glycine rich, suggesting a bias towards this motif making it unsuitable for large scale approaches (Ong et al., 2004; Uhlmann et al., 2012).

There was therefore a clear need for alternative enrichment strategies. Fractionation of peptides by chromatography has been used to identify many different PTM sites including phosphorylation and ubiquitination (Beausoleil et al., 2004; Starita et al., 2012; McNulty and Annan, 2008). These strategies take advantage of the fact that tryptic peptides containing a phosphorylation

or ubiquitination will possess different physico-chemical properties, such as altered charge or hydrophilicity, compared to the majority of other peptides. For example, an ubiquitinated tryptic peptide will have charge +3 or above at acidic pH due to the additional N-terminus from the ubiquitin remnant and can be selected by SCX.

Properties of known arginine methylated tryptic peptides from previous studies were plotted and it was found that arginine methylated peptides in general have a higher charge, are more basic and hydrophilic than the majority of non-methylated peptides (Figure 4.8). The higher charge is due to mis-cleavage by trypsin but the causes of high pI and hydrophilicity were less obvious. Mis-cleavages do not lead to peptides with higher isoelectric points as acidic residues can be present. It seems that methylated arginines tend to be surrounded by non-acidic, hydrophilic residues with some exceptions (e.g. methylation site in TFPT, Table 4.5). Although these properties can make peptides challenging to analyse by mass spectrometry (higher charged peptides can fragment poorly and hydrophilic peptides may not be retained by an in-line reverse phase column). These properties can however also be exploited for chromatographic enrichment.

SCX was used to select for highly charged peptides in both primary T-lymphocytes and Jurkat cells (Thomas Uhlmann) leading to identification of 39 arginine methylation sites (Figure 4.9). Although this was still a relatively low number, 25 of these were novel indicating that this new approach enriched a different subset of sites compared to the antibody. Lysine methylated peptides should also have been enriched using this technique, however only one lysine methylation site was found supporting other findings that

arginine methylation accounts for the majority of protein methylation (Najbauer et al., 1967). Amino acid analysis (Thomas Uhlmann) showed that enrichment for arginine methylation was ∼4 fold when comparing <+3 peptides to ≥+3 peptides. There are several features apart from methylation that might result in a peptide having a charge of ≥+3, it may contain histidine or mis-cleavage may occur due to proline or an acidic residue C-terminal to arginine or lysine. As a result, SCX fractions were still complex, limiting the number of methylation sites identified.

IEF was tested as a method to enrich for basic peptides. This led to a higher number (66) of methylation sites compared to SCX and the antibody, overlap between SCX and IEF was low (Figure 4.17, A). Unfortunately, available IPG strips have a maximum pH of 11, many arginine methylated peptides have pIs above this and are therefore lost from the strip making IEF a sub-optimal method for enrichment. WCX was thus developed as a technique to select all highly basic peptides with pIs above a chosen binding pH. Initial tests indicated that this is an effective method to select basic peptides (Figure 4.12) with no upper limit on pI. To test the method, heavy methyl SILAC Jurkat cells were used, leading to identification of 52 arginine methylation sites, 26 of which were novel. Amino acid analysis showed a low level of enrichment (Figure 4.14) due to the fact that basic peptides were eluted in only one fraction which proved to be still complex when analysed by mass spectrometry. Further fractionation of basic peptides is required to improve coverage of methylated peptides, for example, bound basic peptides could be eluted from the WCX column in salt steps.

A third type of chromatography, HILIC, was used to enrich for arginine

methylated peptides. This was the most effective of all tested methods, yielding 178 arginine methylation sites. HILIC separates peptides according to hydrophilicity. Here, separation was carried out at pH2.7 meaning that peptides would have been positively charged. Consequently, it is likely that peptides separated according to charge as well as hydrophilicity, explaining the higher enrichment levels achieved (amino acid analysis showed hydrophilic fractions had 10 fold higher concentration of methylarginine - Thomas Uhlmann) (Jian et al., 2010).

Trypsin is probably the most widely used enzyme in proteomic studies as it generates peptides that are amenable to analysis by mass spectrometry. However, some peptides produced by trypsin digestion may not be readily detected (for example if they are too large). To achieve increased coverage of the proteome, proteases with alternative cleavage specificities are used (Wang et al., 2008). Here, chymotrypsin was tested as an alternative protease as it has a very different cleavage specificity to trypsin (cleaves C-terminal to tyrosine, phenylalanine and tryptophan). Chymotryptic peptides were separated by HILIC giving an additional 37 arginine methylation sites which were not found using trypsin. The overall number of sites found by chymotrypsin was lower than for trypsin, possibly due to the fact that the former has a more relaxed specificity and will also cleave at methionine, leucine, alanine and valine, generating a more complex peptide mixture and spreading methylation sites over many peptides with different cleavage sites. Many of the novel sites identified were proline rich which may have been missed by tryptic cleavage as proline causes mis-cleavage by trypsin when C-terminal to an arginine or lysine. Future efforts to map arginine methylation sites could also make use

of Lys-C, an extremely robust protease cleaving C-terminal to lysine residues.

Chymotryptic arginine methylated peptides were more basic and hydrophilic than non-methylated peptides (Table 4.2), suggesting that IEF or WCX could be used as enrichment strategies. However, the highly charged nature of methylated peptides is lost (Figure 4.16) so SCX would not be a suitable technique. HILIC may consequently be less effective with chymotryptic arginine methylated peptides as charged based enrichment is lost, possibly partly explaining the lower number of sites observed.

All together, HILIC permitted the identification of 215 arginine methylation sites in 115 proteins. Of these, 171 were new and 94 proteins were not previously known to be modified by arginine methylation. HILIC was therefore the superior enrichment technique, however, 50 sites were uniquely identified using the other enrichment techniques showing that relying on one method cannot give extensive coverage. In an attempt to improve enrichment, WCX was coupled to HILIC, potentially allowing the selection of arginine methylated peptides based on pI, charge and hydrophilicity all in a single work-flow. By this approach, only 5 additional sites were discovered. It is not clear why this approach failed to improve identification of sites since two stage enrichments are commonly carried out in studies of phosphorylation (Brockmeyer et al., 2011). It is possible that losses occurred during the WCX step, or that enrichment by WCX is too weak to have any significant effect on the performance of HILIC.

There appears to be some inter-dependence between charge, pI and hydrophilicity which may also explain the disappointing results when two apparently orthogonal separation methods are combined. Higher charged pep-

tides tend to be more hydrophilic (Figure 4.8) and it was observed that basic peptides are also rather hydrophilic. Thus WCX may not provide additional enrichment prior to HILIC because it selects hydrophilic peptides which are also selected by HILIC.

All methods combined led to the identification of 265 arginine methylation sites in 141 proteins in primary T-lymphocytes and Jurkat cells (work by Thomas Uhlmann). 204 (77%) of these sites are new and 103 (73%) of the proteins are newly identified as being substrates for PRMTs. This more than doubles the number of known sites (137), significantly extending our knowledge of the extent of arginine methylation (Figure 4.17, C and D) and providing much deeper coverage than any previously published study (Ong et al., 2004). There was significant overlap between arginine methylated proteins found in primary T-lymphocytes and those found in the Jurkat cell line. Some sites were found uniquely in primary or Jurkat cells, however it is difficult to conclude anything from this as peptide mixtures were complex leading to stochastic sampling by the mass spectrometer. A low number of lysine methylation sites were identified (13 sites in 11 proteins). This could be due to the fact that lysine methylated peptides have different properties and are not selected with the enrichment methods used here and also lysine methylation is less prevalent (Najbauer et al., 1967). Other types of methylated peptides may have been present in low amounts (e.g. methylation on aspartate or glutamate) but were not searched for.

4.3.3 Alternative approaches to enriching for arginine methylation sites

Chromatography methods are robust and inexpensive but generally only offer very low enrichments for post-translational modifications. This was certainly the case here, for example over 40,000 unique peptides were identified from HILIC fractions from which 215 arginine methylation sites were observed, even combining separation methods could not remove this persistent complexity.

There are several alternative approaches which may be effective for identification of arginine methylation sites on a large scale. *In vivo* enrichment for arginine methylation could be achieved by over-expressing PRMTs, as in Chang et al. who overexpressed PRMT1 in HEK 293 cells leading to elevated levels of arginine methylation (Chang et al., 2010). In this study no actual methylation sites were identified so the effectiveness is uncertain, additionally, over-expressing PRMT1 may lead to generation of spurious sites which may not normally be methylated.

PRMT1 has been mutated to allow it to accept analogues of SAM as a cofactor. The SAM analogues contain an alkene group (C≡C) which is transferred to substrates instead of a methyl group and can be covalently reacted with an azide group. The azide group could be attached to biotin allowing effective enrichment (Wang et al., 2011). Various other SAM analogues have been synthesised which allow endogenous methyltransferases to transfer reactive groups to methylation sites. Willnow et al. synthesised a SAM analogue where selenium replaced the sulfur from which the methyl groups

is normally transferred (Willnow et al., 2012). This led to more stable SAM analogues and increased transfer rates of an alkene group from the selenium centre. These selenium based SAM analogues have only been tested on a small number of substrates and it is not certain how substrate specificity is affected. Furthermore, SAM is not cell permeable in eukaryotic cells so labelling cannot take place *in vivo*. These limitations may mean that this approach is unsuitable for large scale discovery of arginine methylation sites.

The most effective way to enrich arginine methylation sites would be to design a binder that specifically recognises methylated arginine. The arginine residue is positively charged and is frequently found interacting with nucleic acids (Rohs et al., 2009). It should therefore be possible to generate nucleic acid aptamers capable of specifically binding to methylated arginine. The first example of such an aptamer was developed by Hyun et al. to selectively recognise symmetrically dimethylated arginine 8 at the N-terminus of histone H3. The binding affinity was comparable to antibodies recognising modified histone H3, but in order to be useful for large scale identification, a pan-methylarginine specific aptamer would be required which did not rely on a specific sequence context for high affinity binding.

Tudor domains, being naturally occurring readers of methylated arginine, could be recombinantly expressed and used to enrich for methylarginine containing proteins or peptides. Although their affinities for methylarginine are relatively low, the Tudor domains of SMN and TDRD3 could be used for enrichment (Liu et al., 2012). This would also provide information on possible readers of the arginine methylation sites.

Several antibodies against methylated arginine are currently commer-

cially available, these include the 7E6 and 21C7 antibodies, Sym10, Sym11, Asym24 and Asym25 (Boisvert et al., 2003) and an antibody developed by Duan et al. (Duan et al., 2007). The only antibody shown to enrich for arginine methylation prior to identification of sites by mass spectrometry was 7E6, which appears to be biased towards glycine-arginine rich (GAR) sequences (Ong et al., 2004). Recently, extensive efforts have been made to generate monoclonal antibodies against several different types of post-translational modification including arginine methylation. Animals are immunized with an epitope and specific circulating antibodies are captured using the same epitope. These specific antibodies are digested with proteases and peptide sequences identified by mass spectrometry. In parallel, mRNAs coding for immunoglobulins from B-lymphocytes are sequenced. This mRNA sequencing data is used to identify the various light and heavy chains captured by the epitope. Finally, light and heavy chains are paired together in different combinations by expressing them in a single open reading frame. These recombinantly expressed antibodies are then screened for binding. Using this work-flow, antibody mixtures against various forms of methylated arginine have been generated. These mixtures are raised against peptides containing methylarginine surrounded by various amino acids. As a result, they should recognise methylarginine in a broader range of sequence contexts (Cheung et al., 2012).

Results of using these recently developed antibodies against methylated arginine were presented at the ASMS conference in 2012. 10mg of tryptic peptides were immunoprecipitated from a colon cancer cell line resulting in identification of 1428 mono-methylarginine sites in 718 proteins and 444

asymmetric dimethylarginine sites in 188 proteins. However, it is not clear whether labelling by methyl SILAC was used to ensure high confidence. Data was also presented from mouse brain and embryo (378 sites) where labelling by methyl SILAC was not possible (Guo, 2012). This new antibody mixture against methylated arginine is a promising development which could be coupled with methyl SILAC and HILIC to allow high confidence identification of many methylation sites.

The isomethionine methyl SILAC method described here is an improvement on the previously reported heavy methyl SILAC method which should lead to even greater confidence in methylation sites as methyl SILAC pairs cannot be confused with methionine SILAC pairs (Figure 4.24 and Figure 4.25). Furthermore, in isomethionine labelled samples, only methylated peptides will occur in 1:1 pairs. This means that the total number of protein methylation sites could be rapidly estimated by counting the number of methyl SILAC pairs observable by mass spectrometry. This could lead to a targeted approach for methylation where these pairs are selected for fragmentation and identification.

4.3.4 Functional consequences of arginine methylation

The work here has led to the largest set of arginine methylation sites published to date (Uhlmann et al., 2012). Consequently, sites and proteins can be analysed leading to new insights into the biology of arginine methylation. Arginine methylated proteins were found to be present in many cellular

compartments, mainly in the cytoplasm and nucleus (Figure 4.18, A). The largest proportion of proteins interacted with nucleic acids confirming previously published observations and suggesting a widespread role for arginine methylation in modulating protein-nucleic acid interactions (Figure 4.18, B) (Bedford and Clarke, 2009).

Analysing methylation sites for common sequence contexts revealed that 57% of sites contained a C-terminal glycine (Figure 4.20 and Figure 4.21, A) and methylarginine is often embedded in glycine rich regions. This previously noted high glycine frequency often occurs in RNA-binding proteins and further points to a possible widespread role for arginine methylation in regulating protein-nucleic interactions (Denman, 2002). The surrounding glycine residues may make the arginine to be methylated particularly accessible to PRMTs. Another known site targeted by PRMTs is RxR, this motif was also confirmed by data collected here (Figure 4.21, B). Proline was also commonly observed near methylarginine, a feature not previously described on a large scale (Figure 4.20). Thus arginine methylation takes place at glycine-arginine rich (GAR) motifs, RXR motifs and at the proline-rich arginine methylation (PRAM) motifs discovered here. Within the PRAM motifs, proline appeared to be significantly favoured at positions -1 and +4 in relation to the methylated arginine, leading to the consensus sequence (P)Rxxx(P) where proline is present at one or both positions (Figure 4.21, C and Table 4.3). The (P)Rxxx(P) motif accounted for 33% of all methylation sites making it a major site for arginine methylation. The discovery of PRAM motifs provides some insight into the structural context of arginine methylation.

Proline is unique among amino acids as its side chain is linked to its backbone amino group, restricting the φdihedral angle. Consequently, proline-rich sequences tend to adopt a conformation known as a polyproline II (PPII) helix , which is an extended left handed helix with 3 residues per turn. Interaction between adjacent prolyl rings help to stabilise the conformation. PPII helices can often contain other amino acids, in fact sequences without any proline residues such as polyglycine, polyglutamine and polyalanine peptides can adopt a PPII conformation in solution (Rath et al., 2005). Glutamine, glycine, alanine and leucine are among the amino acids frequently found in PPII helices (as well as proline) (Rucker et al., 2003). Small residues such as glycine are thought to permit solvation of the PPII backbone, whilst leucine could form stabilising interactions with prolyl rings (Rucker et al., 2003). In PPII helices, glutamine stabilises the structure by forming a hydrogen bond with the carbonyl oxygen of the next residue. Stretches of glutamine were observed to be present in some methylation sites (e.g. in hnRNP D0) and glycine and alanine were also commonly found (e.g. in snRNPB and PABP-1). The prevalence of these residues is consistent with arginine methylation occurring in PPII helices.

Proline rich sequences commonly mediate protein-protein interactions. The reason for this seems to be that side chains in PPII helices protrude outwards and are readily accessible for the formation of interactions. Backbone carbonyl and N-H groups are also available. Because PPII helices are sterically restricted in conformation, the loss of entropy upon binding is small, making interactions favourable (Rath et al., 2005). Several types of protein domain, such as SH3 and WW domains, are known to bind to proline rich se-

quences. SH3 domains bind to peptides containing xP sequences spaced one residue apart (xPxxP). The xP motifs bind to hydrophobic grooves, thus the residue preceding proline is often hydrophobic. In (P)Rxxx(P) methylation sites, the residue preceding proline at +4 was often found to be hydrophobic (Figure 4.22), supporting the hypothesis that arginine methylation may modulate interactions between proline rich regions and SH3 or SH3-like domains. The proline at position -1 was not found to be preceded by a hydrophobic residue, so this may not form part of the interface but may instead play a role in maintaining a PPII helical conformation (McDonald et al., 2011). The -1 proline may facilitate access of the arginine to a PRMT, as shown by Kölbel et al. who suggested that proline induces a βturn (Kölbel et al., 2012).

The spacing between the methylated arginine and the proline at +4 is more than one complete turn of the PPII helix, indicating that arginine binds to a different interface than the following xP unit. For example, arginine may form a hydrogen bond with an acidic residue as is commonly observed for interactions of proline rich peptides with SH3 domains (McDonald et al., 2011). Alternatively, arginine may form a side chain to main chain hydrogen bond with the carbonyl group of the next residue contributing to stabilising the PPII helix. This would restrict the conformation of the C-terminal residue, favouring small residues (such as glycine) or residues which are already conformationally restricted (proline). 72% of methylation sites found here consisted of RG or RP motifs, supporting this hypothesis. Methylation could disrupt this stabilisation leading to local conformational change.

Arginine methylation adjacent to proline rich regions has been described by Bedford et al.. In this study, it was found that arginine methylation in

arginine-glycine motifs adjacent to proline motifs in Sam68 inhibited binding to an SH3 but not a WW domain (Bedford et al., 2000). However, the methylation sites discovered in this work suggest that arginine methylation frequently occurs within proline rich regions. This may similarly lead to a disruption of binding to SH3 domains as methylation would interfere with hydrogen bonding. Bedford et al. also noted that WASp could be regulated in this way as it contained proline rich regions and arginine-glycine motifs. This prediction appears to be confirmed by the discovery in this work that WASp is indeed arginine methylated.

In addition to its role in regulating protein activity, Fackelmayer suggested that methylation could protect key arginine residues from reacting with endogenous reactive dicarbonyl compounds (Fackelmayer, 2005). However, no direct evidence has so far been found to support this hypothesis.

The occupancy of arginine methylation at three sites was investigated by comparing intensities of non-methylated, monomethylated and dimethylated peptides. Occupancy at these sites was high (17%, 50% and 76%, Figure 4.26). This is not surprising given the weak enrichment achieved with methods used here, only high occupancy sites are likely to have been observable. The high stoichiometry of methylation in hnRNP A2/B1 (76%) is in line with a previous study on an hnRNP (hnRNP K) showing that 5 arginine residues were 100% dimethylated, and a further two were 33% and 10% dimethylated (Ostareck-Lederer et al., 2006). Pahlich et al. demonstrated that arginine methylation did not affect ionization of peptides by MALDI (Pahlich et al., 2005). The effect on ionisation by electrospray is unknown, these values are therefore estimates. Nonetheless, the high stoichiometry

could indicate a constitutive methylation important to maintain protein stability and/or localisation.

4.3.5 Changes in arginine methylation during T-lymphocyte activation

A modified methyl SILAC approach was used here to follow dynamics of arginine (and lysine) methylation during T-lymphocyte activation. This is the first time non-histone arginine methylation dynamics have been followed by mass spectrometry. Zee et al. used heavy methyl SILAC to monitor turnover of histone methylation sites and histone protein, but focused mainly on lysine methylation (Zee et al., 2010). A relatively low number of methylation sites (21) were quantified before and after T-lymphocyte activation (Figure 4.4). Introducing an additional label (Arg10) complicates the peptide mixture and may have led to fewer methylation sites being detected. In addition, cellular material is limited as a single preparation of primary T-lymphocytes is divided into 3 pools, 1 unstimulated and 2 stimulated. Unstimulated primary T-lymphocytes have lower protein content, this will have limited the overall amount of protein digested (since mixing is 1:1:1, Figure 4.27). Many methylation sites did not appear to significantly change in abundance and only 3 arginine methylation sites significantly increased (G3BP-1, hnRNP U and hnRNPUL2). Given the increases in arginine methylation and PRMT1 observed in Chapter 3, this low number may represent only a small fraction of the methylated proteins whose abundances change during activation. The

sites measured here are likely to be of high stoichiometry and occupancy may be constitutive. More powerful enrichment methods (e.g. use of novel antibody mixtures) are required to measure changes in less abundant arginine methylation sites which may be more likely to be regulated following stimulation.

Measuring changes in both protein levels and methylation sites enabled changes in the proportion of protein methylated to be calculated. The largest change in arginine methylation occurred at R460 in G3BP-1 (43 fold increase) (Table 4.4). G3BP-1 is present in stress granules and regulates mRNA stability. A study on the role of G3BP-1 methylation showed that stimulation of the Wnt signalling pathway led to arginine methylation of G3BP-1 by PRMT1. This caused dissociation of β-catenin mRNA from G3BP-1 leading to reduced degradation and may increase protein levels of β-catenin (a transcription factor) (Bikkavilli and Malbon, 2011). The increase in G3BP-1 arginine methylation observed here (Figure 4.28, D) may also lead to dissociation of mRNAs and increased translation of transcription factors important for cellular proliferation and production of cytokines. The large increases in PRMT1 levels (Chapter 3) upon stimulation are also consistent with the increase in G3BP-1 methylation.

Chang et al. noted changes in methylation of ribosomal proteins during the HeLa cell cycle, observing increases in arginine methylation of the 40S and 60S ribosomal subunits at late G_1 and early S phase (Chang et al., 1978). Similarly, an increase in lysine methylation of the 60S ribosomal subunit was observed in the present work after stimulation of T-lymphocytes (Figure 4.28, J and Table 4.4). The role of ribosomal methylation is not clear, but the

methylated residues have been mutated in yeast and found to be functionally important for stress responsive growth (Shirai et al., 2010).

Arginine methylation in the PRxxxP motif of WIPF1 was observed to decrease around 3-fold as a result of activation (Figure 4.28, P and Table 4.4). WIPF1 is a positive regulator of actin polymerisation and is required for proper T-lymphocyte activation. In WIPF1, arginine methylation might negatively regulate activity by reducing binding to SH3 or SH3-like domains. By down-regulating methylation of WIPF1 during T-lymphocyte stimulation, activity of this protein may be increased. The proline-rich region from 321-415 in WIPF1 has been shown to bind Nck (Antón et al., 1998). The proline rich arginine methylation sites in WIPF1 are outside this region and may therefore alter the binding of other proteins. One of the methylation sites (R33) is located in the actin binding region and could have a direct impact on actin polymerisation activity.

The novel arginine methylation sites discovered in this work has led to new insights into the biology of arginine methylation. However, simply generating lists of modification sites does not allow any detailed investigation of molecular consequences of a PTM. To determine molecular and biological effects, arginine methylation sites need to be studied individually. This means a candidate protein must be selected for further investigation. The obvious choice for follow up studies in this case is WASp, since this protein is expressed only in haematopoietic cells and is specifically essential for T-lymphocyte activation.

5. Regulation of WASp activity by arginine methylation

5.1 Introduction

5.1.1 WASp

The cellular actin cytoskeleton consists of filamentous actin (F-actin) which is formed through the polymerisation of globular actin (G-actin). The formation of new actin filaments (nucleation) is kinetically unfavourable because the intermediate states of polymerisation (e.g. actin dimers and trimers) are unstable. Nucleation factors such as the Arp2-Arp3 complex are therefore necessary for polymerisation of actin. On its own, the Arp2-Arp3 complex is an inefficient actin nucleator and requires activation by nucleation promoting factors (NPFs).

Wiskott-Aldrich syndrome protein (WASp) is a member of a family of five NPFs expressed in mammals (WASp, N-WASp, WAVE1, WAVE2, WAVE3) (Kurisu and Takenawa, 2009). It is expressed only in haematopoietic cells and regulates the restructuring of the actin cytoskeleton during several cellular processes. Mutations in the *WAS* gene leading to loss of function are the most frequent cause of the rare immunodeficiency disease Wiskott-Aldrich syndrome which affects 1-10 individuals in 10 million (Derry et al., 1994;

Thrasher, 2009). Those mutations causing ablation of WASp expression give rise to severe symptoms including recurrent infections, thrombocytopenia (low platelet count leading to excessive bleeding) and autoimmunity. Unless treated, these symptoms are usually fatal in early life. WASp acts as a scaffold molecule, integrating and transmitting signals to the actin cytoskeleton enabling cellular morphology to respond to the environment. Being a scaffold protein, WASp has a modular structure and is composed of 5 protein domains. From N- to C- terminus the domain structure of WASp consists of a WH1 (WASP Homology 1) domain, a Basic region, a GTPase binding domain (GBD), a polyproline (PPP) domain and a verprolin homology domain-cofilin homology domain-acidic region (VCA) domain. All WASp family members contain the PPP and VCA domains, but only WASp and N-WASp contain the WH1 domain.

The VCA domain of WASp binds the Arp2-Arp3 complex and monomeric actin. Upon binding the VCA domain, the Arp2-Arp3 complex undergoes a conformational change allowing nucleation and polymerisation of actin.

5.1.2 The role of WASp in T-lymphocyte function

Although WASp is important for the function of many cells in the haematopoietic lineage, its role in T-lymphocytes has been best characterised. When a T-lymphocyte contacts an antigen presenting cell (APC) and recognises the peptide presented by MHC, its shape changes to maximise contact area and promote adhesion. As a result, an area of close contact between the two

cells is formed known as the immunological synapse (IS) or supramolecular activation complex (SMAC). The IS helps to sustain signalling from the TCR and polarise the T-lymphocyte to direct the release of cytokines and other effectors towards the APC. WASp has a crucial role in enabling the formation of the IS. In the absence of T-lymphocyte activation, WASp is present in the cytoplasm in an auto-inhibited form caused by an interaction between the VCA domain and the GBD which occludes the Arp2-Arp3 binding sites. Fyn, a tyrosine kinase related to Lck, is activated after ligation of the TCR and phosphorylates WASp at Y291. Y291 is within the GBD and phosphorylation at this site is thought to destabilise the auto-inhibited form leading to activation. Fyn is anchored to the membrane and recruits WASp to the forming IS via an SH3 domain which binds the PPP domain of WASp (Figure 5.1). WASp is thereby activated near the cell-cell contact area leading to localised polymerisation of actin. At the centre of the IS, actin is depleted while at the edges polymerisation is increased, this ring of polymerising actin expands outwards, mediated by WASp, to drive formation of the contact zone (Kaizuka et al., 2007). There are therefore two zones in the IS: a peripheral SMAC (pSMAC) containing functional signalling complexes and a central SMAC (cSMAC) where signalling is terminated. This has led to a model where actin forms cytoskeletal corrals which may create membrane microdomains containing TCR signalling components (Dustin and Depoil, 2011; Beemiller and Krummel, 2010). Before ligation of the TCR these microdomains are kept separated so signalling complexes cannot form. Re-structuring of the actin cytoskeleton at the IS by WASp then allows microdomains to coalesce leading to formation of signalling complexes.

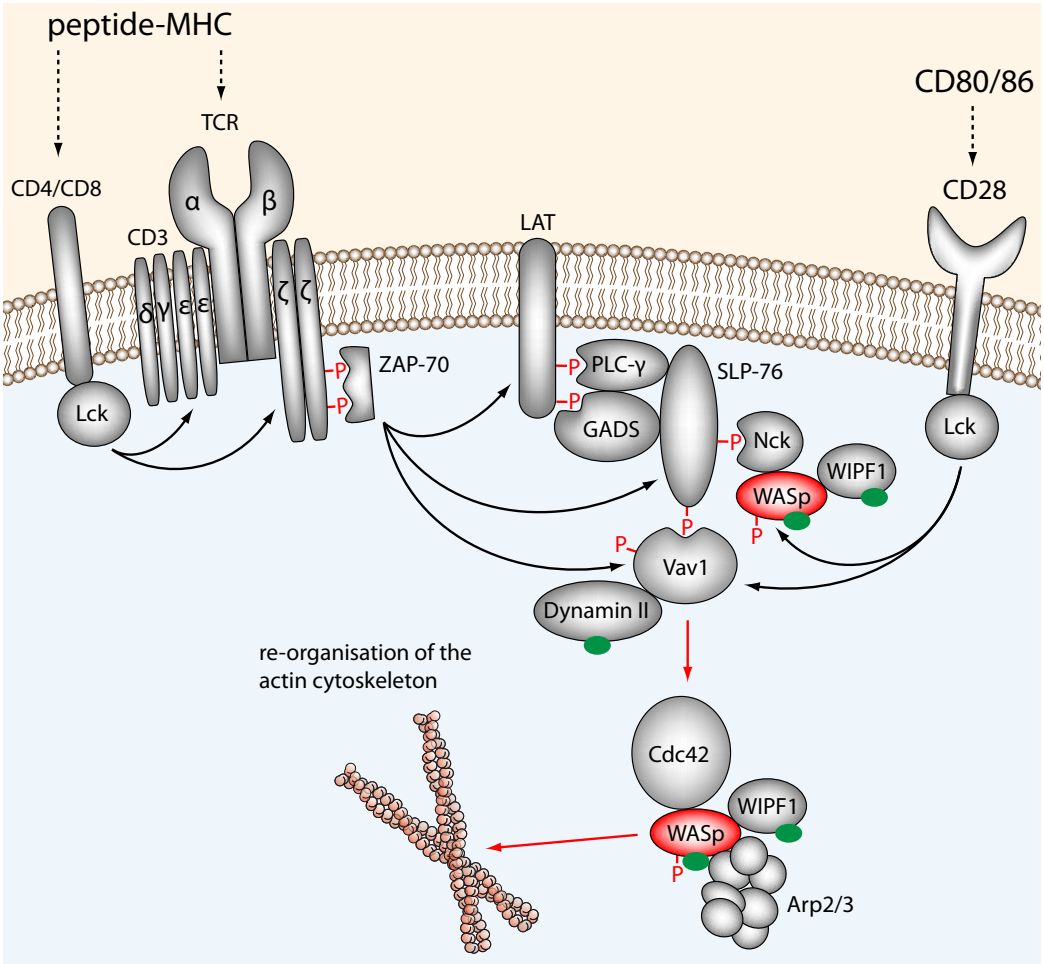


Figure 5.1: Signalling pathway after recognition of an antigenic peptide by the TCR. Ligation of the TCR and the co-receptors CD28, CD4/CD8 leads to an early wave of tyrosine phosphorylation (shown in red). The protein tyrosine kinases Lck and Zap-70 initiate this cascade. The nucleation promoting factors WASp, WIPF1 and Dynamin II are recruited to the immunological synapse. WASp becomes phosphorylated by Lck or Fyn and activated, is able to bind the Arp2-Arp3 complex and promote actin polymerisation leading to establishment of a stable immunological synapse and sustained signalling. Proteins found to be arginine methylated in this work are labelled with a green circle.

Polymerisation of actin at the synapse may also provide the force to push the two opposing membranes together, leading to exclusion of CD45 which has a large ectodomain and assembly of functional TCR-signalling complexes (Kinetic segregation model) (Davis and van der Merwe, 2006; Beemiller and Krummel, 2010). Many TCR signalling factors have binding sites for actin or actin-binding proteins, enabling stabilisation of signalling complexes through binding to the cytoskeleton.

WASp also appears to have an actin-independent role in T-lymphocyte activation. Silvin et al. showed that a WASp construct lacking the VCA domain could enhance NFAT mediated transcription after T-lymphocyte activation and that this activity is dependent on the WH1 domain (Silvin et al., 2001). A subsequent study by Taylor et al. provided evidence (by cellular fractionation and immunofluorescence) that WASp was present in the nucleus in T_H1 cells. In the nucleus, WASp was found to associate with transcription factors such as STAT1 and also proteins involved in histone lysine methylation. Chromatin immunoprecipitation assays indicated that WASp bound to several genomic regions including the *TBX21* gene, coding for the T_H1 lineage-determining transcription factor T-bet. WASp appears to influence histone lysine methylation patterns at this gene leading to increased expression (Taylor et al., 2010).

The essential role of WASp in T-lymphocyte activation was initially shown in $WAS^{-/-}$ mice. In WASp deficient T-lymphocytes from these animals, TCR induced actin polymerisation was deficient, stimulation induced production of IL-2 was reduced and proliferation was impaired (Zhang et al., 1999). As well as its importance to optimal T-lymphocyte activation, WASp is also

required for proper development of T-lymphocytes. Patients with WAS have reduced peripheral CD4+ and CD8+ T-lymphocytes, indicating a partial block in T-lymphocyte maturation (Park et al., 2004).

5.1.3 Regulation of WASp

Numerous interacting partners regulate the activity and localisation of WASp (Figure 5.2). One of the major regulators of WASp is WIPF1 (Wasp interacting family member 1) which binds to the WH1 domain of WASp. This association protects WASp from degradation by calpains and the proteasome. The majority (80%) of WASp is bound to WIPF1 and when WIPF1 is deleted, levels of WASp are severely reduced giving rise to WAS (Le Bras et al., 2009; Lanzi et al., 2012). The activity of WASp is directly regulated by signalling from the TCR. As outlined above, phosphorylation on Y291 is an important regulatory modification. Knock-in mice expressing WASp with the phosphorylation site mutated to phenylalanine display WAS-like symptoms. Phosphorylation also negatively regulates WASp, leading to recruitment of an E3-ubiquitin ligase, ubiquitylation in the WH1 domain and degradation (Blundell et al., 2009; Reicher et al., 2012).

The Rho family GTPase CDC42 is activated by Vav1 downstream of signalling from the TCR and binds to the GBD of WASp. This interaction disrupts the auto-inhibited form of WASp by releasing the VCA domain (Abdul-Manan et al., 1999). The role this interaction plays in WASp activation *in vivo* is unclear as T-lymphocytes expressing WASp with no

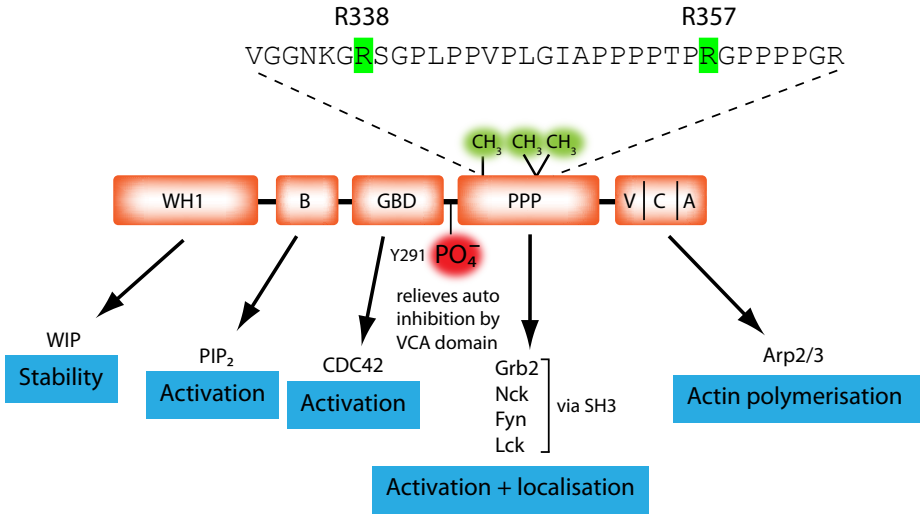


Figure 5.2: Domain structure of WASp and regulation of activity. WASp contains a WASp homology 1 (WH1) domain, a Basic (B) domain, a GTPase Binding Domain (GBD), a Polyproline (PPP) domain and a verprolin homology domain-cofilin homology domain-acidic region (VCA) domain. Arginine methylation sites are highlighted in green and the detected methylation type is shown.

GBD behave normally (Badour et al., 2004). The basic domain of WASp binds phosphatidylinositol (4,5) bisphosphate (PIP_2) which appears to activate WASp synergistically with CDC42 (at least *in vitro*) and may help to recruit WASp to the membrane (Higgs and Pollard, 2000).

The PPP domain of WASp interacts with several proteins via their SH3 domains. Src family tyrosine kinases (e.g. Fyn), Tec family kinases (e.g. ITK), adaptors such as Nck, PSTPIP1 and VASP have been shown to interact with WASp via the PPP domain. WASp is recruited to the IS by binding to the C-terminal SH3 domain of Nck and also to Fyn (Zeng et al., 2003). In addition, PSTPIP1 is important for localisation of WASp to the IS but also recruits the phosphatase PTPN12 which can remove phosphorylation at Y291 thus deactivating WASp. The importance of the PPP domain has been

shown in experiments using T-lymphocytes expressing WASp lacking this region. These T-lymphocytes formed fewer conjugates with APCs and when conjugates were formed the PPP domain deficient WASp was not recruited to the plasma membrane (Cannon et al., 2001). Binding of the PPP domain to multi-SH3 domain containing proteins also leads to oligomerization of WASp enhancing its ability to activate Arp2-Arp3 (Padrick et al., 2008).

Finally, the VCA domain has been shown to be phosphorylated at S483 and S484 by casein kinase 2 which increases binding to the Arp2-Arp3 complex and enhances actin polymerisation. These phosphorylations are of high stoichiometry and appear to be constitutive (Cory et al., 2003).

A large portion of mutations in the *WAS* gene occur in the WH1 domain and several have been shown to disrupt WIPF1 binding leading to WASP degradation and severe symptoms of WAS. Mutations in the VCA domain impair the ability of WASp to mediate actin polymerisation and also lead to symptoms of WAS. Some mutations have been found in the GBD leading to disruption of VCA binding and reduced auto-inhibition. Patients with these latter mutations present with a WAS related disorder known as X-linked neutropenia (XLN) and are predisposed to bacterial infection. These clinically relevant mutations demonstrate the importance of both the positive and negative regulators of WASp.

5.1.4 Aims

In Chapter 4, development of enrichment methods and application of mass spectrometry based proteomics led to the identification of many novel arginine methylated proteins in T-lymphocytes. Among these were Dynamin II, WASp and WIPF1, proteins involved in re-structuring the actin cytoskeleton in response to signalling from the TCR. In order to gain a molecular understanding of the role of arginine methylation in T-lymphocyte activation, WASp was selected for further study as it has a clear and specifically important role during the activation of T-lymphocytes. The aim of this chapter was to begin to dissect the functional consequences of arginine methylation on WASp by mutating the methylated sites to lysine.

5.2 Results

5.2.1 Conservation of methylation sites

Functional PTM sites are more likely to be conserved across species because they are important for the biological function of the protein (Beltrao et al., 2012). To assess the functional significance of the arginine methylation sites discovered in WASp, the human WASp protein sequence was aligned with WASp sequences from three model organisms of increasing evolutionary divergence from humans (*Mus musculus* (mouse), *Denio rerio* (zebra fish) and *Schizosaccharomyces pombe* (fission yeast)). The Clustal Omega sequence alignment program with default settings was used to align protein sequences (Sievers et al., 2011).

Overall conservation of WASp between the organisms studied is low. The R338 methylation site is conserved in mouse and human but not zebra fish or fission yeast (Figure 5.3, A). The R357 methylation site is conserved across all 4 organisms indicating a functional importance. The (P)Rxxx(P) PRAM motif surrounding the R357 site is also conserved as proline is present at one or both positions in all 4 organisms suggesting that this may be important for function. Within the family of WASp homologues in humans, both arginine methylation sites are not conserved indicating that regulation by arginine methylation, at least at these sites, is unique to WASp (Figure 5.3, B). In 3 of the homologues the PRAM motif appears to be conserved suggesting a role independent of methylation.



Figure 5.3: **(A)** Conservation of arginine methylation sites across WASp from human, mouse, zebra fish and fission yeast. **(B)** Conservation across human WASp homologues, WASL (N-WASp), WASF1 (WAVE1), WASF2 (WAVE2) and WASF3 (WAVE3). Methylated arginine is highlighted in green and the (P)Rxxx(P) motif in grey.

5.2.2 PRMT responsible for WASp methylation

In order to identify candidate PRMTs responsible for methylation of WASp a co-transfection and pull-down assay was performed according to Hsu et al. (Hsu et al., 2011). Jurkat T-lymphocytes were co-transfected with PRMT-GFP and Strep-tag WASp plasmids. Expression levels of WASp and the PRMT-GFPs were checked before the pull-down (Figure 5.4, input). WASp was pulled down with streptactin and presence of associated PRMTs was checked by western blot (Figure 5.4, pull-down). None of the expressed PRMTs were detected in the pull-down.

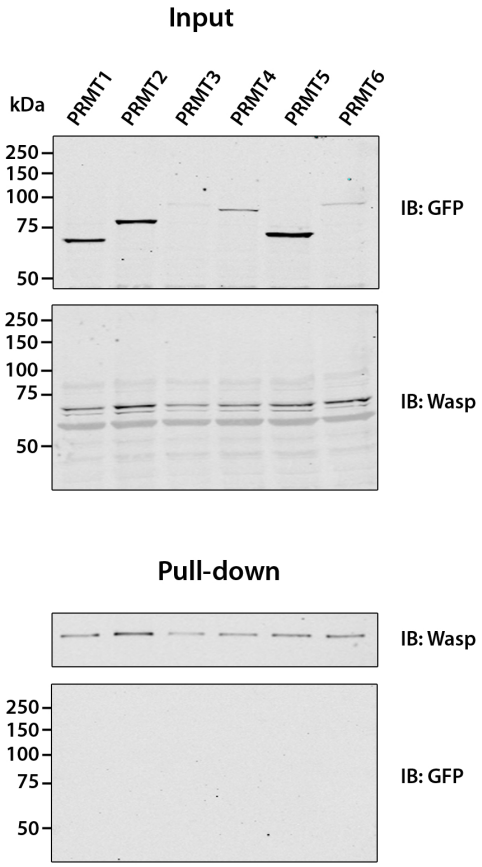


Figure 5.4: Screening for the PRMT responsible for WASp methylation. Jurkat T-lymphocytes were co-transfected with plasmids encoding PRMT-GFP and Strep-tagged WASp (input). WASp was purified by pull-down and the presence of bound PRMT-GFP was assayed (bottom).

5.2.3 Role of WASp methylation sites in interactions with binding partners

Several proteins are known to bind to WASp. To test the importance of the methylation site to interaction partners, WASp mutants were expressed in Jurkat T-lymphocytes, pulled down and bound interaction partners were

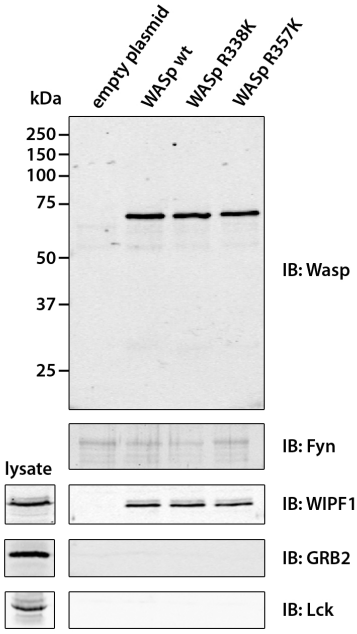


Figure 5.5: Testing the importance of WASp methylation sites for the binding of partners. Jurkat T-lymphocytes were transfected with plasmids encoding Strep-tagged WASp wild type (WASp wt) or plasmids encoding methylation site mutants. WASp was pulled down using Streptactin resin and eluted from the beads by boiling. Presence of binding partners was assayed by western blot, with a sample of whole cell lysate as a positive control.

detected by western blotting (Figure 5.5). WASp mutants were present at equal levels indicating that methylation does not have an impact on protein stability. Known binding partners Fyn, GRB2 and Lck could not be detected but WIPF1 was observed. Association of WIPF1 with WASP was unaffected by mutation of the methylation site to lysine.

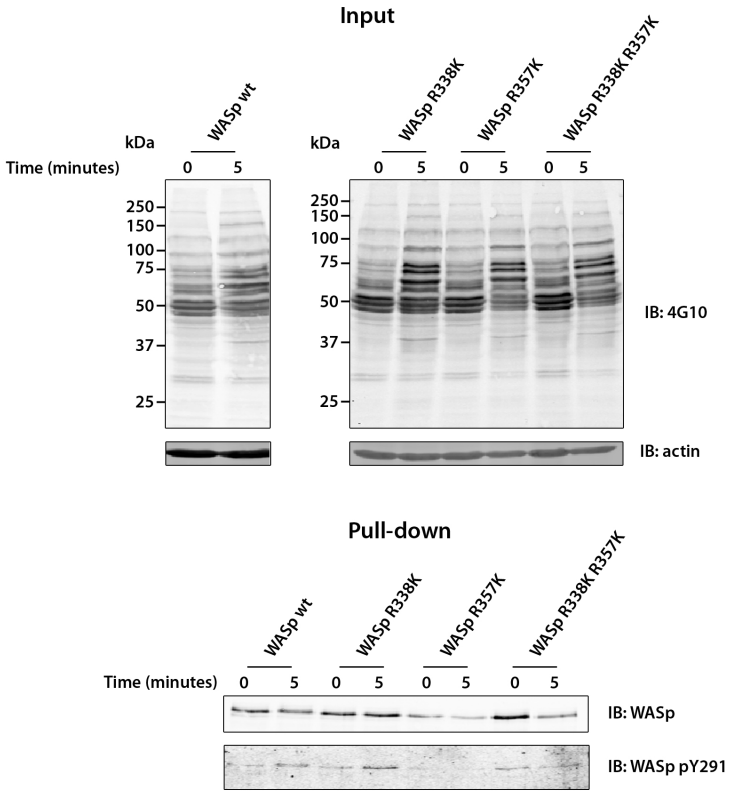


Figure 5.6: Testing the importance of WASp methylation sites for phosphorylation at Y291 (pY291). Jurkat T-lymphocytes were transfected with plasmids encoding Strep-tagged WASp wt or Strep-tagged WASp methylation site mutants. Cells were stimulated with anti-CD3 and anti-CD28 coated beads for 5 minutes. Samples were taken and activation verified by western blotting with 4G10 (input). WASp was pulled-down with streptactin and phosphorylation status was measured using an antibody specific to phosphorylation at Y291 (pull-down).

5.2.4 Role of WASp methylation sites in activation of WASp

The PPP domain of WASp is known to be important for its recruitment to the plasma membrane where it is phosphorylated and activated by Fyn downstream of TCR activation. The arginine methylation sites are within

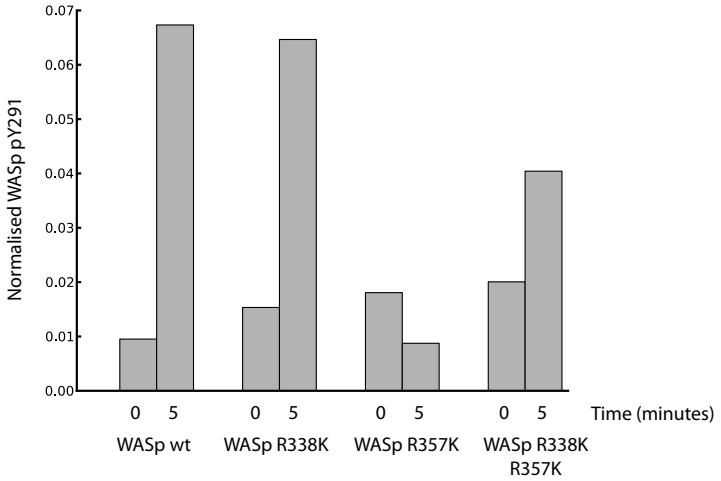


Figure 5.7: Quantification of pull-down in Figure 5.6. Levels of phosphorylated Y291 in WASp were normalised to total WASp before and after stimulation for each WASp construct. Time after stimulation is shown in minutes.

the PPP domain and to determine the affect they might have on WASp activation after TCR signalling, Jurkat T-lymphocytes were transfected with Strep-tagged WASp wt or Strep-tagged WASp mutant plasmids. Cells were stimulated by CD3 and CD28 antibodies for 5 minutes as maximum WASp phosphorylation is observed at this time (Badour et al., 2004). Cells were successfully stimulated as shown by a 4G10 blot on whole cell lysate (Figure 5.6, input), several proteins are phosphorylated on tyrosine. The presence of additional WASp or the WASp mutants does not appear to have a global impact on signalling from the TCR. Strep-tagged WASp was pulled down before and after stimulation and levels of phosphorylation at Y291 were assayed with a specific antibody (Figure 5.6, pull-down). WASp wt and WASp R338K are phosphorylated at Y291 after cellular activation. However WASp R357K is not phosphorylated at Y291 after TCR ligation. The consequences of mutating both sites are less clear due to a higher level of WASp in the

0 min sample compared to 5 min. The pY291 bands in Figure 5.6 were quantified using imaging software and normalised to total WASp in order to objectively measure WASp phosphorylation status (Figure 5.7). This analysis showed that WASp wt and WASp R338K were phosphorylated after cellular stimulation but that WASp R357K was not. There appears to be some induction of phosphorylation in the double mutant, but the high noise in the western blot for the 5 min sample may have led to an overestimate of phosphorylation levels.

5.3 Discussion

WASp provides a link between recognition of a foreign peptide by the TCR and re-structuring of the actin cytoskeleton leading to important changes in membrane dynamics and cellular morphology. These changes are crucial for optimal T-lymphocyte activation, therefore WASp is tightly regulated to ensure that polymerisation of actin occurs at the right time and place. Arginine methylation of WASp at R338 and R357 is a potentially novel form of regulation and its study could lead to a more complete understanding of WASp behaviour.

Sequence alignment studies indicate that the R357 site may be functionally significant as it is conserved in mouse, zebra fish and fission yeast (Figure 5.3, A). The R338 site is less conserved, being present in mouse but not the other model organisms, indicating it may be less functionally significant. WASp contains two GAR motifs with arginine residues conserved across all

4 species (one of these is shown in Figure 5.3 A), these could also be subject to methylation by PRMTs. The (P)Rxxx(P) motif at R357 is found in all 4 species suggesting that it also has an important role.

The arginine methylation sites in WASp are not conserved in other WASp-family members in humans (Figure 5.3 B), indicating that regulation of nucleation promoting factors by arginine methylation may be restricted to those expressed in T-lymphocytes.

A pull-down assay was performed to identify candidate PRMTs responsible for WASp methylation. Although PRMT3 was not expressed, the other PRMTs were visible in the input sample (Figure 5.4). WASp was successfully pulled down, however no PRMTs were observed to be associated. This assay was limited as only PRMT1-6 were available and PRMT3 was not expressed. In addition, this experiment is rather indirect and may not be suitable for detecting transient enzyme-substrate interactions. A better approach would be to over-express the PRMTs and WASp in cells and detect methylation either radioactively or with a specific antibody. PRMT2 contains an SH3 domain which could interact with the PPP domain of WASp, however the catalytic activity of this PRMT is not robust.

Several proteins have been shown to interact with WASp via its PPP domain. To test whether arginine methylation might affect these interactions, methylated arginines were mutated to lysine. This is a common strategy in assessing the effect of a methylation site as it abolishes methylation but maintains the positive charge which may itself be important for interactions. Mutating the methylation sites did not have an affect on binding to WIPF1. This was not surprising as the WH1 domain is distal to the methylation

sites. Binding of other known interactors Grb2, Fyn and Lck could not be detected either to WASp wt or WASp mutants. Cells may need to be stimulated to induce these interactions, although Grb2 was reported to bind constitutively (She et al., 1997). Alternative buffers may need to be used to prevent interactions being disrupted.

Phosphorylation of Y291 is a convenient read-out for recruitment and activation of WASp at the IS. The PPP domain of WASp is known to be essential for recruitment and phosphorylation. To assess whether arginine methylation in this domain has a role in the activation of WASp, Jurkat T-lymphocytes were transfected with Strep-tagged WASp wt and WASp mutants. Cells were then stimulated for 5 minutes with anti-CD3 and anti-CD28 coated beads, WASp was pulled down and phosphorylation at Y291 checked (Figure 5.6). Expression of WASp mutants did not impair total TCR induced phosphorylation as shown by western blotting with 4G10. Levels of pY291 in WASp wt and WASp R338K increased after TCR stimulation demonstrating that the Strep-tagged constructs are recruited and activated. Interestingly, removal of the R357 methylation site appears to have abrogated the ability of WASp to be phosphorylated (Figure 5.7), however the double mutant does appear to show some induction of phosphorylation which is inconsistent with the behaviour of the R357K mutant. A problem with these results is the low signal to noise from the pY291 antibody. Noise may have significantly influenced band intensity measurements, especially for the WASp double mutant where in the 5 min sample significant noise can be seen. This experiment should be repeated, possibly using stable cell lines expressing WASp mutants (generated with lentiviral vectors) so that expression levels can be matched.

The 4G10 antibody could be used to detect WASp phosphorylation.

Further evidence for the functional importance of the R357 methylation site comes from a study of the mutations found in patients suffering from WAS. Mutations in the WASp PPP domain were found in several patients. One of these was a 5 year old boy from Turkey who had a C1115A mutation in the *WAS* gene resulting in a P361T missense mutation in WASp (Jin et al., 2004). P361 is the proline in the PRxxxP motif at the R357 methylation site. The fact that the patient presented with severe WAS symptoms usually associated with absence of WASp, points to an important functional effect of the nearby arginine methylation site and also demonstrates the functional significance of the PRxxxP motif discovered here.

The results shown here suggest an important role for methylation at R357 but there is no evidence for an important role of methylation at R338. The PRxxxP motif at R357 is clearly important for the positive regulation of WASp, but the exact consequences of methylation in this region are less certain. The R357K mutation appeared to negatively impact WASp activation which can be interpreted in two ways. Either the methylation is required for binding to an interaction partner leading to recruitment and activation of WASp, or the arginine residue alone is required and mutation to lysine is enough to disrupt an important interaction (e.g. to Fyn). This highlights the limitations of assessing the functional significance of a PTM site by mutation; it is only possible to conclude that the residue is important, no direct conclusion can be reached regarding the PTM. However, in the context of other results, for example using methylation inhibitors to suppress methylation of WASp, these experiments can provide supporting evidence. The use

of methylation inhibitors also has disadvantages, inhibitors currently available are not very specific for arginine methylation and cells may need to be treated for extended periods to generate a non-methylated pool of protein (depending on the turnover rate of the protein). This can lead to off-target effects making results difficult to interpret. One way of using inhibitors could be to transduce cells with an inducible construct encoding WASp. Expression of WASp could be induced for a short time while applying methylation inhibitors. Any WASp synthesised during this period would be non-methylated and its function could be analysed.

A role for protein methylation in actin cytoskeleton dynamics is not without precedent as actin is known to be methylated at histidine 73 which is thought to be necessary for proper coupling of ATP hydrolysis and actin filament formation (Nyman et al., 2002). Lysine methylation may also be involved in actin polymerisation as a study has shown the presence of a lysine methyltransferase complex in the cytosol and its importance in TCR-induced actin polymerisation, however, to date no nucleation promoting factors have been shown to be lysine methylated (Su et al., 2005). Here, three factors involved in actin polymerisation have been shown to be arginine methylated, indicating a role for this PTM in regulation of the cytoskeleton.

Arginine methylation of WIPF1 in the PRxxxP motif at R211 decreases as a result of T-lymphocyte activation suggesting that methylation of this protein is a negative regulator. Similarly, one possibility is that methylation at R357 negatively regulates WASp function at the IS by disrupting its interaction with an SH3 domain containing protein such as Fyn or Nck. This arginine methylated portion of WASp might be degraded or it might be free

to enter the nucleus to play a role there. Arginine methylation may be a means to stably divide WASp into different functional pools. Further experiments are required to determine whether methylation of WASp is altered after activation of T-lymphocytes, which PRMT is responsible for methylation, which interactions are modulated by methylation and what impact this has on T-lymphocyte function. Such experiments could include immunoprecipitating WASp before and after T-lymphocyte activation to compare methylation status and performing peptide pull-down experiments with non-methylated and methylated peptides corresponding to the R357 methylation site.

6. Discussion

The development of the Jurkat T-lymphocyte and other cell lines over three decades ago, enabled the dissection of early signalling events driving T-lymphocyte activation (Abraham and Weiss, 2004). As a result of intensive study, many of the components of the TCR-signalling pathway are known and this knowledge has led to direct therapeutic benefits. Several drugs and antibodies have been developed to interfere with T-lymphocyte activation by modulating the TCR-signalling pathway. A well known example is cyclosporin which inhibits calcineurin preventing nuclear translocation of NFAT and activation of cytokine genes. Cyclosporin is therefore used as an immunosuppressant to prevent rejection of a transplanted organ (Clipstone and Crabtree, 1992). OKT3 is an example of a therapeutic antibody against the CD3 signalling component of the TCR complex. This antibody causes excessive stimulation of the TCR signalling pathway leading to temporary T-lymphocyte anergy and is used to block acute rejection of a transplant (Cosimi et al., 1981). As a result of these therapies (and others), transplanted organs can be successfully hosted for many years.

A full understanding of T-lymphocyte activation can therefore lead to significant advances in the treatment of immune-related disorders and immunotherapy of cancers. Much attention has been focused on proximal

events in T-lymphocyte activation, such as induction of phosphorylation and calcium increases. The precise molecular detail of later events is less well understood as later stages of activation depend on a complex interplay of many different signalling pathways and are more difficult to dissect. Here, evidence is presented suggesting a role for arginine methylation in the later stages of T-lymphocyte activation, possibly having a role in T-lymphocyte differentiation to subsets.

Arginine methylation is emerging as an important regulator of protein function in general cell biology. Here the focus was on its role in the activation of T-lymphocytes. The application of mass spectrometry based proteomics in this study led to the characterisation of arginine methylation in T-lymphocytes in unprecedented detail and breadth. The enrichment methods allowed many new arginine methylated proteins to be discovered and revealed that arginine methylation may be involved in regulating a wide variety of cellular processes. Many of the proteins discovered to be methylated in primary T-lymphocytes are widely expressed and likely to be methylated in other cell types. Most proteins whose methylation state was altered upon T-lymphocyte activation are also widely expressed, so the changes in arginine methylation observed by western blotting and mass spectrometry may occur in other cell-types. These alterations in arginine methylation may not be specifically associated with T-lymphocyte activation, instead they may be linked to the cell cycle and cellular proliferation.

6.1 Future developments

Being a relatively stable modification, arginine methylation could be well suited to imparting longer term physico-chemical changes on a protein. The majority of arginine methylated proteins identified here are involved in transcription, mRNA processing or translation. By affecting these processes in the long term, arginine methylation may be an important component of cellular differentiation. The role of arginine methylation in transcriptional and splicing programs in differentiating cells is likely to be a rewarding area for further investigation. The use of a conditional knock-out of PRMT1 in mouse thymocytes could be an appropriate model to provide additional insight into the role of arginine methylation in T-lymphocyte differentiation and activation.

Knowledge of arginine methylation has been substantially extended by this study. However, the chromatographic methods used to identify methylation sites only provide weak enrichment. The total number of arginine methylation sites in a cell is unknown. Early studies on phosphorylation using mass spectrometry resulted in identification of hundreds of phosphorylation sites (Ficarro et al., 2002), now identification of many thousands is routine (Wu et al., 2011). The protein kinases are the largest enzyme family in the human genome (kinases are coded by approximately 500 genes). The number of PRMTs is significantly less than this making it likely that arginine methylation is a more specialised and restricted modification. Development of new enrichment reagents such as antibodies should enable identification of many more sites.

Arginine methylation may have a significant role in several aspects of human health. Studies have linked arginine methylation with ageing (Takahashi et al., 2011; Mitchell et al., 2009), cancer (Powers et al., 2011; Yoshimatsu et al., 2011), cardiovascular disease (Leiper et al., 2007), neurodegeneration (Friesen et al., 2001), viral pathogenesis (Iacovides et al., 2007) and immunity (Blanchet et al. (2005); Bonham et al. (2010); Lawson et al. (2007) and this work). In recent years, several improved PRMT inhibitors have been developed (Bonham et al., 2010; Wang et al., 2012) making it possible that modulation of arginine methylation could be a viable therapeutic strategy in various human disorders including those relating to T-lymphocytes.

6. Bibliography

Abdul-Manan, N., Aghazadeh, B., Liu, G. A., Majumdar, A., Ouerfelli, O., Siminovitch, K. A., and Rosen, M. K. (1999). Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. *Nature*, 399(6734):379–383.

Abraham, R. T. and Weiss, A. (2004). Jurkat T cells and development of the T-cell receptor signalling paradigm. *Nat Rev Immunol*, 4(4):301–308.

Abramovich, C., Yakobson, B., Chebath, J., and Revel, M. (1997). A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor. *EMBO J*, 16(2):260–266.

Acuto, O., Di Bartolo, V., and Michel, F. (2008). Tailoring T-cell receptor signals by proximal negative feedback mechanisms. *Nat Rev Immunol*, 8(9):699–712.

Acuto, O. and Michel, F. (2003). CD28-mediated co-stimulation: a quantitative support for TCR signalling. *Nat Rev Immunol*, 3(12):939–951.

Akbar, A. N., Timms, A., and Janossy, G. (1989). Cellular events during memory T-cell activation in vitro: the UCHL1 (180,000 MW) determinant is newly synthesized after mitosis. *Immunology*, 66(2):213–218.

Anderson, P. (2010). Post-transcriptional regulons coordinate the initiation and resolution of inflammation. *Nat Rev Immunol*, 10(1):24–35.

Andreu-Pérez, P., Esteve-Puig, R., de Torre-Minguela, C., López-Fauqued, M., Bech-Serra, J. J., Tenbaum, S., García-Trevijano, E. R., Canals, F., Merlino, G., Avila, M. A., and Recio, J. A. (2011). Protein arginine methyltransferase 5 regulates ERK1/2 signal transduction amplitude and cell fate through CRAF. *Sci Signal*, 4(190):1–14.

Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clin Chem*, 49(7):1041–1044.

Antón, I. M., Lu, W., Mayer, B. J., Ramesh, N., and Geha, R. S. (1998). The Wiskott-Aldrich syndrome protein-interacting protein (WIP) binds to the adaptor protein Nck. *J Biol Chem*, 273(33):20992–20995.

Aston, F. (1919). *Philosophical Magazine*, 38:707.

Badour, K., Zhang, J., Shi, F., Leng, Y., Collins, M., and Siminovitch, K. A. (2004). Fyn and PTP-PEST-mediated regulation of Wiskott-Aldrich syndrome protein (WASp) tyrosine phosphorylation is required for coupling T cell antigen receptor engagement to WASp effector function and T cell activation. *J Exp Med*, 199(1):99–112.

Baldwin, G. and Carnegie, P. (1971). Isolation and partial characterization of methylated arginines from the encephalitogenic basic protein of myelin. *Biochemical Journal*, 123(1):69.

Baldwin, M. A. (2004). Protein identification by mass spectrometry: issues to be considered. *Mol Cell Proteomics*, 3(1):1–9.

Barnhart, B., Kosinski, P. A., Wang, Z., Ford, G. S., Kiledjian, M., and Covey, L. R. (2000). Identification of a complex that binds to the CD154 3' untranslated region: implications for a role in message stability during T cell activation. *J Immunol*, 165(8):4478–4486.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837.

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A*, 101(33):12130–12135.

Bedford, M. T. and Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. *Mol Cell*, 33(1):1–13.

Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000). Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J Biol Chem*, 275(21):16030–16036.

Beemiller, P. and Krummel, M. F. (2010). Mediation of T-cell activation by actin meshworks. *Cold Spring Harb Perspect Biol*, 2(9):1–14.

Beltrao, P., Albanèse, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villén, J., Lim, W. A., Fraser, J. S., Frydman, J., and Krogan, N. J. (2012). Systematic functional prioritization of protein posttranslational modifications. *Cell*, 150(2):413–425.

Bikkavilli, R. K. and Malbon, C. C. (2011). Arginine methylation of G3BP1 in response to Wnt3a regulates -catenin mRNA. *J Cell Sci*, 124(Pt 13):2310–2320.

Blanchet, F., Cardona, A., Letimier, F. A., Hershfield, M. S., and Acuto, O. (2005). CD28 costimulatory signal induces protein arginine methylation in T cells. *J Exp Med*, 202(3):371–377.

Blundell, M. P., Bouma, G., Metelo, J., Worth, A., Calle, Y., Cowell, L. A., Westerberg, L. S., Moulding, D. A., Mirando, S., Kinnon, C., Cory, G. O., Jones, G. E., Snapper, S. B., Burns, S. O., and Thrasher, A. J. (2009). Phosphorylation of WASp is a key regulator of activity and stability in vivo. *Proc Natl Acad Sci U S A*, 106(37):15738–15743.

Bodenmiller, B., Zunder, E. R., Finck, R., Chen, T. J., Savig, E. S., Bruggner, R. V., Simonds, E. F., Bendall, S. C., Sachs, K., Krutzik, P. O., and Nolan, G. P. (2012). Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol*, pages 858–867.

Boffa, L. C., Karn, J., Vidali, G., and Allfrey, V. G. (1977). Distribution of NG, NG,-dimethylarginine in nuclear protein fractions. *Biochem Biophys Res Commun*, 74(3):969–976.

Boisvert, F. M., Cote, J., Boulanger, M. C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002). Symmetrical dimethylarginine methylation is required for the localization of SMN in cajal bodies and pre-mRNA splicing. *J Cell Biol*, 159(6):957–969.

Boisvert, F. M., Côté, J., Boulanger, M. C., and Richard, S. (2003). A proteomic analysis of arginine-methylated protein complexes. *Mol Cell Proteomics*, 2(12):1319–1330.

Boisvert, F. M., Déry, U., Masson, J. Y., and Richard, S. (2005). Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev*, 19(6):671–676.

Bonham, K., Hemmers, S., Lim, Y. H., Hill, D. M., Finn, M. G., and Mowen, K. A. (2010). Effects of a novel arginine methyltransferase inhibitor on T-helper cell cytokine production. *FEBS J*, 277(9):2096–2108.

Boomer, J. S. and Green, J. M. (2010). An enigmatic tail of CD28 signaling. *Cold Spring Harb Perspect Biol*, 2(8).

Borders, C. L., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., and Pett, V. B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci.*, 3(4):541–548.

Brockmeyer, C., Paster, W., Pepper, D., Tan, C. P., Trudgian, D. C., McGowan, S., Fu, G., Gascoigne, N. R., Acuto, O., and Salek, M. (2011). T cell receptor (TCR)-induced tyrosine phosphorylation dynamics identifies THEMIS as a new TCR signalosome component. *J Biol Chem*, 286(9):7535–7547.

Bruins, C., Jeronimus-Stratingh, C., Ensing, K., Van Dongen, W., and De Jong, G. (1999). On-line coupling of solid-phase extraction with mass spectrometry for the analysis of biological samples: I. determination of clenbuterol in urine. *Journal of Chromatography A*, 863(1):115–122.

Bruns, A. F., Grothe, C., and Claus, P. (2009). Fibroblast growth factor 2 (FGF-2) is a novel substrate for arginine methylation by PRMT5. *Biol Chem*, 390(1):59–65.

Cannon, J. L., Labno, C. M., Bosco, G., Seth, A., McGavin, M. H., Siminovitch, K. A., Rosen, M. K., and Burkhardt, J. K. (2001). Wasp recruitment to the T cell:APC contact site occurs independently of Cdc42 activation. *Immunity*, 15(2):249–259.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R. K. (2007). JMJD6 is a histone arginine demethylase. *Science*, 318(5849):444–447.

Chang, F. N., Navickas, I. J., Au, C., and Budzilowicz, C. (1978). Identification of the methylated ribosomal proteins in HeLa cells and the fluctuation of methylation during the cell cycle. *Biochim Biophys Acta*, 518(1):89–94.

Chang, Y. I., Lin, S. W., Chiou, Y. Y., Sung, J. S., Cheng, L. C., Lu, Y. L., Sun, K. H., Chang, K., Lin, C. H., and Lin, W. J. (2010). Establishment of an ectopically expressed and functional PRMT1 for proteomic analysis of arginine-methylated proteins. *Electrophoresis*, 31(23-24):3834–3842.

Chen, C., Nott, T. J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol*, 12(10):629–642.

Chen, D. H., Wu, K. T., Hung, C. J., Hsieh, M., and Li, C. (2004). Effects of adenosine dialdehyde treatment on in vitro and in vivo stable protein methylation in HeLa cells. *J Biochem*, 136(3):371–376.

Chen, H., Xue, Y., Huang, N., Yao, X., and Sun, Z. (2006). MeMo: a web tool for prediction of protein methylation modifications. *Nucleic acids research*, 34(suppl 2):W249–W253.

Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. (2007). The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell*, 25(1):71–83.

Cheng, X., Collins, R. E., and Zhang, X. (2005). Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct*, 34:267–294.

Cheung, W. C., Beausoleil, S. A., Zhang, X., Sato, S., Schieferl, S. M., Wieler, J. S., Beaudet, J. G., Ramenani, R. K., Popova, L., Comb, M. J., Rush, J., and Polakiewicz, R. D. (2012). A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol*, 30(5):447–452.

Cho, J. N., Ryu, J. Y., Jeong, Y. M., Park, J., Song, J. J., Amasino, R. M., Noh, B., and Noh, Y. S. (2012). Control of seed germination by light-induced histone arginine demethylation activity. *Dev Cell*, 22(4):736–748.

Cimato, T. R., Tang, J., Xu, Y., Guarnaccia, C., Herschman, H. R., Pongor, S., and Aletta, J. M. (2002). Nerve growth factor-mediated increases in protein methylation occur predominantly at type I arginine methylation sites and involve protein arginine methyltransferase 1. *J Neurosci Res*, 67(4):435–442.

Claverie, J.-M. (2001). What if there are only 30,000 human genes? *Science*, 291(5507):1255–1257.

Clipstone, N. A. and Crabtree, G. R. (1992). Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature*, 357(6380):695–697.

Cook, J. R., Lee, J. H., Yang, Z. H., Krause, C. D., Herth, N., Hoffmann, R., and Pestka, S. (2006). FBXO11/PRMT9, a new protein arginine methyltransferase, symmetrically dimethylates arginine residues. *Biochem Biophys Res Commun*, 342(2):472–481.

Cory, G. O., Cramer, R., Blanchoin, L., and Ridley, A. J. (2003). Phosphorylation of the WASP-VCA domain increases its affinity for the Arp2/3 complex and enhances actin polymerization by WASP. *Mol Cell*, 11(5):1229–1239.

Cosimi, A. B., Burton, R. C., Colvin, R. B., Goldstein, G., Delmonico, F. L., LaQuaglia, M. P., Tolkoff-Rubin, N., Rubin, R. H., Herrin, J. T., and Russell, P. S. (1981). Treatment of acute renal allograft rejection with OKT3 monoclonal antibody. *Transplantation*, 32(6):535–539.

Côté, J., Boisvert, F. M., Boulanger, M. C., Bedford, M. T., and Richard, S. (2003). Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Mol Biol Cell*, 14(1):274–287.

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, 14(6):1188–1190.

Daily, K., Radivojac, P., and Dunker, A. (2005). Intrinsic disorder and protein modifications: Building an SVM predictor for methylation. In *Computational Intelligence in Bioinformatics and Computational Biology, 2005. CIBCB'05. Proceedings of the 2005 IEEE Symposium on*, pages 1–7. IEEE.

Dalloneau, E., Pereira, P. L., Brault, V., Nabel, E. G., and Hérault, Y. (2011). Prmt2 regulates the lipopolysaccharide-induced responses in lungs and macrophages. *J Immunol*, 187(9):4826–4834.

Davis, S. J. and van der Merwe, P. A. (2006). The kinetic-segregation model: TCR triggering and beyond. *Nat Immunol*, 7(8):803–809.

DeFranco, A. L., Locksley, R. M., and Robertson, M. (2007). *Immunity*. New Science Press, first edition.

Dempster, A. (1918). *Physical Review*, 11:316.

Denman, R. B. (2002). Methylation of the arginine-glycine-rich region in the fragile Xmental retardation protein FMRP differentially affects RNA binding. *Cell Mol Biol Lett*, 7(3):877–883.

Derry, J. M., Ochs, H. D., and Francke, U. (1994). Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell*, 78(4):635–644.

Dong, C. (2011). Genetic controls of Th17 cell differentiation and plasticity. *Exp Mol Med*, 43(1):1–6.

Druker, B. J., Mamon, H. J., and Roberts, T. M. (1989). Oncogenes, growth factors, and signal transduction. *N Engl J Med*, 321(20):1383–1391.

Duan, P., Xu, Y., Birkaya, B., Myers, J., Pelletier, M., Read, L. K., Guarnaccia, C., Pongor, S., Denman, R. B., and Aletta, J. M. (2007). Generation of polyclonal antiserum for the detection of methylarginine proteins. *J Immunol Methods*, 320(1-2):132–142.

Dustin, M. L. and Depoil, D. (2011). New insights into the T cell synapse from single molecule techniques. *Nat Rev Immunol*, 11(10):672–684.

Ea, C. K., Deng, L., Xia, Z. P., Pineda, G., and Chen, Z. J. (2006). Activation of IKK by TNFalpha requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Mol Cell*, 22(2):245–257.

Edman, P. (1950). Method for determination of the amino acid sequence in peptides. *Acta chem. scand*, 4(283-293):7.

El-Andaloussi, N., Valovka, T., Toueille, M., Hassa, P. O., Gehrig, P., Covic, M., Hübscher, U., and Hottiger, M. O. (2007). Methylation of DNA polymerase beta by protein arginine methyltransferase 1 regulates its binding to proliferating cell nuclear antigen. *FASEB J*, 21(1):26–34.

Elias, J. E. and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, 4(3):207–214.

Fackelmayer, F. O. (2005). Protein arginine methyltransferases: guardians of the Arg? *Trends Biochem Sci*, 30(12):666–671.

Feng, Q., He, B., Jung, S. Y., Song, Y., Qin, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (2009). Biochemical control of CARM1 enzymatic activity by phosphorylation. *J Biol Chem*, 284(52):36167–36174.

Feng, Y., Wang, J., Asher, S., Hoang, L., Guardiani, C., Ivanov, I., and Zheng, Y. G. (2011). Histone H4 acetylation differentially modulates arginine methylation by an in Cis mechanism. *J Biol Chem*, 286(23):20323–20334.

Feuerer, M., Hill, J. A., Mathis, D., and Benoist, C. (2009). Foxp3+ regulatory T cells: differentiation, specification, subphenotypes. *Nat Immunol*, 10(7):689–695.

Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol*, 20(3):301–305.

Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol Cell Biol*, 21(24):8289–8300.

Fronz, K., Güttinger, S., Burkert, K., Kühn, U., Stöhr, N., Schierhorn, A., and Wahle, E. (2011). Arginine methylation of the nuclear poly(a) binding protein weakens the interaction with its nuclear import receptor, transportin. *J Biol Chem*, 286(38):32986–32994.

Fujiwara, T., Mori, Y., Chu, D. L., Koyama, Y., Miyata, S., Tanaka, H., Yachi, K., Kubo, T., Yoshikawa, H., and Tohyama, M. (2006). CARM1 regulates proliferation of PC12 cells by methylating HuD. *Mol Cell Biol*, 26(6):2273–2285.

Galichet, A., Hoyerová, K., Kamínek, M., and Gruissem, W. (2008). Farnesylation directs AtIPT3 subcellular localization and modulates cytokinin biosynthesis in Arabidopsis. *Plant Physiol*, 146(3):1155–1164.

Ganesh, L., Yoshimoto, T., Moorthy, N. C., Akahata, W., Boehm, M., Nabel, E. G., and Nabel, G. J. (2006). Protein methyltransferase 2 inhibits NF-kappaB function and promotes apoptosis. *Mol Cell Biol*, 26(10):3864–3874.

Gaspar, H. B., Aiuti, A., Porta, F., Candotti, F., Hershfield, M. S., and Notarangelo, L. D. (2009). How I treat ada deficiency. *Blood*, 114(17):3524–3532.

German, D. C., Bloch, C. A., and Kredich, N. M. (1983). Measurements of s-adenosylmethionine and L-homocysteine metabolism in cultured human lymphoid cells. *J Biol Chem*, 258(18):10997–11003.

Gonsalvez, G. B., Praveen, K., Hicks, A. J., Tian, L., and Matera, A. G. (2008). Sm protein methylation is dispensable for snRNP assembly in Drosophila melanogaster. *RNA*, 14(5):878–887.

Gonsalvez, G. B., Tian, L., Ospina, J. K., Boisvert, F. M., Lamond, A. I., and Matera, A. G. (2007). Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. *J Cell Biol*, 178(5):733–740.

Goulet, I., Gauvin, G., Boisvenue, S., and Côté, J. (2007). Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization. *J Biol Chem*, 282(45):33009–33021.

Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Lüscher, B., and Amati, B. (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature*, 449(7164):933–937.

Guo, A. (2012). MethylScan: Proteomic tool to study protein methylation. In *American Society for Mass Spectrometry*.

He, W., Ma, X., Yang, X., Zhao, Y., Qiu, J., and Hang, H. (2011). A role for the arginine methylation of Rad9 in checkpoint control and cellular sensitivity to DNA damage. *Nucleic Acids Res*, 39(11):4719–4727.

Henriksen, P., Wagner, S. A., Weinert, B. T., Sharma, S., Bacinskaja, G., Rehman, M., Juffer, A. H., Walther, T. C., Lisby, M., and Choudhary, C. (2012). Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in Saccharomyces cerevisiae. *Mol Cell Proteomics*, pages 1510–1522.

Herrmann, F., Pably, P., Eckerich, C., Bedford, M. T., and Fackelmayer, F. O. (2009). Human protein arginine methyltransferases in vivo–distinct properties of eight canonical members of the PRMT family. *J Cell Sci*, 122(Pt 5):667–677.

Higgs, H. N. and Pollard, T. D. (2000). Activation by Cdc42 and PIP(2) of Wiskott-Aldrich syndrome protein (WASp) stimulates actin nucleation by Arp2/3 complex. *J Cell Biol*, 150(6):1311–1320.

Hong, S., Song, H. R., Lutz, K., Kerstetter, R. A., Michael, T. P., and McClung, C. R. (2010). Type II protein arginine methyltransferase 5 (PRMT5) is required for circadian period determination in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 107(49):21211–21216.

Horn, D. M., Zubarev, R. A., and McLafferty, F. W. (2000). Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci U S A*, 97(19):10313–10317.

Hsu, J. M., Chen, C. T., Chou, C. K., Kuo, H. P., Li, L. Y., Lin, C. Y., Lee, H. J., Wang, Y. N., Liu, M., Liao, H. W., Shi, B., Lai, C. C., Bedford, M. T., Tsai, C. H., and Hung, M. C. (2011). Crosstalk between Arg 1175 methylation and Tyr 1173 phosphorylation negatively modulates EGFR-mediated ERK activation. *Nat Cell Biol*, 13(2):174–181.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9(3):90–95.

Hyun, S., Jeong, S., and Yu, J. (2008). Effects of asymmetric arginine dimethylation on RNA-binding peptides. *Chembiochem*, 9(17):2790–2792.

Iacovides, D. C., O'Shea, C. C., Oses-Prieto, J., Burlingame, A., and McCormick, F. (2007). Critical role for arginine methylation in adenovirus-infected cells. *J Virol*, 81(23):13209–13217.

Infantino, S., Benz, B., Waldmann, T., Jung, M., Schneider, R., and Reth, M. (2010). Arginine methylation of the B cell antigen receptor promotes differentiation. *J Exp Med*, 207(4):711–719.

Jansson, M., Durant, S. T., Cho, E. C., Sheahan, S., Edelmann, M., Kessler, B., and La Thangue, N. B. (2008). Arginine methylation regulates the p53 response. *Nat Cell Biol*, 10(12):1431–1439.

Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., and Gygi, S. P. (2011). Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics*, 10(12):99–111.

Jian, W., Edom, R. W., Xu, Y., and Weng, N. (2010). Recent advances in application of hydrophilic interaction chromatography for quantitative bioanalysis. *J Sep Sci*, 33(6-7):681–697.

Jiang, X. and Chen, Z. J. (2012). The role of ubiquitylation in immune defence and pathogen evasion. *Nat Rev Immunol*, 12(1):35–48.

Jin, Y., Mazza, C., Christie, J. R., Giliani, S., Fiorini, M., Mella, P., Gandellini, F., Stewart, D. M., Zhu, Q., Nelson, D. L., Notarangelo, L. D., and Ochs, H. D. (2004). Mutations of the Wiskott-Aldrich syndrome protein (WASP): hotspots, effect on transcription, and translation and phenotype/genotype correlation. *Blood*, 104(13):4010–4019.

Jones, A. W. and Cooper, H. J. (2011). Dissociation techniques in mass spectrometry-based proteomics. *Analyst*, 136(17):3419–3429.

Kaizuka, Y., Douglass, A. D., Varma, R., Dustin, M. L., and Vale, R. D. (2007). Mechanisms for segregating T cell receptor and adhesion molecules during immunological synapse formation in Jurkat T cells. *Proc Natl Acad Sci U S A*, 104(51):20296–20301.

Kaneko, T., Hosokawa, H., Yamashita, M., Wang, C. R., Hasegawa, A., Kimura, M. Y., Kitajiama, M., Kimura, F., Miyazaki, M., and Nakayama, T. (2007). Chromatin remodeling at the Th2 cytokine gene loci in human type 2 helper T cells. *Mol Immunol*, 44(9):2249–2256.

Kang, X. and Frey, D. D. (2004). Chromatofocusing of peptides and proteins using linear ph gradients formed on strong ion-exchange adsorbents. *Biotechnol Bioeng*, 87(3):376–387.

Katz, J. E., Dlakić, M., and Clarke, S. (2003). Automated identification of putative methyltransferases from genomic open reading frames. *Molecular & Cellular Proteomics*, 2(8):525–540.

Kessler, B. M. (2006). Putting proteomics on target: activity-based profiling of ubiquitin and ubiquitin-like processing enzymes. *Expert Rev Proteomics*, 3(2):213–221.

Kim, C., Lim, Y., Yoo, B. C., Won, N. H., Kim, S., and Kim, G. (2010a). Regulation of post-translational protein arginine methylation during HeLa cell cycle. *Biochim Biophys Acta*, 1800(9):977–985.

Kim, D., Lee, J., Cheng, D., Li, J., Carter, C., Richie, E., and Bedford, M. T. (2009). Enzymatic activity is required for the *in Vivo* functions of CARM1. *The journal of biological chemistry*, 285(2):1147–1152.

Kim, J., Lee, J., Yadav, N., Wu, Q., Carter, C., Richard, S., Richie, E., and Bedford, M. T. (2004). Loss of CARM1 results in hypomethylation of thymocyte cyclic AMP-regulated phosphoprotein and deregulated early T cell development. *J Biol Chem*, 279(24):25339–25344.

Kim, S., Gupta, N., and Pevzner, P. A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res*, 7(8):3354–3363.

Kim, S., Mischerikow, N., Bandeira, N., Navarro, J. D., Wich, L., Mohammed, S., Heck, A. J., and Pevzner, P. A. (2010b). The generating

function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics*, 9(12):2840–2852.

Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bähler, J., Green, R. D., and Kouzarides, T. (2007). Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature*, 449(7164):928–932.

Kisielow, J., Nairn, A. C., and Karjalainen, K. (2001). TARPP, a novel protein that accompanies TCR gene rearrangement and thymocyte education. *Eur J Immunol*, 31(4):1141–1149.

Kölbel, K., Ihling, C., Kühn, U., Neundorf, I., Otto, S., Stichel, J., Robaa, D., Beck-Sickinger, A. G., Sinz, A., and Wahle, E. (2012). Peptide backbone conformation affects the substrate preference of protein arginine methyltransferase I. *Biochemistry*, 51(27):5463–5475.

König, H., Ponta, H., and Herrlich, P. (1998). Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J*, 17(10):2904–2913.

Krause, C. D., Yang, Z. H., Kim, Y. S., Lee, J. H., Cook, J. R., and Pestka, S. (2007). Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacol Ther*, 113(1):50–87.

Kuhn, P., Chumanov, R., Wang, Y., Ge, Y., Burgess, R. R., and Xu, W. (2011). Automethylation of CARM1 allows coupling of transcription and mRNA splicing. *Nucleic Acids Res*, 39(7):2717–2726.

Kühn, U., Nemeth, A., Meyer, S., and Wahle, E. (2003). The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem*, 278(19):16916–16925.

Kurisu, S. and Takenawa, T. (2009). The WASP and WAVE family proteins. *Genome Biol*, 10(6):226–226.

Kyte, J., Doolittle, R., et al. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1):105–132.

Lakowski, T. M. and Frankel, A. (2009). Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4. *Biochem J*, 421(2):253–261.

Lanzi, G., Moratto, D., Vairo, D., Masneri, S., Delmonte, O., Paganini, T., Parolini, S., Tabellini, G., Mazza, C., Savoldi, G., Montin, D., Martino, S., Tovo, P., Pessach, I. M., Massaad, M. J., Ramesh, N., Porta, F., Plebani, A., Notarangelo, L. D., Geha, R. S., and Giliani, S. (2012). A novel primary human immunodeficiency due to deficiency in the WASP-interacting protein WIP. *J Exp Med*, 209(1):29–34.

Lawson, B. R., Manenkova, Y., Ahamed, J., Chen, X., Zou, J. P., Baccala, R., Theofilopoulos, A. N., and Yuan, C. (2007). Inhibition of transmethylation down-regulates CD4 T cell activation and curtails development of autoimmunity in a model system. *J Immunol*, 178(8):5366–5374.

Le Bras, S., Massaad, M., Koduru, S., Kumar, L., Oyoshi, M. K., Hartwig, J., and Geha, R. S. (2009). WIP is critical for T cell responsiveness to IL-2. *Proc Natl Acad Sci U S A*, 106(18):7519–7524.

Le Romancer, M., Treilleux, I., Leconte, N., Robin-Lespinasse, Y., Sentis, S., Bouchekioua-Bouzaghou, K., Goddard, S., Gobert-Gosse, S., and Corbo, L. (2008). Regulation of estrogen rapid signaling through arginine methylation by PRMT1. *Mol Cell*, 31(2):212–221.

Lee, H. W., Paik, W. K., and Borun, T. W. (1973). The periodic synthesis of S-adenosylmethionine: protein methyltransferases during the HeLa S-3 cell cycle. *J Biol Chem*, 248(12):4194–4199.

Lee, J., Sayegh, J., Daniel, J., Clarke, S., and Bedford, M. T. (2005). PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family. *J Biol Chem*, 280(38):32890–32896.

Leiper, J., Nandi, M., Torondel, B., Murray-Rust, J., Malaki, M., O'Hara, B., Rossiter, S., Anthony, S., Madhani, M., Selwood, D., Smith, C., Wojciak-Stothard, B., Rudiger, A., Stidwill, R., McDonald, N. Q., and Vallance, P. (2007). Disruption of methylarginine metabolism impairs vascular homeostasis. *Nat Med*, 13(2):198–203.

Li, H., Kobayashi, M., Blonska, M., You, Y., and Lin, X. (2006). Ubiquitination of RIP is required for tumor necrosis factor alpha-induced NF-kappaB activation. *J Biol Chem*, 281(19):13636–13643.

Li, H., Park, S., Kilburn, B., Jelinek, M., Henschen-Edman, A., Aswad, D., Stallcup, M., and Laird-Offringa, I. (2002). Lipopolysaccharide-induced methylation of HuR, an mRNA-stabilizing protein, by CARM1. *Science's STKE*, 277(47):44623–44630.

Li, P., Wang, D., Yao, H., Doret, P., Hao, G., Shen, Q., Qiu, H., Zhang, X., Wang, Y., Chen, G., and Wang, Y. (2010). Coordination of PAD4 and HDAC2 in the regulation of p53-target gene expression. *Oncogene*, 29(21):3153–3162.

Li, S. S. (2005). Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem J*, 390(Pt 3):641–653.

Li, Z. and Stock, J. B. (2009). Protein carboxyl methylation and the biochemistry of memory. *Biol Chem*, 390(11):1087–1096.

Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S., and Herschman, H. R. (1996). The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J Biol Chem*, 271(25):15034–15044.

Lindstein, T., June, C. H., Ledbetter, J. A., Stella, G., and Thompson, C. B. (1989). Regulation of lymphokine messenger RNA stability by a surface-mediated T cell activation pathway. *Science*, 244(4902):339–343.

Liu, K., Chen, C., Guo, Y., Lam, R., Bian, C., Xu, C., Zhao, D. Y., Jin, J., MacKenzie, F., Pawson, T., and Min, J. (2010). Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci U S A*, 107(43):18398–18403.

Liu, K., Guo, Y., Liu, H., Bian, C., Lam, R., Liu, Y., Mackenzie, F., Rojas, L. A., Reinberg, D., Bedford, M. T., Xu, R. M., and Min, J. (2012). Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One*, 7(2):e30375.

Liu, Q. and Dreyfuss, G. (1995). In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol*, 15(5):2800–2808.

Loyet, K. M., Ouyang, W., Eaton, D. L., and Stults, J. T. (2005). Proteomic profiling of surface proteins on Th1 and Th2 cells. *J Proteome Res*, 4(2):400–409.

Lynch, K. W. (2004). Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol*, 4(12):931–940.

Lynch, K. W. and Weiss, A. (2000). A model system for activation-induced alternative splicing of CD45 pre-mRNA in T cells implicates protein kinase C and Ras. *Mol Cell Biol*, 20(1):70–80.

Mann, M. and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nat Biotechnol*, 21(3):255–261.

Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M. T., and Milcarek, C. (1998). Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. *Proc Natl Acad Sci U S A*, 95(19):11095–11100.

Martinez, N. M., Pan, Q., Cole, B. S., Yarosh, C. A., Babcock, G. A., Heyd, F., Zhu, W., Ajith, S., Blencowe, B. J., and Lynch, K. W. (2012). Alternative splicing networks regulated by signaling in human T cells. *RNA*, 18(5):1029–1040.

McAlister, G. C., Phanstiel, D. H., Brumbaugh, J., Westphall, M. S., and Coon, J. J. (2011). Higher-energy collision-activated dissociation without a dedicated collision cell. *Mol Cell Proteomics*, 10(5):1–6.

McBride, A. E., Weiss, V. H., Kim, H. K., Hogle, J. M., and Silver, P. A. (2000). Analysis of the yeast arginine methyltransferase Hmt1p/Rmt1p and its in vivo function. cofactor binding and substrate interactions. *J Biol Chem*, 275(5):3128–3136.

McDonald, C. B., Seldeen, K. L., Deegan, B. J., Bhat, V., and Farooq, A. (2011). Binding of the cSH3 domain of Grb2 adaptor to two distinct RXXK motifs within Gab1 docker employs differential mechanisms. *J Mol Recognit*, 24(4):585–596.

McNeill, L., Cassady, R. L., Sarkardei, S., Cooper, J. C., Morgan, G., and Alexander, D. R. (2004). CD45 isoforms in T cell signalling and development. *Immunol Lett*, 92(1-2):125–134.

McNeill, L., Salmond, R. J., Cooper, J. C., Carret, C. K., Cassady-Cain, R. L., Roche-Molina, M., Tandon, P., Holmes, N., and Alexander, D. R. (2007). The differential regulation of Lck kinase phosphorylation sites by CD45 is critical for T cell receptor signaling responses. *Immunity*, 27(3):425–437.

McNulty, D. E. and Annan, R. S. (2008). Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics*, 7(5):971–980.

Mempel, T. R., Henrickson, S. E., and Von Andrian, U. H. (2004). T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature*, 427(6970):154–159.

Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics*, 10(9):1–15.

Michel, F., Attal-Bonnefoy, G., Mangino, G., Mise-Omata, S., and Acuto, O. (2001). CD28 as a molecular amplifier extending TCR ligation and signaling capabilities. *Immunity*, 15(6):935–945.

Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J., Girault, H. H., and Rossier, J. S. (2003). Protein fractionation in a multicompartment device using Off-gel isoelectric focusing. *Electrophoresis*, 24(1-2):3–11.

Migliori, V., Müller, J., Phalke, S., Low, D., Bezzi, M., Mok, W. C., Sahu, S. K., Gunaratne, J., Capasso, P., Bassi, C., Cecatiello, V., De Marco, A., Blackstock, W., Kuznetsov, V., Amati, B., Mapelli, M., and Guccione, E. (2012). Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol*, 19(2):136–144.

Miller, M. J., Wei, S. H., Cahalan, M. D., and Parker, I. (2003). Autonomous T cell trafficking examined in vivo with intravital two-photon microscopy. *Proc Natl Acad Sci U S A*, 100(5):2604–2609.

Miller, M. J., Wei, S. H., Parker, I., and Cahalan, M. D. (2002). Two-photon imaging of lymphocyte motility and antigen response in intact lymph node. *Science*, 296(5574):1869–1873.

Miller, S. A., Huang, A. C., Miazgowicz, M. M., Brassil, M. M., and Weinmann, A. S. (2008). Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. *Genes Dev*, 22(21):2980–2993.

Miller, S. A., Mohn, S. E., and Weinmann, A. S. (2010). Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol Cell*, 40(4):594–605.

Mire-Sluis, A. R., Wickremasinghe, R. G., Hoffbrand, A. V., Timms, A. M., and Francis, G. E. (1987). Human T lymphocytes stimulated by phytohaemagglutinin undergo a single round of cell division without a requirement for interleukin-2 or accessory cells. *Immunology*, 60(1):7–12.

Mitchell, T. R., Glenfield, K., Jeyanthan, K., and Zhu, X. D. (2009). Arginine methylation regulates telomere length and stability. *Mol Cell Biol*, 29(18):4918–4934.

Moore, R. E., Young, M. K., and Lee, T. D. (2002). Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom*, 13(4):378–386.

Mowen, K. A., Schurter, B. T., Fathman, J. W., David, M., and Glimcher, L. H. (2004). Arginine methylation of NIP45 modulates cytokine gene expression in effector T lymphocytes. *Mol Cell*, 15(4):559–571.

Murphy, K. (2012). *Janeway's immunobiology*. Garland Science, eighth edition.

Najbauer, J., Johnson, B. A., Young, A. L., and Aswad, D. W. (1967). Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. *Biochemical and Biophysical Research Communications*, 268(14):10501–10509.

Nausch, N., Bourke, C. D., Appleby, L. J., Rujeni, N., Lantz, O., Trottein, F., Midzi, N., Mduluza, T., and Mutapi, F. (2012). Proportions of CD4+ memory T cells are altered in individuals chronically infected with Schistosoma haematobium. *Sci Rep*, 2:472–472.

Neault, M., Mallette, F. A., Vogel, G., Michaud-Levesque, J., and Richard, S. (2012). Ablation of prmt6 reveals a role as a negative transcriptional regulator of the p53 tumor suppressor. *Nucleic Acids Res*, 40(19):9513–9521.

Nishioka, K. and Reinberg, D. (2003). Methods and tips for the purification of human histone methyltransferases. *Methods*, 31(1):49–58.

Nyman, T., Schüler, H., Korenbaum, E., Schutt, C. E., Karlsson, R., and Lindberg, U. (2002). The role of MeH73 in actin polymerization and ATP hydrolysis. *J Mol Biol*, 317(4):577–589.

Oberdoerffer, S., Moita, L. F., Neems, D., Freitas, R. P., Hacohen, N., and Rao, A. (2008). Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science*, 321(5889):686–691.

Obianyo, O. and Thompson, P. R. (2012). Kinetic mechanism of protein arginine methyltransferase 6 (PRMT6). *J Biol Chem*, 287(8):6062–6071.

Ogilvie, R. L., Abelson, M., Hau, H. H., Vlasova, I., Blackshear, P. J., and Bohjanen, P. R. (2005). Tristetraprolin down-regulates IL-2 gene expression through AU-rich element-mediated mRNA decay. *J Immunol*, 174(2):953–961.

Oh-hora, M. and Rao, A. (2008). Calcium signaling in lymphocytes. *Curr Opin Immunol*, 20(3):250–258.

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*, 4(9):709–712.

Olsen, J. V., Ong, S. E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*, 3(6):608–614.

Ong, S. E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods*, 1(2):119–126.

Ostareck-Lederer, A., Ostareck, D. H., Rucknagel, K. P., Schierhorn, A., Moritz, B., Huttelmaier, S., Flach, N., Handoko, L., and Wahle, E. (2006). Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein k by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. *J Biol Chem*, 281(16):11115–11125.

Padrick, S. B., Cheng, H. C., Ismail, A. M., Panchal, S. C., Doolittle, L. K., Kim, S., Skehan, B. M., Umetani, J., Brautigam, C. A., Leong, J. M., and Rosen, M. K. (2008). Hierarchical regulation of WASP/WAVE proteins. *Mol Cell*, 32(3):426–438.

Pahlich, S., Bschir, K., Chiavi, C., Belyanskaya, L., and Gehring, H. (2005). Different methylation characteristics of protein arginine methyltransferase 1 and 3 toward the ewing sarcoma protein and a peptide. *Proteins*, 61(1):164–175.

Paik, W. K. and Kim, S. (1967). Enzymatic methylation of protein fractions from calf thymus nuclei. *Biochemical and Biophysical Research Communications*, 29(1):14–20.

Paizs, B. and Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*, 24(4):508–548.

Pak, M. L., Lakowski, T. M., Thomas, D., Vhuiyan, M. I., Hüsecken, K., and Frankel, A. (2011). A protein arginine N-methyltransferase 1 (PRMT1) and 2 heteromeric interaction increases PRMT1 enzymatic activity. *Biochemistry*, 50(38):8226–8240.

Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P., and Sif, S. (2004). Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol Cell Biol*, 24(21):9630–9645.

Pang, C. N., Gasteiger, E., and Wilkins, M. R. (2010). Identification of arginine- and lysine-methylation in the proteome of Saccharomyces cerevisiae and its functional implications. *BMC Genomics*, 11:92–92.

Park, J. Y., Kob, M., Prodeus, A. P., Rosen, F. S., Shcherbina, A., and Remold-O'Donnell, E. (2004). Early deficit of lymphocytes in Wiskott-Aldrich syndrome: possible role of WASP in human lymphocyte maturation. *Clin Exp Immunol*, 136(1):104–110.

Pawlak, M. R., Scherer, C. A., Chen, J., Roshon, M. J., and Ruley, H. E. (2000). Arginine N-methyltransferase 1 is required for early postimplantation mouse development, but cells deficient in the enzyme are viable. *Mol Cell Biol*, 20(13):4859–4869.

Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J., Swaney, D. L., Tervo, M. A., Bolin, J. M., Ruotti, V., Stewart, R., Thomson, J. A., and Coon, J. J. (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods*, 8(10):821–827.

Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem*, 76(14):3935–3943.

Pintucci, G., Quarto, N., and Rifkin, D. B. (1996). Methylation of high molecular weight fibroblast growth factor-2 determines post-translational

increases in molecular weight and affects its intracellular distribution. *Mol Biol Cell*, 7(8):1249–1258.

Powers, M. A., Fay, M. M., Factor, R. E., Welm, A. L., and Ullman, K. S. (2011). Protein arginine methyltransferase 5 accelerates tumor growth by arginine methylation of the tumor suppressor programmed cell death 4. *Cancer Res*, 71(16):5579–5587.

Rath, A., Davidson, A. R., and Deber, C. M. (2005). The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers*, 80(2-3):179–185.

Rawlings, J. S., Gatzka, M., Thomas, P. G., and Ihle, J. N. (2011). Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence. *EMBO J*, 30(2):263–276.

Reicher, B., Joseph, N., David, A., Pauker, M. H., Perl, O., and Barda-Saad, M. (2012). Ubiquitylation-dependent negative regulation of WASp is essential for actin cytoskeleton dynamics. *Mol Cell Biol*, 32(15):3153–3163.

Richard, S., Morel, M., and Cléroux, P. (2005). Arginine methylation regulates IL-2 gene expression: a role for protein arginine methyltransferase 5 (PRMT5). *Biochem J*, 388(Pt 1):379–386.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., Macneill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., and Comb, M. J. (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*, 131(6):1190–1203.

Roepstorff, P. and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*, 11(11):601–601.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature*, 461(7268):1248–1253.

Rowland-Jones, S. L. and McMichael, A. J. (2000). *Lymphocytes: A Practical Approach*. Oxford University Press, second edition.

Roy, J., Zheng, B., Rymond, B. C., and Woolford, J. L. (1995). Structurally related but functionally distinct yeast Sm D core small nuclear ribonucleoprotein particle proteins. *Mol Cell Biol*, 15(1):445–455.

Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E., and Creamer, T. P. (2003). Host-guest scale of left-handed polyproline II helix formation. *Proteins*, 53(1):68–75.

Sakamaki, J., Daitoku, H., Ueno, K., Hagiwara, A., Yamagata, K., and Fukamizu, A. (2011). Arginine methylation of BCL-2 antagonist of cell death (BAD) counteracts its phosphorylation and inactivation by Akt. *Proc Natl Acad Sci U S A*, 108(15):6085–6090.

Sanchez, S. E., Petrillo, E., Beckwith, E. J., Zhang, X., Rugnone, M. L., Hernando, C. E., Cuevas, J. C., Godoy Herz, M. A., Depetris-Chauvin, A., Simpson, C. G., Brown, J. W., Cerdán, P. D., Borevitz, J. O., Mas, P., Ceriani, M. F., Kornblihtt, A. R., and Yanovsky, M. J. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature*, 468(7320):112–116.

Sayegh, J., Webb, K., Cheng, D., Bedford, M. T., and Clarke, S. G. (2007). Regulation of protein arginine methyltransferase 8 (PRMT8) activity by its N-terminal domain. *J Biol Chem*, 282(50):36444–36453.

Scheu, S., Stetson, D. B., Reinhardt, R. L., Leber, J. H., Mohrs, M., and Locksley, R. M. (2006). Activation of the integrated stress response during T helper cell differentiation. *Nat Immunol*, 7(6):644–651.

Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, 18(20):6097–6100.

Schneider, U., Schwenk, H. U., and Bornkamm, G. (1977). Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer*, 19(5):621–626.

Schulz, E. G., Mariani, L., Radbruch, A., and Höfer, T. (2009). Sequential polarization and imprinting of type 1 T helper lymphocytes by interferon-gamma and interleukin-12. *Immunity*, 30(5):673–683.

Schwartz, D. and Gygi, S. P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol*, 23(11):1391–1398.

Searle, B. C., Dasari, S., Wilmarth, P. A., Turner, M., Reddy, A. P., David, L. L., and Nagalla, S. R. (2005). Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J Proteome Res*, 4(2):546–554.

Shao, J., Xu, D., Tsai, S., Wang, Y., and Ngai, S. (2009). Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PLoS One*, 4(3):e4920.

Shaw, G. and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, 46(5):659–667.

She, H. Y., Rockow, S., Tang, J., Nishimura, R., Skolnik, E. Y., Chen, M., Margolis, B., and Li, W. (1997). Wiskott-Aldrich syndrome protein is associated with the adapter protein Grb2 and the epidermal growth factor receptor in living cells. *Mol Biol Cell*, 8(9):1709–1721.

Shen, E. C., Henry, M. F., Weiss, V. H., Valentini, S. R., Silver, P. A., and Lee, M. S. (1998). Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev*, 12(5):679–691.

Shi, S., Qiu, J., Sun, X., Suo, S., Huang, S., and Liang, R. (2012). PMes: Prediction of methylation sites based on enhanced feature encoding scheme. *PloS one*, 7(6):e38772.

Shien, D., Lee, T., Chang, W., Hsu, J., Horng, J., Hsu, P., Wang, T., and Huang, H. (2009). Incorporating structural characteristics for identification of protein methylation sites. *Journal of Computational Chemistry*, 30(9):1532–1543.

Shirai, A., Sadaie, M., Shinmyozu, K., and Nakayama, J. (2010). Methylation of ribosomal protein L42 regulates ribosomal function and stress-adapted cell growth. *J Biol Chem*, 285(29):22448–22460.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*, 7:539–539.

Silvin, C., Belisle, B., and Abo, A. (2001). A role for Wiskott-Aldrich syndrome protein in T-cell receptor-mediated transcriptional activation independent of actin polymerization. *J Biol Chem*, 276(24):21450–21457.

Sims, R. J., Rojas, L. A., Beck, D., Bonasio, R., Schüller, R., Drury, W. J., Eick, D., and Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science*, 332(6025):99–103.

Smith, W. A., Schurter, B. T., Wong-Staal, F., and David, M. (2004). Arginine methylation of RNA helicase a determines its subcellular localization. *J Biol Chem*, 279(22):22795–22798.

Snijders, A. P., Hung, M. L., Wilson, S. A., and Dickman, M. J. (2010). Analysis of arginine and lysine methylation utilizing peptide separations at neutral pH and electron transfer dissociation mass spectrometry. *J Am Soc Mass Spectrom*, 21(1):88–96.

Starita, L. M., Lo, R. S., Eng, J. K., von Haller, P. D., and Fields, S. (2012). Sites of ubiquitin attachment in Saccharomyces cerevisiae. *Proteomics*, 12(2):236–240.

Steen, H. and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol*, 5(9):699–711.

Sterck, L., Rombauts, S., Vandepoele, K., Rouzé, P., and Van de Peer, Y. (2007). How many genes are there in plants (. . . and why are they there)? *Current Opinion in Plant Biology*, 10(2):199–203.

Stoecklin, G., Hahn, S., and Moroni, C. (1994). Functional hierarchy of AUUUA motifs in mediating rapid interleukin-3 mRNA decay. *J Biol Chem*, 269(46):28591–28597.

Stracker, T. H. and Petrini, J. H. (2011). The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol*, 12(2):90–103.

Su, I. H., Dobenecker, M. W., Dickinson, E., Oser, M., Basavaraj, A., Marqueron, R., Viale, A., Reinberg, D., Wülfing, C., and Tarakhovsky, A. (2005). Polycomb group protein ezh2 controls actin polymerization and cell signaling. *Cell*, 121(3):425–436.

Sun, L., Wang, M., Lv, Z., Yang, N., Liu, Y., Bao, S., Gong, W., and Xu, R. M. (2011). Structural insights into protein arginine symmetric dimethylation by PRMT5. *Proc Natl Acad Sci U S A*, 108(51):20538–20543.

Swiercz, R., Cheng, D., Kim, D., and Bedford, M. T. (2007). Ribosomal protein rpS2 is hypomethylated in PRMT3-deficient mice. *J Biol Chem*, 282(23):16917–16923.

Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*, 101(26):9528–9533.

Takahashi, Y., Daitoku, H., Hirota, K., Tamiya, H., Yokoyama, A., Kako, K., Nagashima, Y., Nakamura, A., Shimada, T., Watanabe, S., Yamagata, K., Yasuda, K., Ishii, N., and Fukamizu, A. (2011). Asymmetric arginine dimethylation determines life span in C. elegans by regulating forkhead transcription factor DAF-16. *Cell Metab*, 13(5):505–516.

Tang, J., Frankel, A., Cook, R. J., Kim, S., Paik, W. K., Williams, K. R., Clarke, S., and Herschman, H. R. (2000). PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *J Biol Chem*, 275(11):7723–7730.

Taouatas, N., Drugan, M. M., Heck, A. J., and Mohammed, S. (2008). Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods*, 5(5):405–407.

Taylor, G. A., Carballo, E., Lee, D. M., Lai, W. S., Thompson, M. J., Patel, D. D., Schenkman, D. I., Gilkeson, G. S., Broxmeyer, H. E., Haynes, B. F., and Blackshear, P. J. (1996). A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity*, 4(5):445–454.

Taylor, M. D., Sadhukhan, S., Kottangada, P., Ramgopal, A., Sarkar, K., D'Silva, S., Selvakumar, A., Candotti, F., and Vyas, Y. M. (2010). Nuclear role of WASp in the pathogenesis of dysregulated TH1 immunity in human Wiskott-Aldrich syndrome. *Sci Transl Med*, 2(37):37–37.

Tee, W. W., Pardo, M., Theunissen, T. W., Yu, L., Choudhary, J. S., Hajkova, P., and Surani, M. A. (2010). Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev*, 24(24):2772–2777.

Thrasher, A. J. (2009). New insights into the biology of Wiskott-Aldrich syndrome (WAS). *Hematology Am Soc Hematol Educ Program*, pages 132–138.

Trudgian, D. C., Singleton, R., Cockman, M. E., Ratcliffe, P. J., and Kessler, B. M. (2012). Modls:post-translational modification localization scoring with automatic specificity expansion.

Trudgian, D. C., Thomas, B., McGowan, S. J., Kessler, B. M., Salek, M., and Acuto, O. (2010). CPFP: a central proteomics facilities pipeline. *Bioinformatics*, 26(8):1131–1132.

Tsugita, A. (1987). Developments in protein microsequencing. *Advances in biophysics*, 23:81–113.

Uhlmann, T., Geoghegan, V. L., Thomas, B., Ridlova, G., Trudgian, D. C., and Acuto, O. (2012). A method for large-scale identification of protein arginine methylation. *Mol Cell Proteomics*, 11:1489–1499.

Vagin, V. V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G. J., and Aravin, A. A. (2009). Proteomic analysis of murine piwi proteins reveals a role for arginine methylation in specifying interaction with tudor family members. *Genes Dev*, 23(15):1749–1762.

van der Merwe, P. A. and Dushek, O. (2011). Mechanisms for T cell receptor triggering. *Nat Rev Immunol*, 11(1):47–55.

Villarino, A. V., Katzman, S. D., Gallo, E., Miller, O., Jiang, S., McManus, M. T., and Abbas, A. K. (2011). Posttranscriptional silencing of effector cytokine mRNA underlies the anergic phenotype of self-reactive T cells. *Immunity*, 34(1):50–60.

Voet, D. and Voet, J. (2004). *Biochemistry*. John Wiley and Sons, third edition.

Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., Cox, J., Mann, M., and Choudhary, C. (2011). A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics*, 10(10):M111.013284.

Waldmann, T., Izzo, A., Kamieniarz, K., Richter, F., Vogler, C., Sarg, B., Lindner, H., Young, N. L., Mittler, G., Garcia, B. A., and Schneider, R. (2011). Methylation of H2AR29 is a novel repressive PRMT6 target. *Epigenetics Chromatin*, 4:11–11.

Wang, J., Chen, L., Sinha, S. H., Liang, Z., Chai, H., Muniyan, S., Chou, Y. W., Yang, C., Yan, L., Feng, Y., Kathy Li, K., Lin, M. F., Jiang, H., George Zheng, Y., and Luo, C. (2012). Pharmacophore-based virtual screening and biological evaluation of small molecule inhibitors for protein arginine methylation. *J Med Chem*, 55(18):7978–7987.

Wang, L., Pal, S., and Sif, S. (2008). Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. *Mol Cell Biol*, 28(20):6262–6277.

Wang, R., Zheng, W., Yu, H., Deng, H., and Luo, M. (2011). Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J Am Chem Soc*, 133(20):7648–7651.

Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates, J. R., Delahunty, C. M., Hahn, P., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J., and Böttger, A. (2009). Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science*, 325(5936):90–93.

Willnow, S., Martin, M., Lüscher, B., and Weinhold, E. (2012). A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. *Chembiochem*, 13(8):1167–1173.

Wolos, J. A., Frondorf, K. A., Davis, G. F., Jarvi, E. T., McCarthy, J. R., and Bowlin, T. L. (1993). Selective inhibition of T cell activation by an inhibitor of S-adenosyl-L-homocysteine hydrolase. *J Immunol*, 150(8 Pt 1):3264–3273.

Wu, J., Cui, N., Wang, R., Li, J., and Wong, J. (2012). A role for CARM1-mediated histone H3 arginine methylation in protecting histone acetylation by releasing corepressors from chromatin. *PLoS One*, 7(6):34692.

Wu, R., Haas, W., Dephoure, N., Huttlin, E. L., Zhai, B., Sowa, M. E., and Gygi, S. P. (2011). A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods*, 8(8):677–683.

Wysocki, V. H., Tsaprailis, G., Smith, L. L., and Breci, L. A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom*, 35(12):1399–1406.

Xu, W., Cho, H., Kadam, S., Banayo, E. M., Anderson, S., Yates, J. R., Emerson, B. M., and Evans, R. M. (2004). A methylation-mediator complex in hormone signaling. *Genes Dev*, 18(2):144–156.

Xu, Y., Ding, J., Huang, Q., and Deng, N. (2012). Prediction of protein methylation sites using conditional random field. *Protein and peptide letters*, pages 403–412.

Yadav, N., Lee, J., Kim, J., Shen, J., Hu, M. C., Aldaz, C. M., and Bedford, M. T. (2003). Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. *Proc Natl Acad Sci U S A*, 100(11):6464–6468.

Yang, Y., Lu, Y., Espejo, A., Wu, J., Xu, W., Liang, S., and Bedford, M. T. (2010). TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell*, 40(6):1016–1023.

Yoshimatsu, M., Toyokawa, G., Hayami, S., Unoki, M., Tsunoda, T., Field, H. I., Kelly, J. D., Neal, D. E., Maehara, Y., Ponder, B. A., Nakamura, Y., and Hamamoto, R. (2011). Dysregulation of PRMT1 and PRMT6, type I arginine methyltransferases, is involved in various types of human cancers. *Int J Cancer*, 128(3):562–573.

Yoshimoto, T., Boehm, M., Olive, M., Crook, M. F., San, H., Langenickel, T., and Nabel, E. G. (2006). The arginine methyltransferase PRMT2 binds RB and regulates E2F function. *Exp Cell Res*, 312(11):2040–2053.

Yu, M. C., Bachand, F., McBride, A. E., Komili, S., Casolari, J. M., and Silver, P. A. (2004). Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes Dev*, 18(16):2024–2035.

Yu, Z., Chen, T., Hébert, J., Li, E., and Richard, S. (2009). A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Mol Cell Biol*, 29(11):2982–2996.

Zee, B. M., Levin, R. S., Xu, B., LeRoy, G., Wingreen, N. S., and Garcia, B. A. (2010). In vivo residue-specific histone methylation dynamics. *J Biol Chem*, 285(5):3341–3350.

Zeng, R., Cannon, J. L., Abraham, R. T., Way, M., Billadeau, D. D., Bubeck-Wardenberg, J., and Burkhardt, J. K. (2003). SLP-76 coordinates Nck-dependent Wiskott-Aldrich syndrome protein recruitment with Vav-1/Cdc42-dependent Wiskott-Aldrich syndrome protein activation at the T cell-APC contact site. *J Immunol*, 171(3):1360–1368.

Zhang, G., Deinhardt, K., Chao, M. V., and Neubert, T. A. (2011). Study of neurotrophin-3 signaling in primary cultured neurons using multiplex stable isotope labeling with amino acids in cell culture. *J Proteome Res*, 10(5):2546–2554.

Zhang, J., Shehabeldin, A., da Cruz, L. A., Butler, J., Somani, A. K., McGavin, M., Kozieradzki, I., dos Santos, A. O., Nagy, A., Grinstein, S., Penninger, J. M., and Siminovitch, K. A. (1999). Antigen receptor-induced activation and cytoskeletal rearrangement are impaired in Wiskott-Aldrich syndrome protein-deficient lymphocytes. *J Exp Med*, 190(9):1329–1342.

Zhang, X., Wen, H., and Shi, X. (2012). Lysine methylation: beyond histones. *Acta Biochim Biophys Sin (Shanghai)*, 44(1):14–27.

Zhang, X., Zhou, L., and Cheng, X. (2000). Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J*, 19(14):3509–3519.

Zheng, W. and Flavell, R. A. (1997). The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell*, 89(4):587–596.

Zhong, H. and Li, L. (2005). An algorithm for interpretation of low-energy collision-induced dissociation product ion spectra for de novo sequencing of peptides. *Rapid Commun Mass Spectrom*, 19(8):1084–1096.

Zhou, M., Morgner, N., Barrera, N. P., Politis, A., Isaacson, S. C., Matak-Vinkovi, D., Murata, T., Bernal, R. A., Stock, D., and Robinson, C. V. (2011). Mass spectrometry of intact v-type atpases reveals bound lipids and the effects of nucleotide binding. *Science*, 334(6054):380–385.

Zhu, J., Yamane, H., Cote-Sierra, J., Guo, L., and Paul, W. E. (2006). GATA-3 promotes Th2 responses through three different mechanisms: induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of Th1 cell-specific factors. *Cell Res*, 16(1):3–10.

Zobel-Thropp, P., Gary, J. D., and Clarke, S. (1998). Methylarginine is a novel posttranslational modification of arginine residues in yeast proteins. *The journal of biological chemistry*, 273(45):29283–29286.

Zurita-Lopez, C. I., Sandberg, T., Kelly, R., and Clarke, S. G. (2012). Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming -NG-monomethylated arginine residues. *J Biol Chem*, 287(11):7859–7870.

⌘ Author's Choice

© 2012 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

A Method for Large-scale Identification of Protein Arginine Methylation*[S]

Thomas Uhlmann‡¶‖, Vincent L. Geoghegan‡¶, Benjamin Thomas§, Gabriela Ridlova‡§, David C. Trudgian§, and Oreste Acuto‡‡‡**

The lack of methods for proteome-scale detection of arginine methylation restricts our knowledge of its relevance in physiological and pathological processes. Here we show that most tryptic peptides containing methylated arginine(s) are highly basic and hydrophilic. Consequently, they could be considerably enriched from total cell extracts by simple protocols using either one of strong cation exchange chromatography, isoelectric focusing, or hydrophilic interaction liquid chromatography, the latter being by far the most effective of all. These methods, coupled with heavy methyl-stable isotope labeling by amino acids in cell culture and mass spectrometry, enabled in T cells the identification of 249 arginine methylation sites in 131 proteins, including 190 new sites and 93 proteins not previously known to be arginine methylated. By extending considerably the number of known arginine methylation sites, our data reveal a novel proline-rich consensus motif and identify for the first time arginine methylation in proteins involved in cytoskeleton rearrangement at the immunological synapse and in endosomal trafficking. ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M112.020743, 1489–1499, 2012.***

Knowledge of the type, extent and dynamics of post-translational modifications (PTMs)[1] reveals the changing network of protein interactions and the regulation of cellular functions. Methylated arginines (Meth-R) are relatively frequent in cellular proteins (*e.g.* 0.7–1% of total arginines (1) and this work) and have a very slow turnover (2), likely to confer lasting functional properties to proteins. Meth-Rs are often found at glycine-arginine-rich (GAR) sequences and modulate protein-protein and protein-nucleic acid interactions by reducing hydrogen-bonding and local hydrophilicity (2). Methylation at arginine can weaken interactions but also enhance the binding of adaptors called Tudor domains to Meth-R-containing sequences (3, 4).

Meth-R occurs in three forms: asymmetric dimethylarginine (ADMA), bearing two methyl groups on one nitrogen of the guanidino group; symmetric dimethylarginine (SDMA), where both nitrogens are singly methylated and monomethylarginine (MMA), a reaction intermediate. In humans, these reactions are catalyzed by a family of protein arginine methyltransferases (PRMTs), with PRMT1, 3, 4, 6, and 8 forming ADMA and PRMT5 and 7 producing SDMA. In mice, PRMT1 and PRMT4 deficiency causes embryonic (5) or Perinatal (6) lethality, respectively, whereas lack of PRMT3 leads to growth defect (7). Arginine methylation has been investigated for its role in the regulation of gene expression at multiple levels and in DNA repair (2). PRMT1, 4, 5, 6, and 7 control the epigenetic code by methylating histone N-terminal tails (H1, H2A, 3, and 4) (4, 8–12) and tune the activity of transcription factors (2, 13–15), coactivators (16, 17) and corepressors (8, 11). Factors involved in mRNA splicing (18, 19), elongation (20), transport and translation (19, 21) often contain Meth-R, suspected to be involved in protein-RNA interaction (2). Moreover, T cell CD28 co-stimulation (22), TNFα (15), TLR4 (15), NGF (23) and NFkB (15) signaling pathways may be regulated by PRMTs. Protein arginine methylation is implicated in pathogenic processes, including oncogenesis (24), cardiovascular disease (25), autoimmunity and viral infections (26), raising the possibility that abnormally methylated proteins can be disease markers and PRMTs potential therapeutic targets (27).

Only a very limited number of arginine methylated proteins and sites have been identified with certainty to date (see UniProtKB). *In vitro* elegant screenings were developed to search for PRMT substrates (21) but identifying the Meth-R concerned is usually cumbersome. The existence of Meth-R sites is in many cases only suspected by protein similarity such as the presence of GAR motifs, which are substrates for some but not all PRMTs (2). These vexing limitations hamper progress of a very promising field in biology. Previous work developed a powerful approach using heavy methyl-SILAC coupled to MS to identify with high certainty peptides containing Meth-R (28). An anti-ADMA antibody was used to

From the ‡T Cell Signalling Laboratory and §Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom

Received May 30, 2012, and in revised form, July 30, 2012
⌘ *Author's Choice*—Final version full access.
Published, MCP Papers in Press, August 3, 2012, DOI 10.1074/mcp.M112.020743

[1] The abbreviations used are: PTMs, post-translational modifications; Meth-R, methylated arginines; ADMA, asymmetric dimethylarginine, GAR, glycine-arginine-rich; SDMA, symmetric dimethylarginine; MMA, monomethylarginine; PRMTs, protein arginine methyltransferases; HILIC, hydrophilic interaction liquid chromatography; SCX, strong cation exchange; SILAC, stable isotope labeling by amino acids in cell culture; PRAM, proline-rich arginine methylation; FDR, false discovery rate.

enrich for Meth-R-containing proteins but only a limited number of Meth-R were identified (28). The relative low abundance of Meth-R-containing proteins and peptides and lack of adequate methods for their enrichment remains therefore a major hurdle for both large- and small-scale discovery of Meth-R sites, leaving the heavy methyl-SILAC method essentially unexploited. Here, we have uncovered distinct physico-chemical properties of Meth-R-containing tryptic peptides and showed that hydrophilic interaction liquid chromatography (HILIC) is a simple method providing excellent enrichment of Meth-R-containing peptides. When coupled to heavy methyl-stable isotope labeling by amino acids in cell culture (SILAC) based MS, HILIC allowed us to identify hundreds of Meth-R sites, providing for the first time a means for "methylome"-scale investigation that vastly outperforms the use of anti-ADMA antibody.

EXPERIMENTAL PROCEDURES

SILAC Labeling and Cell Lysis for MS Analysis—Custom-made RPMI 1640 medium lacking L-Methionine, L-Arginine, and L-Lysine (Invitrogen, Carlsbad, CA) was supplemented with 10% dialyzed fetal calf serum (Invitrogen), L-Arginine and L-Lysine and either a) L-Methionine (CK Gas Products Ltd.), or b) L-Methionine-methyl-^{13}C-methyl-D_3 (Sigma, ISOTEC™) at a final concentration of 0.1 mM L-Methionine, 0.29 mM L-Arginine and 0.219 mM L-Lysine and filter sterilized (0.22 μm pore size, Millipore, Billerica, MA). Jurkat cells were grown at 37 °C in a humidified 5% CO_2-containing atmosphere for 5–7 cell doublings in labeling media. Human primary CD4+ lymphocytes were isolated from whole blood of healthy donors by negative selection using a Dynal isolation kit (Invitrogen) according to the manufacturer's protocol. 1×10^6 to 2×10^6 CD4+ lymphocytes in 10 ml of light or heavy medium were then stimulated by plate bound anti-CD3 at 10 μg/ml and soluble CD28.2 (BioLegend, San Diego, CA) at 1 μg/ml. IL-2 (AbDSerotec) was added to a concentration of 50 ng/ml. Cells were grown at 37 °C in a humidified 5% CO_2-containing atmosphere for 7–9 days. Cells were harvested and lysed for either a) 10 min in ice-cold lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1% Nonidet P-40, 0.5% Deoxicholate, protease inhibitor mixture (Roche), or b) 20 min in 8 M Urea in 25 mM Tris pH 8. Lysates were cleared by centrifugation at 14,000 $\times$ *g* for 20 min and mixed in a 1:1 ratio of protein content (measured by Bradford method and absorbance at 280 nm). Lysates were used for a) immunoprecipitation, or b) strong Cation exchange (SCX) chromatography, HILIC and OFFGEL isoelectric focusing.

Immunoprecipitation and In-gel Digestion—Immunoprecipitation was performed on the mixed cell lysates overnight at 4 °C on Sepharose protein G beads (GE Healthcare) pre-incubated with an anti-DMA mAb 7E6 (Abcam, Cambridge, UK). Beads were washed four times with fresh ice-cold lysis buffer and eluted with 100 mM triethylamine pH 11.5 for 10 min at RT. The eluates were neutralized with $^1/_{20}$ volume of 1 M phosphate buffer pH 6.8, boiled in reducing SDS NuPAGE sample buffer (Invitrogen), alkylated with 55 mM iodoacetamide (Sigma) and separated on 4–12% gradient Bis-Tris NuPAGE gels (Invitrogen). The gel was washed with distilled water and lightly stained with Colloidal Blue (Invitrogen). The gel lane was divided into 15 slices and subject to GeLCMS/MS.

SCX, HILIC, and OFFGEL Fractionation—Cleared Urea lysate was acetone precipitated and resuspended in 1.6 M Urea in 25 mM Tris pH8. Overnight digestion at 37 °C was carried out with 12.5 ng/μl trypsin (Proteomics grade, Sigma) or chymotrypsin (Sequencing grade, Promega, Madison, WI). For SCX, 4–6 mg peptides were acidified with 0.1% trifluoroacetic acid (Sigma) and cleared by centrifuging at 20,000 $\times$ *g* for 10 min. Peptides were applied to a 1 ml SCX column (Resource S, GE Healthcare) using Buffer A (50 mM KH_2PO_4 20% acetonitrile, pH2.7). Bound peptides were eluted with increasing Buffer B (50 mM KH_2PO_4, 20% acetonitrile, 1 M KCl, pH2.7). In a first region of separation the gradient was stepped up to 7% Buffer B to elute 1+ and most of the 2+ charged peptides. A shallow gradient to 35% Buffer B separated the remaining 2+ from 3+ and higher charged peptides. In the last stage the Buffer B was changed to pH7 and a steep gradient to 100% Buffer B ensured complete elution of bound peptides. Following C18 desalting (Empore Octadecyl C18, Sigma), the fractions were analyzed by LC-MS/MS.

For HILIC 1 mg of peptides were acidified with 0.1% trifluoroacetic acid (TFA) and cleared by centrifuging at 20,000g for 10 min. Peptides were loaded onto a reverse phase C2/C18 column (GE Healthcare) with Buffer A (0.1% formic acid, pH2.7). Elution of peptides was carried out with a steep gradient to 100% Buffer B (0.1% formic acid, 95% acetonitrile, pH2.7). Desalted peptides were then dried under vacuum, re-suspended in Buffer B and applied to a HILIC column (Merck). Bound peptides were eluted with a shallow gradient of Buffer A up to 80%.

For IEF on OFFGEL, 1 mg of peptides were loaded onto an IPG strip (13 cm pH7–11, GE Healthcare) with IPG buffer (pH7–11, GE Healthcare) according to manufacturer's instructions. Peptides were focused in solution with an Agilent 3100 OFFGEL fractionator (Agilent Technologies, Santa Clara, CA) for 20 kVh up to a maximum voltage of 8000 V. Focused peptides were removed and desalted with C18 (Empore Octadecyl C18, Sigma). Peptides were also extracted by immersing the paper wicks from the cathode end in 0.1% TFA for 1–2 h, followed by brief sonication. Peptides from the respective fractions were collected and adjusted to 0.1% TFA, 5% acetonitrile and applied to a homemade C18 stage tip for desalting. Bound peptides were washed with 0.1% TFA, 2% acetonitrile and eluted in 0.1% TFA, 60% acetonitrile. Eluted peptides were dried and sent to the PNAC facility in Cambridge (http://www.bioc.cam.ac.uk/pnac/aaa.hml) for amino acid analysis on a Biochrome 30 Analyzer.

Mass Spectrometry Data Acquisition—Samples were analyzed on an Ultimate 3000 nano HPLC (Dionex, Camberley, UK) system run in direct injection mode coupled to either a LTQ XL Orbitrap mass spectrometer, or a Q Exactive mass spectrometer (Thermo Electron, Hemel Hempstead, UK). Samples were resolved on a 15 cm by 75 μm inner diameter picotip analytical column (New Objective, Woburn, MA), which was packed in-house with Reprosil-Pur C18-AQ phase, 3 μm bead (Dr. Maisch, Germany). A 120 min gradient was used to separate the peptides and each sample was typically injected three times and data merged in order to increase sample coverage. The LTQ XL Orbitrap mass spectrometer was operated in a "Top 5" data dependent acquisition mode and the Q Exactive mass spectrometer was operated in a "Top 10" data dependent acquisition mode. Precursor scans were performed at a resolving power of 60,000, from which either five precursor ions (LTQ XL Orbitrap) or 10 precursor ions (Q Exactive) were selected and fragmented. Charge state +1 ions were rejected.

Mass Spectrometry Data Analysis—MS/MS peak lists were converted to mzXML format using ReAdW version 4.4.1 (LTQ XL Orbitrap data) or to mgf format using ProteoWizard msconvert release 3.0.3535 (Q Exactive data). Both tools were used with default parameters except that zlib compression of spectral data was enabled. Data was uploaded to the central proteomics facilities pipeline (CPFP at: https://cpfp-master.molbiol.ox.ac.uk/cpfp_demo/auth/login) (29). Files were searched using Mascot version 2.3.01 (Matrix science), X!TANDEM version 2008.12.01.1 (30) and OMSSA version 2.1.8 (31) against a concatenated target and reversed decoy version of the IPI human sequence database (EBI), versions 3.78 and 3.79 containing

86,702 and 86,635 target protein sequences respectively. Enzyme was set to trypsin (or chymotrypsin for some experiments) allowing for up to 3 missed cleavages. Carbamidomethyl cysteine was set as a fixed modification. For analysis of arginine methylation oxidized methionine, heavy oxidized methionine, monomethylarginine, heavy monomethylarginine, dimethylarginine, and heavy dimethylarginine were set as variable modifications. For analysis of lysine methylation, oxidized methionine, heavy oxidized methionine, monomethyllysine, heavy monomethyllysine, dimethyllysine, heavy dimethyllysine, trimethyllysine, and heavy trimethyllysine were set as variable modifications. Mass tolerances for MS and MS/MS peak identifications were 10 ppm and 0.5 Da respectively (LTQ XL Orbitrap) or 20 ppm and 0.1 Da respectively (Q Exactive). Mass spectrometry data are available at http://www.peptideatlas.org/PASS/PASS00081. Assignment of a methylation site required identification by searches in the Central proteomics facility pipeline (CPFP) at 1% FDR and the presence of a confirming peptide in the precursor spectrum of equal intensity separated by a mass difference introduced by the light/heavy methyl group(s). The actual false positive rate of identification using the heavy methyl SILAC criteria (1.2%, supplemental Fig. S5) was calculated on a representative dataset by analyzing all false positive hits for arginine methylation for the presence of confirming heavy methyl SILAC pairs. This number was then doubled and divided by the total number of hits for arginine methylation. The requirement for the presence of a methylSILAC pair to corroborate each methylated peptide increases confidence and enables even low scoring peptides to be considered (28). Modification localization scoring was performed using the ModLS algorithm (unpublished[2]) which extends the PTMScore method to incorporate automatic specificity expansion. For each variable modification chosen for the database search all amino acid specificities defined in the Unimod database (www.unimod.org) are considered during localization. This allows the correct localization of *e.g.* dimethylation to lysine, even if only arginine methylation was chosen for the search. Ambiguous modification localizations are annotated as such in supplemental Table S5. Within CPFP, inference of protein identifications from peptide sequences was performed using ProteinProphet (32). Default parameters were used for protein grouping and the assignment of isoform identifications. Precursor *m/z*, charge, InterProphet probability for all methylated peptides found in this work are listed in supplemental Table S4.

RESULTS

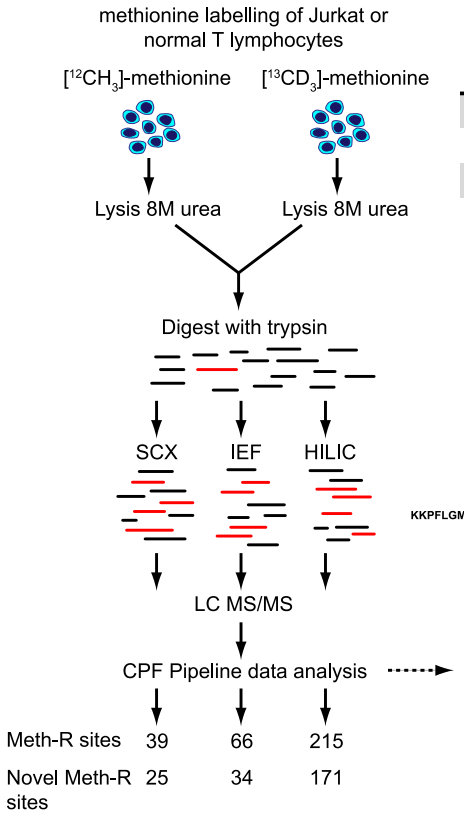
Rationale and Methods for Enrichment of Meth-R Containing Peptides—Biosynthetic labeling with methionine (M0) and [$^{13}CD_3$]-methionine (M4) (heavy methyl-SILAC) (Fig. 1*A*) enables detection of protein methylation sites by LC-MS/MS with high confidence (28). In this study, we labeled Jurkat T cells by heavy methyl-SILAC and extended this approach to primary cells by showing efficient SILAC labeling of antigen receptor stimulated T-cells (supplemental Figs. S1, S2). In a previous study, enrichment of Meth-R-containing proteins from Hela cells with the monoclonal antibody (mAb) 7E6 to MMA and ADMA led to the identification of 59 Meth-R sites (28). By a similar approach (see Methods), we identified 41 Meth-R sites in Jurkat T cells, only 6 of which were novel (*e.g.* not annotated in the UniProtKB database) (supplemental Table S1). However, cell proteomes should comprise higher numbers of Meth-R-containing proteins, since 0.7–1.0% of arginine in proteins extracted from cell lines or tissues is methylated (1) (and our unpublished data). Moreover, similar to previous investigations (28), we identified Meth-R sites almost invariably at RG sites (supplemental Table S1), pointing again at the limited usefulness of 7E6 mAb for proteomic-scale studies. These limitations and the small number of Meth-R sites (137) known to-date (*e.g.* in UniProtKB), spurred us to develop effective, unbiased approaches that enrich for Meth-R-containing peptides for large-scale identification of Meth-R-containing sites using LC-MS/MS analysis (Fig. 1*A*).

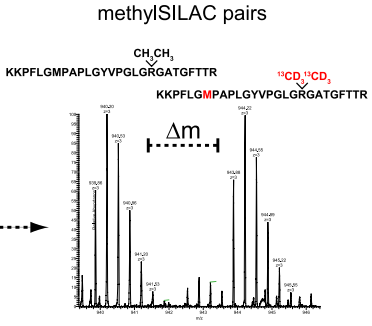
SCX Chromatography—Approximately 70% of tryptic peptides in proteomes exhibit a +2 net charge at pH 2.7 and can be separated from other peptides according to single charge differences using SCX chromatography (33). Because trypsin cuts poorly at Meth-R, the missed cleavages produce peptides with charge $\geq+3$ at pH 2.7. Thus, most Meth-R peptides should be present in the $\geq+3$ peptide-containing fractions after SCX chromatography. SCX fractions enriched for tryptic peptides of charge $<+3$ from total proteins of Jurkat cell, exhibited ~1.0% of Meth-R (as revealed by amino acid analysis, see Methods and supplemental Fig. S3*A*) but no Meth-R peptide could be identified by LC-MS/MS (data not shown). In contrast, in SCX fractions of tryptic peptides of charge $\geq+3$, ~4.0% of arginine was present as Meth-R (supplemental Fig. S3*A*) allowing us to identify 39 Meth-R sites from Jurkat and primary human T cells in four independent experiments (Fig. 1*A* and supplemental Table S1), of which 25 were new and, interestingly, 4 were not at RG motifs (supplemental Table S1). SCX fractionation enabled us to discover 10 new Meth-R containing proteins and Meth-R within sequences rarely selected by the 7E6 mAb (supplemental Table S1). In addition, as predicted by the poor trypsin cleavage at Meth-K, a known Meth-K peptide was also found (histone H4, supplemental Table S1).

IEF—Histidine-containing tryptic peptides (charge $\geq+3$ at pH 2.7) were the major contaminant (60–70%) within the SCX fractions in which most Meth-R peptides were detected (not shown). This likely contributed to the analyte complexity, limiting the number of Meth-R peptides identified. However, these data brought to our attention key physico-chemical properties of Meth-R tryptic peptides. Arginines prone to be methylated are likely to be fully exposed to solvent and therefore most often embedded within sequences containing hydrophilic and/or neutral amino acid residues. Moreover, we noticed that sequences surrounding Meth-R sites (from UniProtKB and those found by 7E6 Ab and SCX enrichment) infrequently contained negatively charged amino acids. This suggested that tryptic peptides carrying one or more Meth-R should frequently display basic isoelectric points (pIs). This was indeed the case for Meth-R peptides found by 7E6 mAb and SCX enrichment (Fig. 1*B*) with a median pI$>$ 11.00. To test this prediction on a larger scale we fractionated tryptic

[2] David C. Trudgian, Rachelle Singleton, Matthew E. Cockman, Peter J. Ratcliffe, and Benedikt M. Kessler, ModLS: Post-translational modification localization scoring with automatic specificity expansion 2012 (submitted).

A



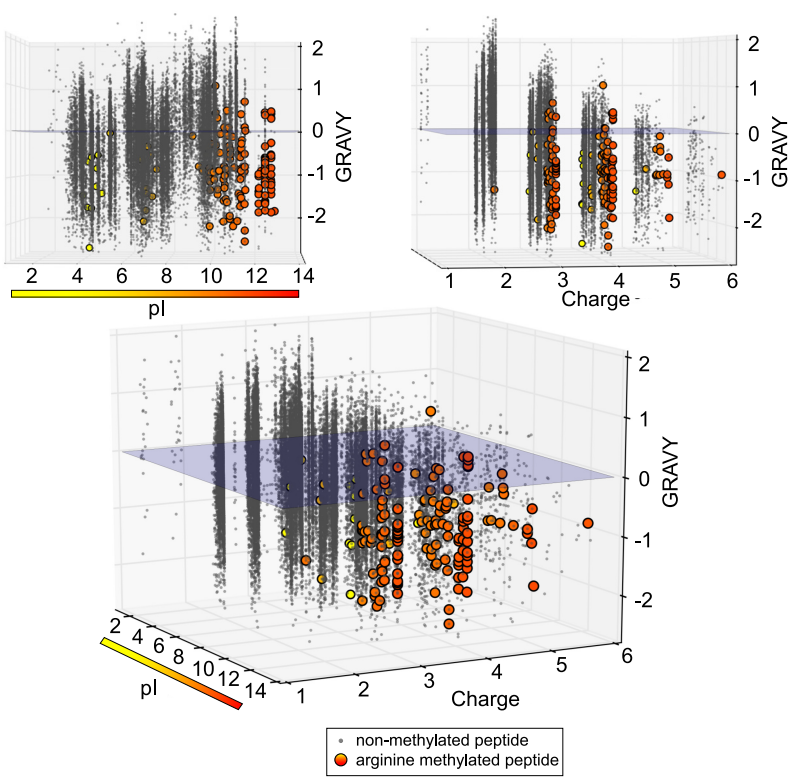


B

	pI			GRAVY			
	Max	Min	Median	Max	Min	Median	n
DMA-IP	12.78	4.50	11.51	0.06	-2.21	-1.05	24
SCX	12.95	3.94	11.25	0.30	-2.47	-1.05	45
IEF	12.78	4.08	11.31	0.66	-2.46	-1.05	50
HILIC	12.79	4.05	11.19	0.65	-2.61	-0.84	140

n = number of arginine methylated peptides

C



digests from heavy methyl-SILAC-labeled Jurkat or primary CD4+ lymphocytes by isoelectric focusing on IPG strips (Fig. 1*A*, see Methods). The results of four independent experiments consistently showed that > 90% of Meth-R peptides displayed a pI ≥ 9 with a median pI of 11 (Figs. 1*B* and 1*C*), indicating that Meth-R tryptic peptides are largely highly basic in nature. Interestingly, at the basic end of the IEF gradient, (pH 9–11 and in the electrode pad itself) the enrichment for Meth-R peptides was substantial, reaching values of ≥ 5% of peptides containing Meth-R as revealed by LC-MS/MS analysis (not shown). However, the lack of IPG strips with a suitable pH range precluded good recovery of most of the highly basic peptides. Using IEF we were able to detect 66 Meth-R sites of which 34 were new (Fig. 1*A* and supplemental Table S1) and found 20 that did not contain an RG motif (supplemental Table S1). IEF fractionation permitted the identification of 19 proteins previously unknown to contain Meth-R and 3 Meth-K sites.

HILIC—HILIC has recently been shown to be an excellent method to complement reverse-phase fractionation (34) by enriching for highly hydrophilic peptides (35). Because the vast majority of Meth-R-containing tryptic peptides identified after 7E6 mAb, SCX or IEF enrichment were highly hydrophilic with a median GRAVY score (36) of –1.05 (Fig. 1*B*), we reasoned that HILIC fractionation should also enable selection for Meth-R containing peptides. Trypsin digests of Methyl-SILAC Jurkat or primary CD4+ lymphocytes were loaded onto a HILIC column in 90% acetonitrile. Bound peptides were eluted with increasing concentration of aqueous buffer (supplemental Fig. S3*B*) and were subjected to amino acid and LC-MS/MS analysis. Fractions containing highly hydrophilic peptides (see supplemental Fig. S3*B*) were enriched ~10 fold for Meth-R, as compared with fractions with fewer hydrophilic peptides, in which no Meth-R-containing peptide could be detected (not shown). Meth-R peptides identified after HILIC separation had a median GRAVY score of −0.84 (Fig. 1*B*), a value similar to that of Meth-R peptides found by anti-DMA, SCX and IEF. In five independent experiments, LC-MS/MS analysis of the twelve most hydrophilic HILIC fractions led to the identification of 215 Meth-R sites (Fig. 1*A* and supplemental Table S1). Notably, HILIC enabled also the detection in chymotrypsin-digested peptides of 37 Meth-R sites not found after tryptic digestion (supplemental Table S1). HILIC alone identified 3 to 5 times higher numbers than 7E6 mAb, SCX and IEF individually and covered ~70% of the 101 unique sites identified by the other three methods together (supplemental Table S1). Fig. 2*A* illustrates the power of HILIC method to discover Meth-R sites as compared with SCX, 7E6 and IEF individually. Importantly, 171 (~ 80%) of Meth-R sites found with HILIC were new (Fig. 1*A* and Table S1), the highest discovery rate among the four different approaches. 87 (~50%) of these new sites were at non-RG motifs. Monomethylated-Rs were more frequently seen with HILIC than with the three other methods. HILIC alone allowed us to assign 215 Meth-R sites to 115 proteins, 94 of which were not previously known to be substrates of PRMTs. In total, 190 (77%) of the 249 unique Meth-R sites identified by 7E6 mAb, SCX, IEF and HILIC together were new (Fig. 2*B* and supplemental Table S1). Unique sites were assigned to 131 different proteins, 93 (72%) of which were previously unknown to be methylated on arginine (Fig. 2*C*, supplemental Table S1 and Fig. 5). Our study brings the total number of Meth-R sites and Meth-R proteins in the human proteome to 327 and 149 (Fig. 2*C*), respectively, substantially widening, in a single study, our insight into the regulation of cellular functions by PRMTs. HILIC was vastly superior to each of the three other enrichment methods alone or combined for detecting known and new Meth-R sites.

Bioinformatic Analysis of Meth-R Sites and Meth-R Proteins—Heavy methyl-SILAC validation (28) greatly increases the confidence in the identification of Meth-R (and Meth-K) containing peptides. Excluding heavy methyl-SILAC validation, reduced the confidence for Meth-R containing peptides to 67% FDR (despite the fact that for the overall dataset of modified and unmodified peptides the FDR was 1%). However, inclusion of putative Meth-R peptides requiring a corresponding light/heavy methyl-SILAC partner raised the confidence of identifying Meth-R sites to ~1% FDR (supplemental Fig. S5). Therefore we feel highly confident about the Meth-R sites identified in our study, which were all manually validated by inspecting individual light/heavy pairs. This is further supported by the fact that our investigation found 43% of the known Meth-R sites.

Of the 190 new Meth-R sites discovered in our study, 57% occurred at RG sites, at times within GAR motifs (Fig. 3*A* and supplemental Table S1), thought to be mainly substrates for

FIG. 1. ***A*, Schematic representation of our workflow. Cells were labeled with light or heavy methionine, lysed and combined in a 1:1 protein ratio.** Proteins were digested with trypsin or, in some experiments, with chymotrypsin. Arginine methylated peptides (depicted as *red lines*) were enriched through the use of SCX, IEF (in an OFFGEL apparatus) or HILIC and analyzed by mass spectrometry. MS spectra for putative arginine methylated peptides were manually verified for the presence of a 1:1 methyl-SILAC pair. An example methyl-SILAC pair is shown. For each separation method, the total number of Meth-R sites identified as well as the novel ones is indicated; *B*, Meth-R-containing tryptic peptides have distinct physico-chemical properties. The Table indicates maximum, minimum and median values for isoelectric points (pI) and GRAVY scores of Meth-R peptides identified with the indicated enrichment methods. The number of unique peptides (n) from each method used to calculate values is shown. *C*, 3D scatter plot of 33000 non-methylated tryptic peptides (*gray circles*) identified from pH3–11 IPG strip and 128 arginine methylated tryptic peptides (*yellow-red circles*) identified from anti-DMA-IP, HILIC, SCX and IEF. Three different views of the same plot are shown. The peptides are plotted according to pI, GRAVY score and charge state at pH 2.7, representing the different selection criteria exploited by the enrichment methods. *Yellow* represents a low pI and *red* a high pI.

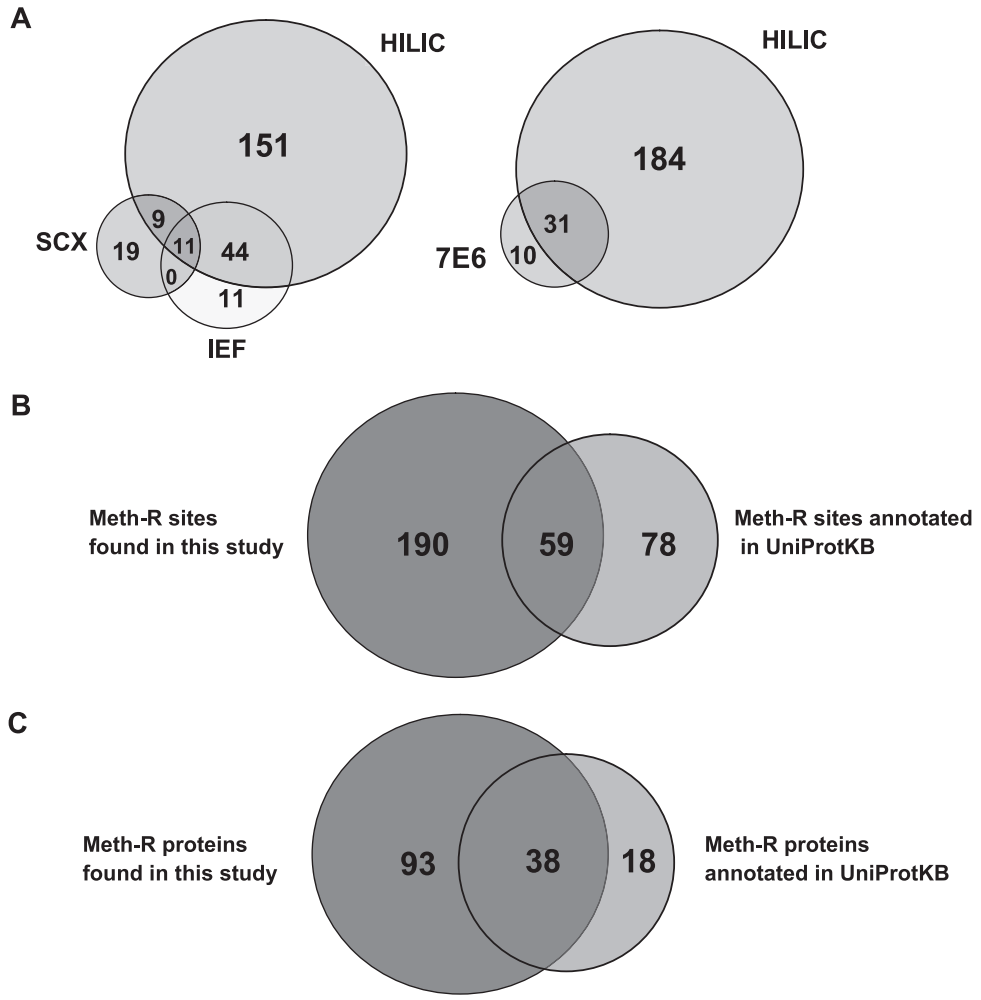


FIG. 2. ***A*, Venn diagrams of the number of Meth-R sites identified by HILIC, IEF, and SCX individually and combined (*left*) or unique in HILIC or 7E6 mAb or found by both methods (*right*) in Jurkat cells and primary human CD4+ lymphocytes combined.** *B*, Venn diagram showing the number of Meth-R sites found only in this study, the Meth-R sites annotated in UniProtKB database and those in common. *C*, Venn diagram of the number of arginine methylated proteins identified in this study, those previously reported in the UniProtKB database and those in common. Circles are approximately scaled relative to number of sites.

PRMT1 (2). The remaining 43% were found at non-RG motifs. This is a considerable increase in the frequency of non-RG containing sequences among Meth-R sites when compared with 25% within the 137 sites known to date (supplemental Fig. S6). In ~ 50% of non-RG sites found here, proline occurred quite frequently before and/or after R (Fig. 3*A*). Such a feature was also observable, though less frequently (32%), in previously identified non-GAR containing methylation sites (not shown). Inspecting the set of non-RG sites for the occurrence of proline at a particular position flanking Meth-R (see Methods), revealed that proline at positions −1 (R being 0) and +4, was significantly favored, the sequences containing proline at either or both of these positions (x(P)Rxxx(P)) accounting for 33% of all the methylation sites we found (Fig. 3*B*). We have provisionally designated this subset of new Meth-R consensus sites as Proline-Rich Arginine Methylation (PRAM) motif. Although glycine and proline often followed the Meth-R, we found a number of sites where the Meth-R was followed by leucine and alanine (Fig 3*C*). Clearly, with growing numbers of Meth-R sites to be identified in future proteomics studies, it should be possible to refine consensus site subsets and help assign PRMT specificities.

The newly discovered Meth-R proteins appear to reside in the cytoplasm and/or nucleus with no preference for either compartment Fig. 4*A*). As previously noticed (2), a high proportion of Meth-R proteins are known to bind nucleic acids and are involved in regulating chromatin structure, transcription, RNA processing and translation. Our study considerably increases the number of such proteins regulated by PRMTs, including five implicated in chromatin structure, 22 participating in the transcription process and 21 in RNA processing (see Figs. 4*B* and 5, blue circles). Most interestingly, we also found proteins implicated in other cellular functions not previously reported to be subjected to regulation by PRMTs. Among

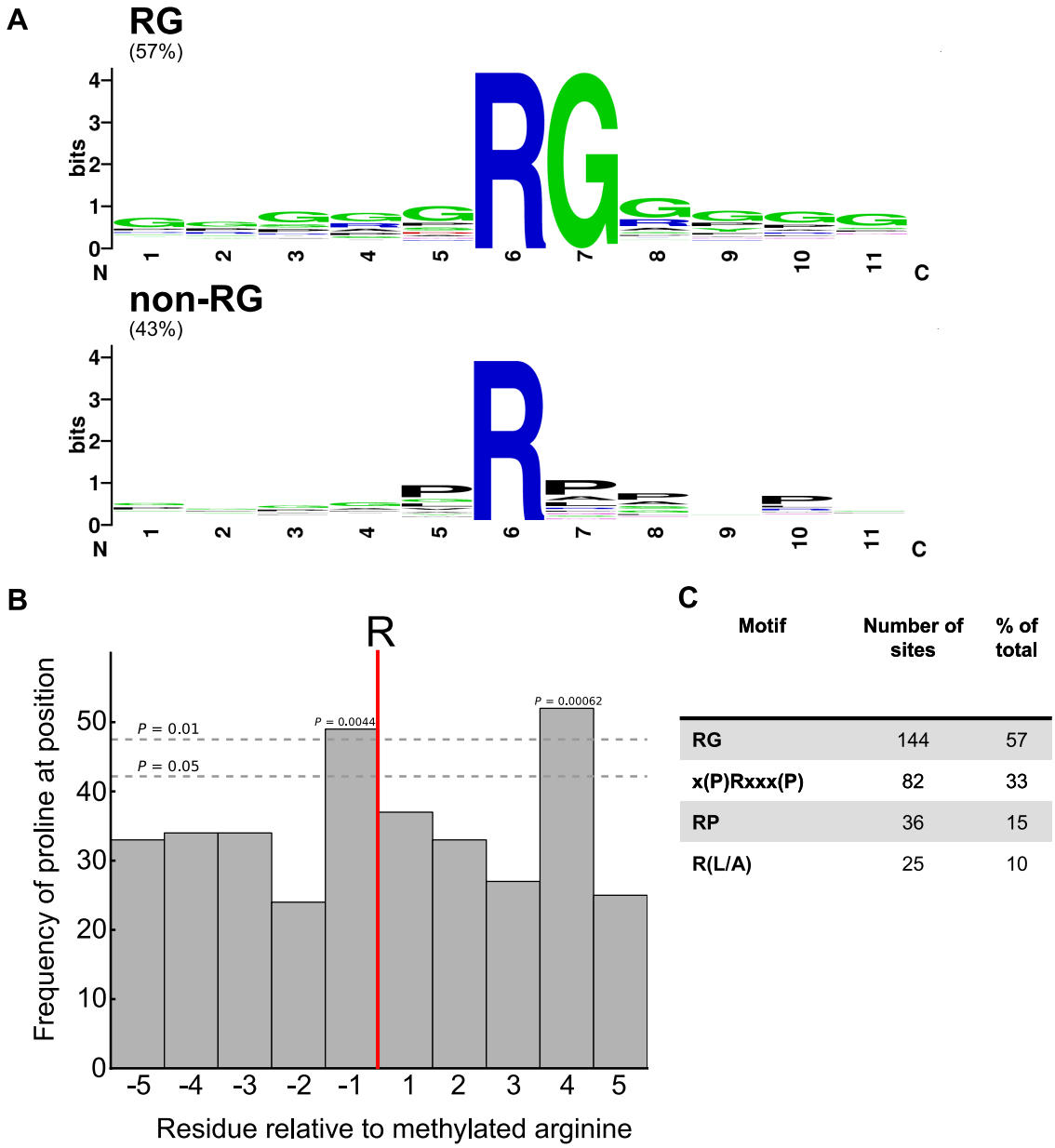


FIG. 3. ***A*, Sequence logo (41) representation of all arginine methylation sites identified in this study.** Of methylated arginines, 57% were followed by a glycine and were generally flanked by a glycine rich sequence. The remaining 43% of sites reveal a previously unreported proline-rich arginine methylation (PRAM) motif. *B*, Frequency of proline at each position relative to the methylated arginine. Dashed lines indicate frequency level of proline with a probability of occurrence by chance of 0.05 or 0.01. *p* values for proline frequency at position –1 and +4 are shown. *p* values were calculated by bootstrapping. *C*, Most common motifs found at Meth-R sites. Note that motifs can be composed of combinations of motifs listed.

these (see Fig. 5) were two nuclear pore components (RANBP2 and TPR) and 6 proteins (DNM2, PICALM, SEC24C, GGA3, SNX3, and CLINT1 (see, UniProtKB) implicated in endosomal trafficking and three, WASP/WIP complex (37) and DNM2 (38), involved in actin cytoskeletal re-arrangement at the immunological synapse (IS). A considerable number of Meth-R sites were found in proteins of unknown function (Fig. 5).

Although SCX and HILIC could in principle enrich also for Meth-K proteins, we found only 10 lysine methylation sites (supplemental Table S1 and Fig. 5). Perhaps Meth-K occurs less frequently than Meth-R in proteins and/or Meth-K- and Meth-R-containing peptides present different physico-chemical properties and thus both would not be enriched by the methods used here. Nevertheless, we found three proteins previously unknown to be Meth-K, including two heat shock proteins, HSPA5 and HSPA8.

Overall, the methods explored here, with HILIC being by far the most effective, allowed us to considerably enrich for

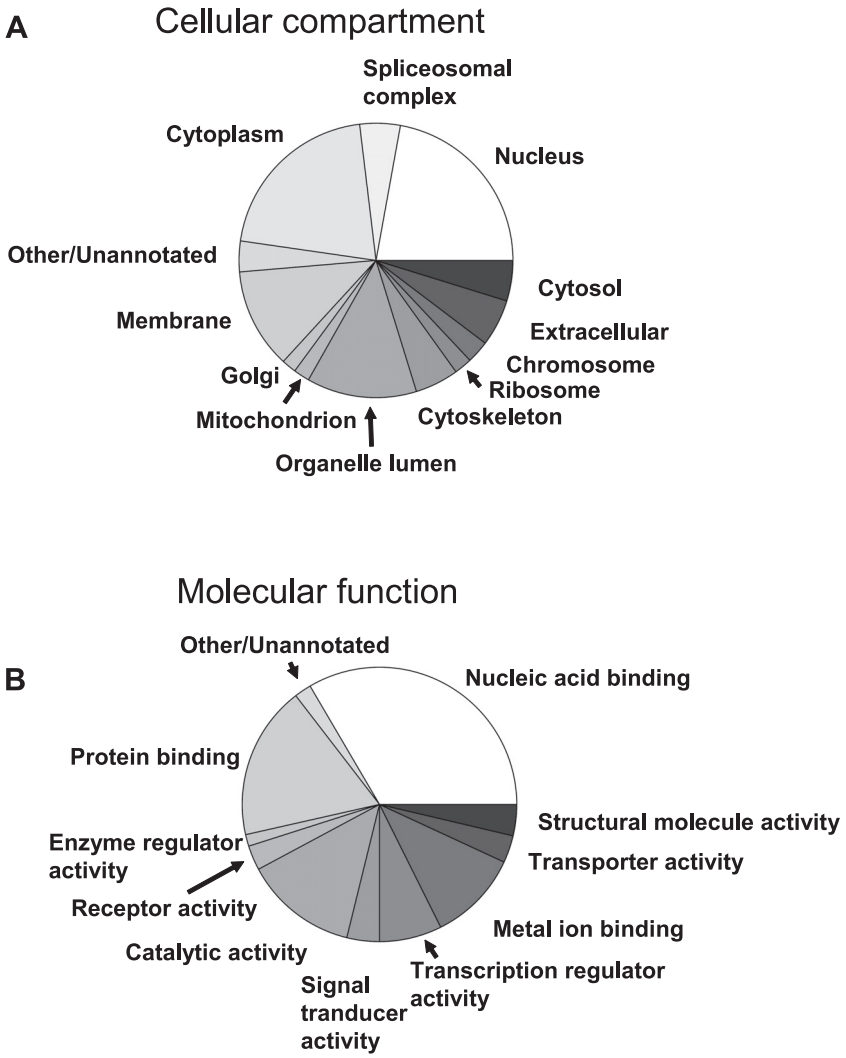


FIG. 4. **Breakdown of arginine methylated proteins according to distribution in cellular compartments (*A*) or by molecular function (*B*) according to Gene Ontology (GO) annotation.**

Meth-R peptides other than those in canonical RG/GAR motifs and uncover a substantial number of new proteins and functions likely to be regulated by PRMTs.

DISCUSSION

To date, no effective and unbiased method for the enrichment of protein and/or peptide containing Meth-R exists. An mAb (7E6) directed at ADMA and MMA permitted the identification of only a handful of Meth-R sites (28) (and this work) almost exclusively at RG sites. Because antibodies are unlikely to effectively discriminate (and with high affinity) the subtle chemical difference between Meth-R and unmodified R, use of pan Meth-R antibodies may not be suitable. A major obstacle in the identification of MS-based Meth-R, namely, being incorrectly mistaken for other isobaric chemical modifications, was recently solved by using a heavy methyl-SILAC approach (28), raising the possibility of accurate large-scale studies of protein arginine methylation. Despite this, a simple and robust method was missing for effective enrichment of Meth-R-containing peptide for small- and proteomic-scale analysis of arginine methylation. The exploratory phase of this investigation revealed that >90% of Meth-R-containing tryptic peptides displayed high basic pI and high hydrophilicity. Enrichment using IEF was barred by the unavailability of IPG strips with adequate high pH range. We found however that HILIC had excellent capacity to enrich for Meth-R peptides. HILIC alone allowed us to identify from trypsin or chymotrypsin digests of heavy methyl-SILAC cell lysates Jurkat cells and primary CD4+ lymphocytes, 215 Meth-R sites, 80% of which were new and covered almost 90% of the total 190 new Meth-R sites discovered here by all four methods. Our study more than doubles the 137 Meth-R sites known to date and adds 93 Meth-R-containing proteins to the 59 previously known in humans, substantially increasing in a single effort our knowledge of the extent and biological significance of arginine methylation. Although these numbers seem limited in comparison to those reported for other PTMs, it is possible that arginine methylation is more restricted and selective, also considering that in this study we explored only the T cell proteome. Although

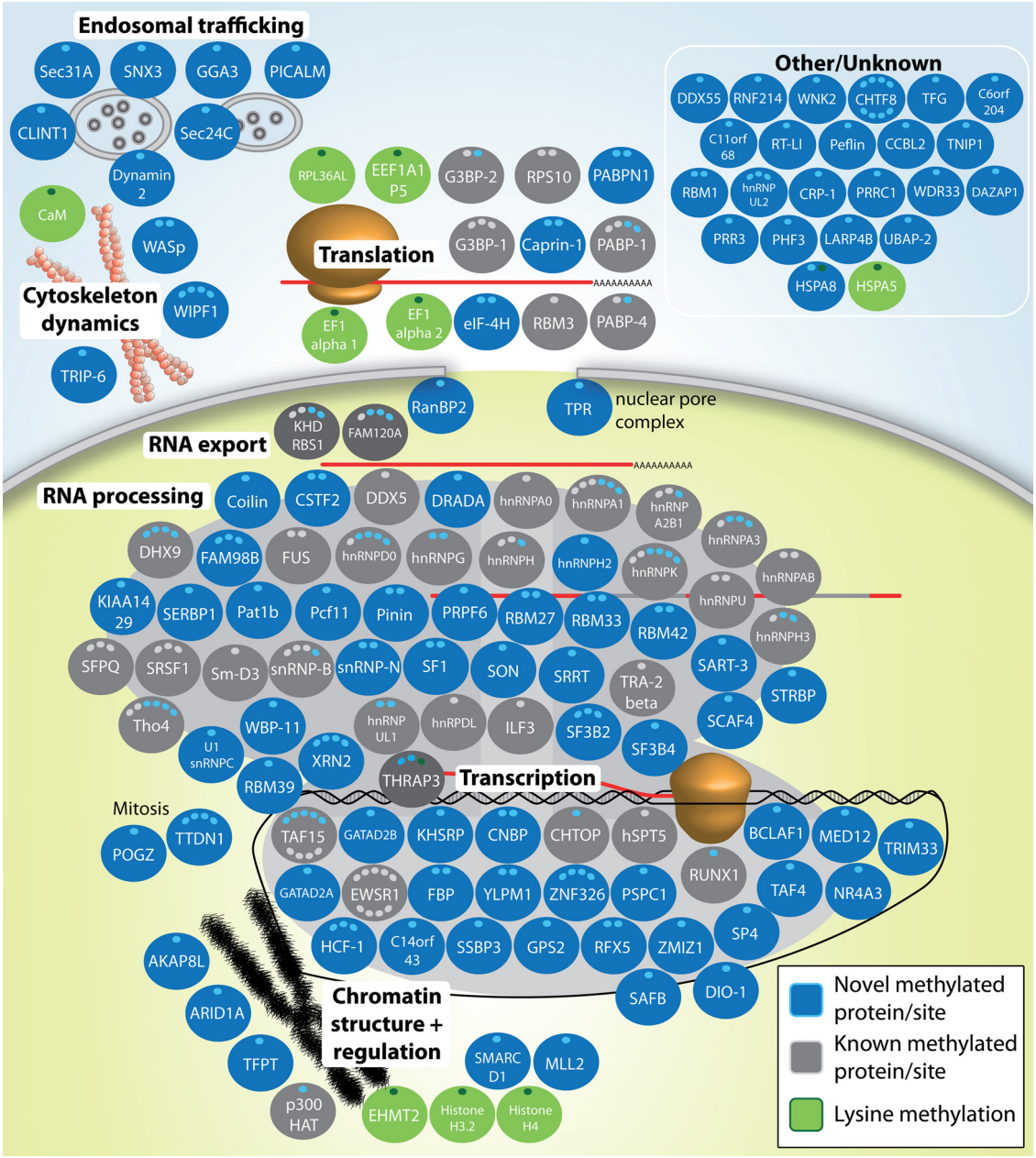


FIG. 5. **Overview of identified proteins containing methylated arginine/lysine.** Proteins are grouped according to annotated function in the UniprotKB database. Proteins newly identified as PRMT substrates in this study are drawn in *dark blue*. Previously known PRMT substrates are shaded in *gray*. Individual arginine methylation sites are marked as *dots*, *blue* representing novel Meth-R sites and *gray* representing previously known Meth-R sites. Green denotes a protein methylated on lysine.

HILIC performance overshadowed SCX and IEF, it missed 40 Meth-R sites identified by the latter two methods, despite a similar number of analysis performed. However, pI and GRAVY score of HILIC and SCX/IEF Meth-R peptides did not significantly differ, suggesting that the major cause of missed-hits by HILIC could be a persistent complexity of the peptide mixture in the HILIC fractions. In an effort to discover higher numbers of Meth-R sites after HILIC enrichment, we made use of a Q Exactive Orbitrap (see Methods), a powerful new MS/MS instrument that allows faster fragmentation rates, hence higher numbers of peptide identified (39). This allowed us to confirm in a single experiment about half of Meth-R sites identified by several analyses of HILIC fractions on an LTQ XL Orbitrap and led us to discover 11 new ones, indicating that the Q Exactive is largely preferable for proteomics scale discovery of Meth-R. Meth-R proteins detected in Jurkat cells and primary CD4+ lymphocytes overlapped only partially (supplemental Fig. S4). Possible differences in protein expression and/or PRMT levels in the two cell types may explain this divergence, considering that

Jurkat is a T cell lymphoma and that cancer cells may express modified repertoires of Meth-R proteins (24). Similarly, differences in tissue expression may explain why our data missed 18 Meth-R proteins in the mammalian proteome.

Only 25% of Meth-R sites annotated to date are at non-RG sites (supplemental Table S2). Some contain prolines and may be CARM1 substrates (21). In contrast to the use of 7E6 mAb, HILIC (as well as SCX and IEF) enrichment seems to be unbiased, as almost half of the sites we identified were at non-RG sites (supplemental Table S1 and Fig. 3*A*). The relatively low number of Meth-R sites known to date are largely derived from the most abundant proteins implicated in RNA metabolism (supplemental Table S3), as RG sites tend to occur at repeated GAR sequences of RNA binding proteins. However, only 7 such proteins (of 56 total) cover half of the 137 Meth-R sites known to date (see supplemental Tables S2 and S3). Thus, HILIC alone removes the above biases and reveals that PRMTs control proteins implicated in many diverse cellular functions.

Meth-R occurring at GAR sequences of RNA-binding proteins may contribute to protein-RNA interaction (2). The PRAM sequences defined in our study could be part of protein-protein interaction sites and Meth-R may modulate such interactions (40). The PRxxxP motif is predicted to be a poly-proline II helix able to interact with SH3 and other protein domains. Inspection of methylated PRxxxP motif revealed that a hydrophobic amino acid often precedes the +4 proline (PRxxΦP) (supplemental Fig. S7), a signature for poly-proline II helix interacting with SH3 domains and supports a role for PRMTs in modulating protein-protein interactions (40). Interestingly, our data revealed that the WASP/WIP complex and DNM-2, both involved in the formation and maintenance of the IS induced by TCR signaling, contained PRAM sequences known to be part of SH3 binding sites for several binding partners (37, 38). These data, together with the identification of novel Meth-R proteins implicated in endosomal trafficking, chromatin remodeling, transcription, RNA processing and translation open new perspectives to further uncover the role of PRMTs in the T cell activation process.

While HILIC represents a simple and robust method to enrich for Meth-R peptides, a few technical improvements lie ahead for routine "methylome"-scale studies. One immediate possibility to de-complex further peptide mixtures prior to MS analysis is orthogonal combination of two chromatographic methods according to the physico-chemical features of Meth-R-containing peptides highlighted in this study. Large/highly charged peptides might escape detection by MS, a problem that can be partly overcome using proteases other than trypsin. Indeed, chymotrypsin enabled us to identify a sizable number of Meth-R-containing peptides not found in trypsin digests. Moreover, SCX, IEF and, preferentially HILIC, can be readily employed in small-scale formats to discover arginine methylation sites in low complexity peptide mixtures such as those derived from digests of purified protein complexes or individual proteins, proven or suspected to interact with PRMTs.

The discovery of methods for unbiased enrichment of Meth-R-containing peptides, such as HILIC, will also allow us to ask why and to what extent abnormal protein arginine methylation plays a role in carcinogenesis (*e.g.* prostate cancer and breast cancer) and several other, important pathologies, including cardiovascular and infectious diseases, contributing perhaps to opening avenues for new therapeutic tools and drugs.

* This work was supported by Wellcome Trust Grant GR076558MA to O.A. V.G. was supported by a PhD studentship from the BBSRC, T.U. was supported by a Swiss National Science Foundation fellowship and a Roche Research Foundation fellowship.

S This article contains supplemental Figs. S1 to S7 and Tables S1 to S5.

¶ T.U. and V.G. contributed equally to this work.

‖ Current address: Dualsystems Biotech AG, Grabenstrasse 11a, 8952 Schlieren, Switzerland.

** Current address: Biochemistry Dept and Proteomics Core UT Southwestern Medical Center at Dallas 6001 Forest Park Rd, Dallas, TX 75390-8816, USA.

‡‡ To whom correspondence should be addressed: T Cell Signalling Laboratory, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom. Tel.: ++44 1865 275615; Fax: ++44 1865 275515; E-mail: oreste.acuto@path.ox.ac.uk.

REFERENCES

1. Bulau, P., Zakrzewicz, D., Kitowska, K., Wardega, B., Kreuder, J., and Eickelberg, O. (2006) Quantitative assessment of arginine methylation in free versus protein-incorporated amino acids in vitro and in vivo using protein hydrolysis and high-performance liquid chromatography. *BioTechniques* **40,** 305–310
2. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33,** 1–13
3. Sprangers, R., Groves, M. R., Sinning, I., and Sattler, M. (2003) High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J. Mol. Biol.* **327,** 507–520
4. Yang, Y., Lu, Y., Espejo, A., Wu, J., Xu, W., Liang, S., and Bedford, M. T. (2010) TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol. Cell* **40,** 1016–1023
5. Yu, Z., Chen, T., Hébert, J., Li, E., and Richard, S. (2009) A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Mol. Cell. Biol.* **29,** 2982–2996
6. Kim, J., Lee, J., Yadav, N., Wu, Q., Carter, C., Richard, S., Richie, E., and Bedford, M. T. (2004) Loss of CARM1 results in hypomethylation of thymocyte cyclic AMP-regulated phosphoprotein and deregulated early T cell development. *J. Biol. Chem.* **279,** 25339–25344
7. Swiercz, R., Person, M. D., and Bedford, M. T. (2005) Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochem. J.* **386,** 85–91
8. Ancelin, K., Lange, U. C., Hajkova, P., Schneider, R., Bannister, A. J., Kouzarides, T., and Surani, M. A. (2006) Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat. Cell Biol.* **8,** 623–630
9. Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Lüscher, B., and Amati, B. (2007) Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449,** 933–937
10. Iberg, A. N., Espejo, A., Cheng, D., Kim, D., Michaud-Levesque, J., Richard, S., and Bedford, M. T. (2008) Arginine methylation of the histone H3 tail impedes effector binding. *J. Biol. Chem.* **283,** 3006–3010

11. Jelinic, P., Stehle, J. C., and Shaw, P. (2006) The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.* **4,** e355
12. Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., Cerruti, L., Curtis, D. J., Patel, D. J., Allis, C. D., Cunningham, J. M., and Jane, S. M. (2009) PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* **16,** 304–311
13. Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Nimer, S. D. (2008) Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev.* **22,** 640–653
14. Jansson, M., Durant, S. T., Cho, E. C., Sheahan, S., Edelmann, M., Kessler, B., and La Thangue, N. B. (2008) Arginine methylation regulates the p53 response. *Nat. Cell Biol.* **10,** 1431–1439
15. Covic, M., Hassa, P. O., Saccani, S., Buerki, C., Meier, N. I., Lombardi, C., Imhof, R., Bedford, M. T., Natoli, G., and Hottiger, M. O. (2005) Arginine methyltransferase CARM1 is a promoter-specific regulator of NF-kappaB-dependent gene expression. *EMBO J.* **24,** 85–96
16. Chen, D., Ma, H., Hong, H., Koh, S. S., Huang, S. M., Schurter, B. T., Aswad, D. W., and Stallcup, M. R. (1999) Regulation of transcription by a protein methyltransferase. *Science* **284,** 2174–2177
17. Fathman, J. W., Gurish, M. F., Hemmers, S., Bonham, K., Friend, D. S., Grusby, M. J., Glimcher, L. H., and Mowen, K. A. (2010) NIP45 controls the magnitude of the type 2 T helper cell response. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3663–3668
18. Liu, Q., and Dreyfuss, G. (1995) In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* **15,** 2800–2808
19. Ohkura, N., Takahashi, M., Yaguchi, H., Nagamura, Y., and Tsukada, T. (2005) Coactivator-associated arginine methyltransferase 1, CARM1, affects pre-mRNA splicing in an isoform-specific manner. *J. Biol. Chem.* **280,** 28927–28935
20. Kwak, Y. T., Guo, J., Prajapati, S., Park, K. J., Surabhi, R. M., Miller, B., Gehrig, P., and Gaynor, R. B. (2003) Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol. Cell* **11,** 1055–1066
21. Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* **25,** 71–83
22. Blanchet, F., Cardona, A., Letimier, F. A., Hershfield, M. S., and Acuto, O. (2005) CD28 costimulatory signal induces protein arginine methylation in T cells. *J. Exp. Med.* **202,** 371–377
23. Cimato, T. R., Tang, J., Xu, Y., Guarnaccia, C., Herschman, H. R., Pongor, S., and Aletta, J. M. (2002) Nerve growth factor-mediated increases in protein methylation occur predominantly at type I arginine methylation sites and involve protein arginine methyltransferase 1. *J. Neurosci. Res.* **67,** 435–442
24. Cheung, N., Chan, L. C., Thompson, A., Cleary, M. L., and So, C. W. (2007) Protein arginine-methyltransferase-dependent oncogenesis. *Nat. Cell. Biol.* **9,** 1208–1215
25. Wang, Z., Tang, W. H., Cho, L., Brennan, D. M., and Hazen, S. L. (2009) Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. *Arterioscler. Thromb. Vasc. Biol.* **29,** 1383–1391
26. Bedford, M. T., and Richard, S. (2005) Arginine methylation an emerging regulator of protein function. *Mol. Cell* **18,** 263–272
27. Copeland, R. A., Solomon, M. E., and Richon, V. M. (2009) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug discovery* **8,** 724–732
28. Ong, S. E., Mittler, G., and Mann, M. (2004) Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–126
29. Trudgian, D. C., Thomas, B., McGowan, S. J., Kessler, B. M., Salek, M., and Acuto, O. (2010) CPFP: a central proteomics facilities pipeline. *Bioinformatics* **26,** 1131–1132
30. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
31. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3,** 958–964
32. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75,** 4646–4658
33. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12130–12135
34. Jian, W., Edom, R. W., Xu, Y., and Weng, N. (2010) Recent advances in application of hydrophilic interaction chromatography for quantitative bioanalysis. *J. Separation Sci.* **33,** 681–697
35. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **7,** 971–980
36. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132
37. Zhang, J., Dong, B., and Siminovitch, K. A. (2009) Contributions of Wiskott-Aldrich syndrome family cytoskeletal regulatory adapters to immune regulation. *Immunolog. Rev.* **232,** 175–194
38. Gomez, T. S., Hamann, M. J., McCarney, S., Savoy, D. N., Lubking, C. M., Heldebrant, M. P., Labno, C. M., McKean, D. J., McNiven, M. A., Burkhardt, J. K., and Billadeau, D. D. (2005) Dynamin 2 regulates T cell activation by controlling actin polymerization at the immunological synapse. *Nat. Immunol.* **6,** 261–270
39. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11,** 10.1074/mcp.M111.013722
40. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J. Biol. Chem.* **275,** 16030–16036
41. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097–6100