# Identification and Initial Characterisation of Protein Methylation Sites in *Saccharomyces cerevisiae*


Michael Plank

Merton College

University of Oxford

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy
Trinity Term, 2015

# Identification and Initial Characterisation of Protein Methylation Sites in *Saccharomyces cerevisiae*

Michael Plank

Merton College

University of Oxford

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Trinity Term, 2015

## Abstract

The field of protein arginine methylation has seen a surge in the number of known methylation sites in higher eukaryotes over recent years, but our understanding of the functions of these modifications is lagging behind. A better knowledge of arginine methylation in a model organism like *Saccharomyces cerevisiae* that is relatively amenable to experimental manipulation may inform on arginine methylation also in higher eukaryotes.

The proteome-wide identification of methylation sites in *S. cerevisiae*, based on liquid chromatography tandem mass spectrometry in combination with several enrichment and fractionation strategies and on the use of heavy methyl SILAC is presented here. The majority of sites identified are in good agreement with previous reports of arginine methylation, i.e. occur in glycine-arginine-rich regions, locate to proteins involved in RNA-processing and overlap with known sites. Additionally, the data obtained suggest a bias for localisation of arginine methylation in unstructured protein regions and asparagine-rich contexts. It is further shown that the use of heavy methyl SILAC is useful for confident site identification.

To overcome challenges in the global approach, the detection of methylation sites on overexpressed candidate proteins was also employed. This strategy yielded methylation sites on more than half of the 38 successfully purified proteins, but also brought to light challenges in the localisation of sites.

Subsequently, it was investigated which methyltransferases are responsible for methylating mapped sites on two of these proteins. These experiments strongly indicate that Lsm4 is neither exclusively methylated by Hmt1 nor Hsl7, while Hmt1 appears to be the sole methyltransferase responsible for methylating Scd6. It was further shown that the methylation of the glycine-arginine-rich region of Scd6 is reduced upon prolonged sodium azide stress which may have potential implications in stress-induced translational repression.

Overall, these results both inform on methods for the assignment of arginine methylation and improve its biological characterisation.

# Acknowledgements

First and foremost, I would like to thank my supervisors Benedikt, Catherine and Chris for their time and effort dedicated to supervising my PhD project.

Several people have have helped me with their advice on experimental techniques and data analysis. In particular, I would like to thank Francoise, Anitha, Chris, Sarah, Milan, Rebecca, Roman, Zhanru, Phil, Michael, Keith, Shabaz, Cynthia, Dave, Katherine, Andrea (University of Oxford), Karin (University of Manchester), David (EBI, Cambridge) and all others who have contributed to this work, but are not listed here. I`m grateful to Vincent for his help at several instances during this project and for allowing me to use MethylQuant and to Peter for his advice and for sharing reagents. I would also like to acknowledge the support I have received via online communities and by suppliers of instruments and software, especially by Jenny (Thermo), Will (Agilent), Katherine and Dan (Bioinformatics Solutions). I thank Katalin for peptide synthesis and Heidi for proof-reading. I am thankful for plasmids and materials that were kindly provided by the Green (University of Oxford), Bonaldi (European Institute of Oncology), Wang (Chinese Academy of Sciences) and Gozani (Stanford University) groups. I thank Heather, Louise, Ray and the other members of the Oxford Protein Production facility for their help with recombinant protein production and the MRC for funding this part of my work.

Finally, I gratefully acknowledge the funding received from the Biochemical Society, the BMSS, Merton College and the EPSRC.

# Contents

I certify that this is entirely my own work unless otherwise stated.

Some figures and tables as indicated in their legends are reproduced and parts of the text are adapted from Plank et al., 2015 with permission of John Wiley and Sons.

## List of Figures

## List of Tables

## List of Abbreviations

| | |
|---|---|
| 6xHis-tag | hexa-histidine tag |
| ATP | adenosine tri-phosphate |
| BCA | bicinchoninic acid |
| CID | collision induced dissociation |
| DTT | dithiothreitol |
| ETD | electron transfer dissociation |
| FDR | false discovery rate |
| Glc | glucose |
| GO | gene ontology |
| GST | glutathione S-transferase |
| HCD | higher energy collisional dissociation |
| HILIC | hydrophilic interaction liquid chromatography |
| HIV | human immunodeficiency virus |
| hnRNP | heterogeneous nuclear ribonucleoprotein |
| IP | immunoprecipitation |
| LB | lysogeny broth |
| LC-MS/MS | liquid chromatography tandem mass spectrometry |
| MBP | maltose binding protein |
| MBT | malignant brain tumour |
| MS | mass spectrometry |
| MTase | methyltransferase |
| m/z | mass-to-charge ratio |
| PRMT | protein arginine methyltransferase |
| PSM | peptide spectrum match |
| PTM | post-translational modification |
| Rme1 | NG-mono-methyl arginine |
| Rme2as | NG,NG-di-methyl arginine |
| Rme2sym | NG,NG'-di-methyl arginine |
| SAM | S-adenosyl methionine |
| SC | synthetic complete |
| SD | synthetic defined |
| SILAC | stable isotope labelling with amino acids in cell culture |
| snRNA | small nuclear ribonucleic acid |
| snRNP | small nuclear ribonucleoprotein |
| TFA | trifluoroacetic acid |
| wt | wild-type |
| YPD | yeast extract, peptone, dextrose |
| δ-Rme1 | δ-mono-methyl arginine |

# 1. Introduction

Post-translational modifications (PTMs) to proteins play important roles in regulating their catalytic activity, interactions, stability and sub-cellular localisation. More than 400 types of chemical modifications to proteins, including phosphorylation, acetylation, ubiquitinylation and methylation, have been described (Magrane & Consortium 2011). The methylation of proteins on lysine-, arginine-, aspartate-, glutamate-, asparagine-, glutamine-, histidine-, cysteine-, isoaspartate-, N- and C-terminal residues has been observed (see "3.1.4. Methylation on Residues other than Lysine and Arginine"). An overview of the properties and biology of protein arginine methylation, the main focus of the experimental work described in this thesis, is presented in the following sections.

## 1.1. Forms and Chemical Properties of Protein Arginine Methylation

Arginine methylation can occur as the addition of a single methyl-group to one of the ω-amino groups of arginine (mono-methyl arginine; Rme1) or with two methyl-groups added to the terminal amino groups, either both on the same ω-amino group (asymmetric di-methyl arginine; Rme2as) or one on each (symmetric di-methyl arginine; Rme2sym) (Paik & Kim 1968). Additionally, mono-methylation of the δ-position (δ-Rme1) has been reported in *S. cerevisiae*, but not in higher eukaryotes (Fig. 1.1) (Zobel-Thropp et al. 2000). Methyl-groups are transferred to proteins enzymatically by protein arginine methyltransferases (PRMTs) using S-adenosyl methionine (SAM) as the methyl-donor. It has been estimated in a rat cell line that Rme2as accounts for ~70%, Rme1 to ~20% and Rme2sym ~3% of cellular protein N-methylation (Najbauer & Aswad 1990). Arginine methylation has been found in organisms from yeast to man, but appears to be absent in prokaryotes (Bachand 2007).

While N-methylation does not alter the positive charge of arginine at physiological pH, one of its five potential hydrogen-bond donors is lost with the addition of each methyl-group.

Furthermore, methylation leads to a slight increase in size and hydrophobicity of the guanidino-group of the arginine side chain.

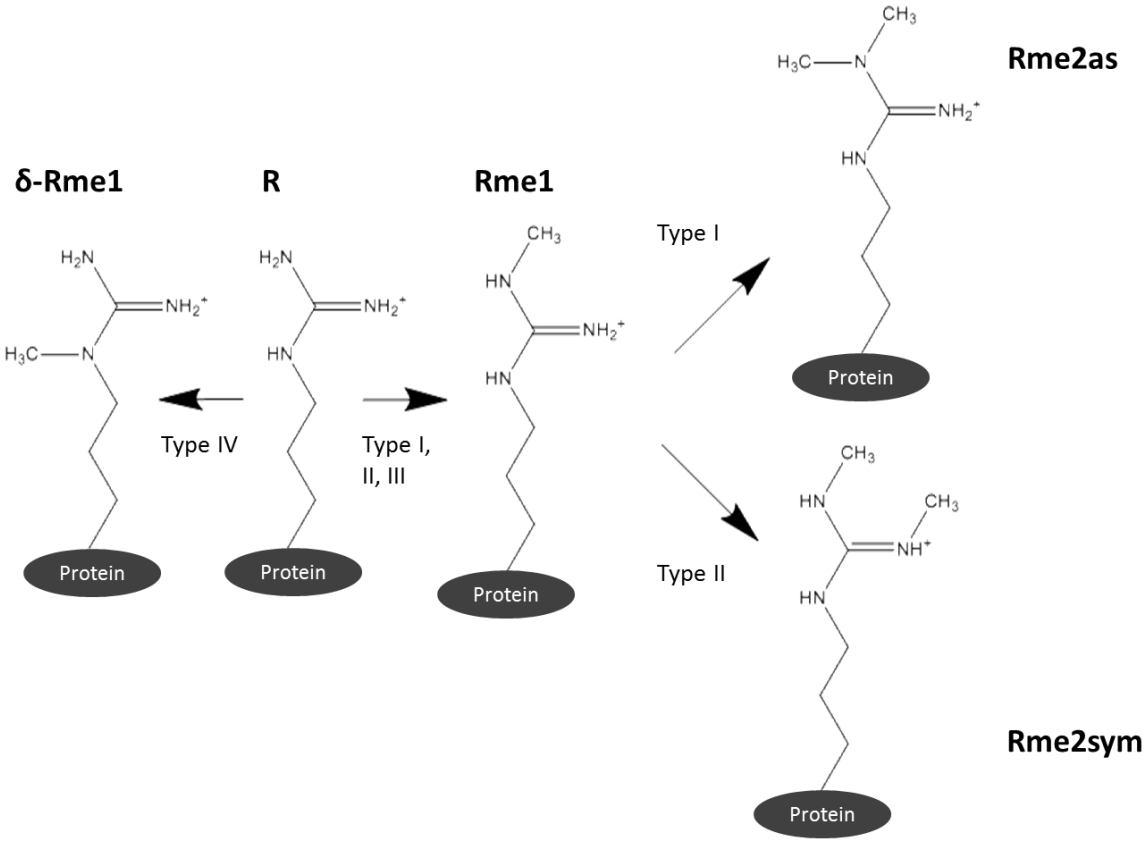



***Figure 1.1: Forms of protein arginine methylations.*** *The formation of δ-mono-methyl arginine (δ-Rme1) from arginine (R) is catalysed by type IV and of ω-mono-methyl arginine (Rme1) by type I, II and III PRMTs. Type I PRMTs are further responsible for the formation of asymmetric di-methyl arginine (Rme2as) and type II PRMTs of symmetric di-methyl arginine (Rme2sym).*

## 1.2. Mechanistic Role of Arginine Methylation

### 1.2.1. Protein-Protein Interactions

While lysine methylation is recognized e.g. by chromo- and malignant brain tumour (MBT) domains (Li et al. 2007), tudor domains may bind to methylated lysines or arginines (Dawson et al. 2012; Lu & Wang 2013). For example, the tudor domain of SMN has been observed to bind to symmetrically di-methylated arginines of Sm proteins and the Sm-like protein Lsm4 (see also "1.3. Functional Role of Arginine Methylation in mammals") (Brahms et al. 2001). The interaction of the tudor domains of SPF30 and TDRD3 with methylated arginines in glycine-arginine rich regions also has been reported (Côté & Richard 2005), as has the

binding of the SMN tudor domain to asymmetrically di-methylated arginine (Cheng et al. 2007). The binding of methyl-arginine to the tudor domain of SMN is thought to occur via cation-pi interactions and arginine methylation has been shown to increase this type of interaction in an *in vitro* model (Sprangers et al. 2003; Hughes & Waters 2006). It should be noted, that no tudor domain containing protein has been detected in *S. cerevisiae* to date (Pfam v. 28.0; PF00567).

Another reader domain for methylated arginine residues is the WD40 domain which has been reported to bind Rme2sym (Migliori et al. 2012).

On the other hand, a negative role for arginine methylation on protein-protein interactions has also been described. For example, arginines of the RG-repeats neighbouring the proline rich region of Sam68 can be asymmetrically di-methylated. This di-methylation impairs interactions with SH3 domains (Bedford et al. 2000). In yeast, the methylation of the RNA-binding protein Npl3 inhibits its interaction with Tho2, a member of the transcription/export (TREX) complex (Yu et al. 2004). Similarly, the interaction between Npl3 and the small nuclear ribonucleoprotein (snRNP) protein Snp1 has been reported to be increased in the absence of PRMT Hmt1 (Chen et al. 2010). A later report however contradicted these findings, observing a slightly higher interaction between Npl3 and Snp1 in the presence than in the absence of Hmt1 in a bacterial two-hybrid system (Erce et al. 2013). The same study reported increased Npl3-Air2 and Npl3-Ded1 interactions when Hmt1 was expressed in the two-hybrid system.

Antagonistic functions of symmetric and asymmetric di-methylation have been reported in the case of histone mark H3R2 in human cells: While H3R2me2sym promotes the binding of WDR5, a member of the co-activator complex, to nucleosomes, H3R2me2as blocks binding of the same protein. Consequently H3R2sym marks are enriched in euchromatic, while H3R2as is more prevalent in heterochromatic regions (Migliori et al. 2012; Iberg et al. 2008).

Interestingly, opposing effects of the same type of methylation on the interaction of different proteins with the same target have also been observed: While symmetric di-methylation of Sm proteins facilitates the interaction with the tudor domain of SMN (Brahms et al. 2001), this domain binds to EWS in the unmethylated, but not the symmetrically di-methylated state (Young et al. 2003).

### 1.2.2. Protein-RNA Interactions

Arginine is an important residue for protein-nucleic acid interaction due to its positive charge as well as up to five hydrogen bonds that are frequently observed between arginine and the bases and backbone of DNA and RNA (Luscombe et al. 2001). For example, in a study of the human immunodeficiency virus (HIV) Tat protein binding to its RNA target (TAR), the overall charge density of the relevant protein region, as well as the presence of a specific arginine, were found to be important for the interaction. Based on these results an 'arginine-fork' model was proposed in which arginine forms two hydrogen bonds each with two adjacent phosphates of an RNA bulge (Calnan et al. 1991) (Fig. 1.2).

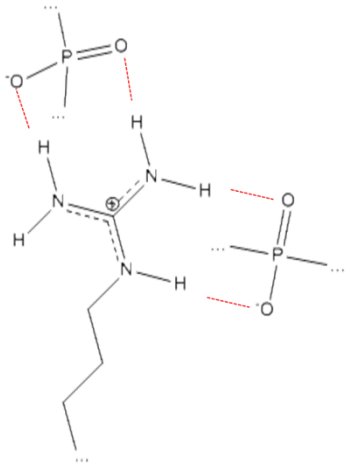


***Figure 1.2: Model of protein-RNA interaction via an arginine-fork.*** *The planar guanidino-group of an arginine residue forms two hydrogen bonds each (red, dotted lines) with two phosphates of RNA. Omitted parts of the RNA and protein chain are indicated by '…'. Adapted from Calnan et al. (1991).*

These hydrogen bonds can be abolished by arginine methylation: For example, asymmetric di-methylation of several arginine residues of the HIV Rev protein in its basic RNA-binding region, which is similar to the one found in Tat, leads to a decreased interaction with the Rev responsive element (Hyun et al. 2008). Further cases of reduced protein interaction with RNA

upon arginine methylation have been reported, e.g. for the heterogeneous nuclear ribonucleoprotein (hnRNP) A1 (Rajpurohit et al. 1994), CNBP (Wei et al. 2014) and FMRP (Denman 2002). A weakening of its interaction with RNA upon arginine methylation has further been described for REF/ALY, the human homologue of yeast Yra1. The process has been suggested to be important for the handing over of the RNA to the export factor TAP/NXF1 (Hung et al. 2010). It should be noted that protein-RNA interactions are often affected by the methylation of RGG-boxes which constitute both arginine methylation target sequences and RNA-binding motifs (Kiledjian & Dreyfuss 1992). However, they also function in protein-protein interactions via disordered regions (Thandapani et al. 2013). For example, tudor domains frequently interact with methyl-arginines in RGG-boxes and the high affinity of the interaction is attributed to the increased local concentration of methylated arginines (Tripsianes et al. 2011).

### 1. 3. Functional Role of Arginine Methylation

Diverse biological consequences may result from the altered protein-protein and protein-RNA interactions caused by arginine methylation. The main known functions of arginine methylation are summarized in the following sections.

#### 1.3.1. Transcriptional regulation

Much effort has been invested in understanding the effect of PTMs on histone tails on transcription regulation. Arginine methylation has been observed in eukaryotes on H4R3, H3R2, H3R8, H3R17 and H3R26 (Di Lorenzo & Bedford 2011).

Interestingly, in *S. cerevisiae* the asymmetric di-methylation of H4R3 by Hmt1 leads to silencing of chromatin, while in humans it is associated with activation (Lacoste et al. 2002; Wang et al. 2001, p.4). Arginine methylation-dependent chromatin silencing in S. *cerevisiae* involves the recruitment of histone deacetylase Sir2 and Hmt1 activity (Yu et al. 2006). Transcriptional repression based on the H3R2me2as mark has been observed in *S. cerevisiae*,

which likely relies on the inhibition of the formation of the activating H3K4me3 mark (Kirmizis et al. 2007). In contrast, the H3R2me1 and H3R2me2sym modifications constitute transcription activating marks (Kirmizis et al. 2009).

Several factors involved in transcription other than histones are arginine methylated in *S. cerevisiae*. The heavily mono- and asymmetrically di-methylated hnRNP protein Npl3 acts as a transcriptional anti-terminator, possibly by inhibiting access of a component (Hrp1) of the cleavage/termination factor CFI to the RNA. The recruitment of Npl3 to RNA relies on Hmt1 activity and consequently the rate of termination at cryptic terminators is increased in *hmt1* deletion strains (Wong et al. 2010).

Additionally, Hmt1 activity increases PolII processivity. Tho2 is also required for PolII processivity and less Tho2 is recruited to transcripts in *hmt1* deletion strains. Based on the finding that Npl3 methylation reduces the interaction between Npl3 and Tho2, it has been suggested that a decrease in processivity in *hmt1* deletion strains may be due to the sequestration of available Tho2 by a blocked release from hnRNPs (Wong et al. 2010).

### 1.3.2. Splicing

The spliceosome is formed of five small nuclear ribonucleoprotein particles (snRNPs) and several non-snRNP proteins. The snRNPs in turn consist of seven Sm-proteins assembled as a ring on small nuclear ribonucleic acids (snRNAs) (Yong et al. 2004). This cytoplasmic assembly is mediated by the SMN protein which in human cells binds symmetrically di-methylated arginines in the C-terminal RG-rich region of SmD1 and SmD3 via its tudor domain (Brahms et al. 2000). Subsequently, a similar role for symmetric di-methyl arginines was also found in SmB/B' and Lsm4 (Brahms et al. 2001). The involvement of di-methylated arginines in the cytoplasmic maturation of snRNPs is supported by the finding that Sm proteins accumulate in the cytoplasm upon inhibition of methylation (Côté & Richard 2005).

Also the further development of snRNPs in Cajal bodies of the nucleus requires arginine methylation (Boisvert et al. 2002).

In *S. cerevisiae*, the methylation of Sm or Sm-like (Lsm) proteins has not been observed to date, potentially reflecting the much lower prevalence of RNA splicing in this organism (Meyer & Vilardell 2009). The methylation of a non-Sm snRNP protein, Snp1, however has been established and there are some indications that this methylation is required for the recruitment of splicing factors to RNA (Chen et al. 2010).

### 1.3.3. Nuclear shuttling

Arginine methylation facilitates the nuclear export of the *S. cerevisiae* protein Npl3 (Shen et al. 1998). Interestingly, the nuclear export of Npl3 does not rely on methylated arginines on the protein, but rather the absence of unmethylated arginines. This was demonstrated by the fact that a mutant version of the protein with arginines converted to lysines could be exported independently of the presence of Hmt1, while in the wild-type (wt) Hmt1 activity was required (Xu & Henry 2004). There are indications that this shuttling is facilitated by disruption of interactions between Npl3 with Tho2 and of Npl3 self-association via arginine methylation by Hmt1 (McBride et al. 2005). The export of the poly-A binding protein Nab2 is also dependent on Hmt1 activity (Green et al. 2002). Finally, the nuclear export of Hrp1 requires the methylation of Npl3, rather than depending on its own methylation state (Xu & Henry 2004). It should be noted that the export of Npl3, Nab2 and Hrp1 depends on on-going transcription, consistent with the fact that these proteins are co-transcriptionally recruited to RNA (Green et al. 2002; Wong et al. 2010; Shen et al. 1998).

In contrast, for mammalian proteins arginine methylation appears to be important for their localisation to the nucleus (Passos et al. 2006). For example, the human hnRNP A2 protein is re-localized from the nucleus to the cytoplasm if the methylation of its RGG-domain is inhibited (Nichols et al. 2000). Further examples are Sam68 and RNA helicase A (Côté et al.

2003; Smith et al. 2004). Finally, the export of viral RNA of HIV is impaired upon overexpression of the arginine methyltransferase PRMT6, which is likely due to a reduction of the association of the RNA with Rev protein upon methylation (Invernizzi et al. 2006).

### 1.3.4. Further Functions of Arginine Methylation

Interestingly, it has been observed that the HIV Tat, NC and Rev proteins are arginine methylated by a host methyltransferase (PRMT6), leading to a reduction of their respective functions (Boulanger et al. 2005; Invernizzi et al. 2007; Invernizzi et al. 2006). Arginine methylation may therefore also constitute a mechanism of the host-defence system.

In higher eukaryotes, roles of arginine methylation in DNA damage repair (Auclair & Richard 2013) and signalling (Biggar & Li 2015) have also been established. Finally, arginine methylation plays a role in protein translation which is discussed in section "3.1.3. Methylation of Proteins involved in Translation".

## 1.4. Protein Arginine Methyltransferases

Methyl-groups are added to polypeptide chains by protein methyltransferases (MTases) which use SAM as a methyl donor. Most PRMTs are members of the seven-beta strand protein structural family, while most lysine MTases belong to the SET-domain family (Clarke 2013). In contrast to lysine MTases, which appear to be relatively specific with respect to their target proteins, sites and type of (i.e. mono-, di- and tri-) methylation, PRMTs generally methylate more than one target site (Clarke 2013). They are classified into four groups by their activities, with type I catalysing the generation of $N^G$-mono-methyl arginine (Rme1) and asymmetric $N^G,N^G$-di-methyl arginine (Rme2as), type II of Rme1 and symmetric $N^G,N^{G'}$-di-methyl arginine (Rme2sym) and type III of Rme1 only. Finally, type IV PRMTs catalyse the mono-methylation of the δ-position of arginine. Of the known human PRMTs, PRMT1 (type I), PRMT3 (type I) and PRMT5 (type II) are conserved down to unicellular eukaryotes

(Bachand 2007; Krause et al. 2007). A type IV MTase has only been reported in *S. cerevisiae* to date (Niewmierzycka & Clarke 1999; Chern et al. 2002) (Fig. 1.1).

While no PRMT3 homologue is known in *S. cerevisiae*, Hmt1 (Rmt1) is homologous to PRMT1 and Hsl7 to PRMT5 (Gary et al. 1996; Henry & Silver 1996; Miranda et al. 2006; Lee et al. 2000). Rmt2 is an additional seven-beta strand PRMT present in *S. cerevisiae* which belongs to type IV, while Sfm1 is structurally distinct and, as it has only been found to mono-methylate its sole known target is proposed to be a type III PRMT (Niewmierzycka & Clarke 1999; Young et al. 2012). There are indications that suggest the existence of further uncharacterised PRMTs in *S. cerevisiae* (Low & Wilkins 2012; Low et al. 2013).

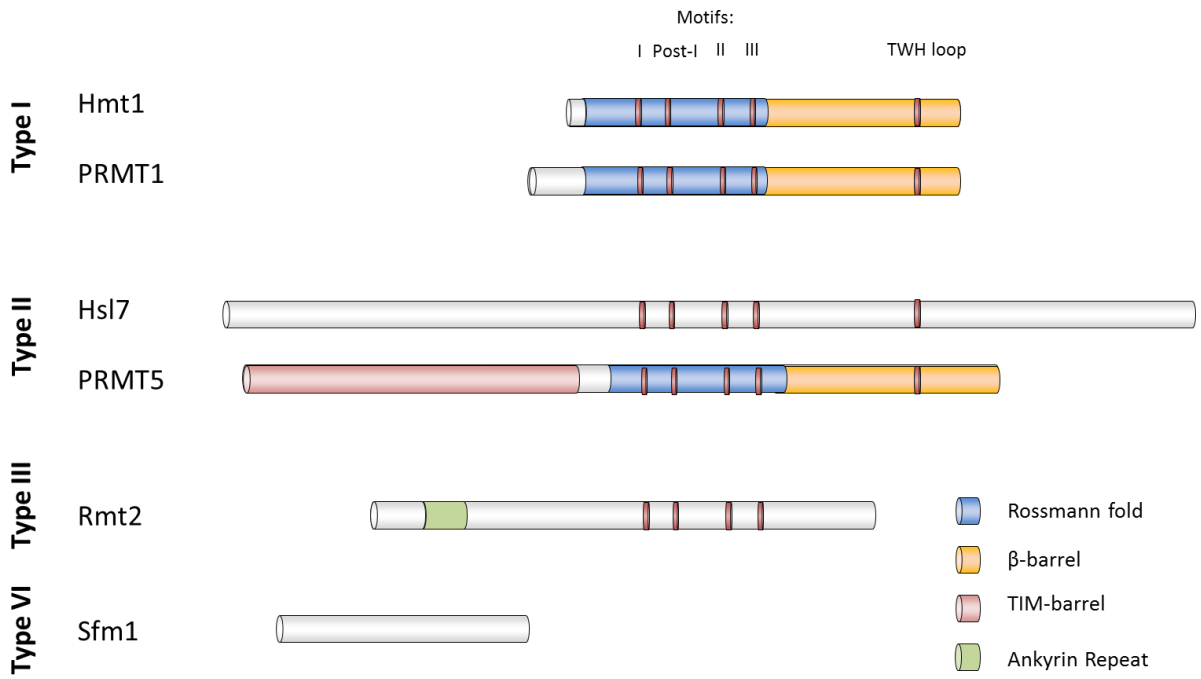



***Figure 1.3: Schematic representation of S. cerevisiae PRMTs and their human homologues.*** *The motifs I, post-I, II and III, the TWH loop and the ankyrin repeat domain of Rmt2 are indicated based on the protein sequences. The Rossmann fold, β-barrel and TIM-barrel are shown on PRMTs with solved structures based on Zhang & Cheng* (2003), *Sun et al.* (2011) *and Weiss et al.* (2000).

PRMTs of the seven-beta strand family share a MTase domain of a Rossmann-fold (Rossmann et al. 1974) which harbours the conserved motifs I (VLD/EVGxGxG), post-I (V/IxG/AxD/E), II (F/I/VDI/L/K) and III (LR/KxxG) implicated in SAM-binding (Zhang et

al. 2000; Bedford 2007) (Fig. 1.3). In PRMT1, Hmt1 and Hsl7, for which the structure has been solved, this domain was found joined to a C-terminal β-barrel via a dimerisation loop. Hmt1 dimerises to form a ring, with the active site near the central cavity. This dimerisation is important for the catalytic function of Hmt1. At least at high concentrations (0.1-4 mg/ml) *in vitro*, further association of three dimers occurs (Weiss et al. 2000). Hsl7 also appears to self-associate, but the exact oligomerisation state is unknown (Cid et al. 2001, p.7).

Based on the finding that hnRNPs are often heavily arginine methylated in higher eukaryotes, the PRMT Hmt1 was identified via its synthetic lethality with functionally impaired hnRNP Npl3 (Henry & Silver 1996). Hmt1, which mainly localizes to the nucleus, is considered the predominant PRMT in *S. cerevisiae* and methylates both histone and non-histone proteins (Gary et al. 1996; Lacoste et al. 2002), however is not essential for cell viability (Henry & Silver 1996). While no major phenotypic effect was observed upon *hmt1* deletion, a delay in the processing of certain RNAs and impaired nuclear export of hnRNPs was reported (Yu et al. 2004).

Hsl7 has been shown to mono- and symmetrically di-methylate calf histone H2A *in vitro*, while no *in vivo* substrate is known (Miranda et al. 2006; Sayegh & Clarke 2008; Lee et al. 2000). Hsl7 dynamically localizes to the bud-neck and spindle pole body and is implicated in cell cycle progression via Swe1 inhibition (Cid et al. 2001, p.7). However, the elongated-bud phenotype of *hsl7Δ* strains has been found to be independent of its MTase function (Ruault & Pillus 2006).

Rmt2 is an unusual PRMT as it mono-methylates the δ-amino group of arginine, which has only been observed in *S. cerevisiae* (Chern et al. 2002). An interesting structural feature of Rmt2 are its N-terminal ankyrin-repeats, a module that may be employed for the recognition of methylated lysine (Collins et al. 2008). This may suggest a potential interplay between

lysine and arginine methylation. Rmt2 appears to localize to speckles in both the cytoplasm and nucleus and co-purifies with nuclear pore proteins (Olsson et al. 2007).

Even though SPOUT family proteins generally methylate RNA, one member, Sfm1, has been reported to be a PRMT in *S. cerevisiae* (Tkaczuk et al. 2007; Young et al. 2012). Sfm1 appears not to be conserved in metazoa (Young et al. 2012).

No PRMT3 homologue has been described in *S. cerevisiae,* despite this protein`s conservation in *S. pombe* (Bachand 2007). It is interesting in this light that Hmt1 methylates Rps3, while human RPS3 is methylated by PRMT3. This suggests that Hmt1 took over the function of a PRMT3 homologue in *S. cerevisiae.*

### 1.5. Reversibility and Regulation of Arginine Methylation

Reversibility is an important feature for many PTMs to exert dynamic control of signalling. Unlike for lysine (Tsukada et al. 2006), no arginine demethylase has been proposed in *S. cerevisiae* to date. However, the nuclear import of Npl3 requires its phosphorylation by Sky1, which is impeded by Npl3 methylation (Yun & Fu 2000). The fact that Npl3 is exported in the methylated state may suggest demethylation in the cytoplasm (Shen et al. 1998). An arginine demethylase activity of the human JMJD6 enzyme has been described (Chang et al. 2007, p.6) and this finding was supported by some (Poulard et al. 2014; Gao et al. 2015) and challenged by other subsequent reports (Webby et al. 2009, p.6; Boeckel et al. 2011; Böttger et al. 2015). Additionally, arginine demethylation by two JmjC-domain containing proteins has been observed in *A. thaliana* (Cho et al. 2012). It should however be noted that the hydroxylation of lysines by JmjC-domain proteins (Webby et al. 2009) may interfere with antibody-based assays recognizing neighbouring methylated arginines. Unmethylated and methylated arginines can also be enzymatically deiminated to citrulline, however this also only has been described in higher eukaryotes (Cuthbert et al. 2004).

The regulation of methylation reactions is largely unexplored to date. In the case of Hrp1, arginine methylation is inhibited by the interaction of the protein with RNA (Valentini et al. 1999). A similar inhibition may exist for Gar1 as *in vitro* methylation of this protein occurred at a higher extent in cell extracts treated with RNase than in untreated samples (Frankel & Clarke 1999). The methylation of Npl3 is repressed by the binding of the Air1 protein to Hmt1 (Inoue et al. 2000). In contrast, the binding of certain human PRMTs to interacting proteins enhances their activity or alters their substrate preference (Pak et al. 2011; Lacroix et al. 2008). While the human PRMT4 (CARM1) is inhibited through phosphorylation reactions which either inhibit its dimerisation or SAM-binding , regulation of *S. cerevisiae* PRMTs by PTMs has not been reported to date (Feng et al. 2009, p.1; Higashimoto et al. 2007).

### 1.6. Aim of this Work

While the regulation and function of several methylation sites in higher eukaryotes has been elucidated over the last twenty years, the number of sites with known function is low compared to the several hundred sites identified in large-scale proteomics studies (Boisvert et al. 2003; Uhlmann et al. 2012; Bremang et al. 2013; Guo et al. 2014). An increased knowledge of methylation sites in a lower model organism that lends itself more readily to experimental manipulation may therefore be helpful in fostering our understanding of the biology of arginine methylation.

In contrast to the human system, the number of known methylated non-histone proteins in *S. cerevisiae* remains low and limited efforts have been made to detect methylation sites on a global scale. In particular, the methylation sites known in *S. cerevisiae* largely represent one aspect of arginine methylation, i.e. methylation of RNA-binding proteins, in RG-rich regions and by PRMT Hmt1. This may constitute an inherent property of the yeast system, however it might also be attributed to limited efforts towards the identification of further sites.

Here, several approaches aimed at the detection of methylation sites in *S. cerevisiae* are presented. Both proteome-wide strategies (chapter 2) and experiments on the level of individual proteins (chapter 3) were employed. Subsequently, the MTase activities responsible for the deposition of some of the newly identified methylation sites were explored (chapter 4). Finally, steps towards the understanding of the biological consequences of arginine methylation on a selected protein were taken (chapter 5).

# 2. Identification of Protein Methylation by Proteomics Approaches

## 2.1. Introduction

In the following section, some of the early efforts for the detection of arginine methylation sites are described and the disadvantages of the classical methods used are outlined. Subsequently, the challenges associated with the identification of methylation sites by liquid chromatography tandem mass spectrometry (LC-MS/MS) and a strategy for overcoming one of these challenges are presented. Finally, LC-MS/MS based studies that have been performed in higher eukaryotes for the detection of methylation sites are summarized.

### 2.1.1. Methods for the Identification of Protein Arginine Methylation

Methylated arginine was originally identified after the incubation of isolated nuclei with radioactively labelled SAM or methionine and amino acid analysis (Paik & Kim 1967; Paik & Kim 1968). Two sources of this methylated arginine were identified as histones and myelin basic protein (Paik & Kim 1970; Baldwin & Carnegie 1971). Subsequently, asymmetric di-methylation of arginines in hnRNP proteins was also initially observed by amino acid analysis (Beyer et al. 1977). Liu (1995) determined the identities of the methylated proteins by radioactive *in vivo* and *in vitro* methylation assays on hnRNP complexes immunopurified from HeLa cells and subsequent 2D gel electrophoresis.

In yeast, an *in vivo* methylation assay based on the transfer of methyl-groups from radioactive SAM was used to demonstrate the *in vivo* methylation of several proteins and their identity was inferred as Gar1, Nop1 and Nsr1 based on their migration in a SDS gel (Xu et al. 2003).

The disadvantage of *in vivo* radioactive methylation assays for the identification of methylated proteins is that further effort is required to determine the identity of a protein yielding signal. Once a methylated protein is purified, amino acid analysis can be used to determine the amino acid identity (e.g. lysine or arginine) and type (e.g. mono-, symmetric or asymmetric arginine di-methylation) of methylation, however this does not provide the

position of the methylated residue. Also, $^{14}C$ and $^{3}H$ are relatively weakly radioactive, so that long exposure times on the order of weeks are required. The sensitivity of this assay depends on the abundance of a protein in the cell, its number of methylation sites and the methylation stoichiometry of each of the sites. For example, Xu et al (2003) noted that the known arginine methylated protein Hrp1 could only be detected in their assay once its expression level was increased.

The methylation of the yeast protein Npl3 was first detected by an *in vitro* assay in which purified Npl3 was incubated with PRMT Hmt1 and radioactively labelled SAM (Henry & Silver 1996). However, this approach is limited to situations in which candidates for a PRMT and target protein are known. In the above case, Npl3 was chosen due to its structural and functional similarities to mammalian hnRNPs and Hmt1 had been identified in a genetic interaction screen with Npl3. While the *in vitro* generally yields higher signal than the *in vivo* methylation assay, it also does not provide the position of methylation sites. Additionally, as individual proteins are tested, its throughput is limited.

The use of pan methyl antibodies distinguishes different types of methylation and the detection of methylated proteins by western blotting is faster than by autoradiography. However, the task of determining the protein identity of an observed band and the localisation of the methylation site remains. Additionally, antibodies may not bind to methylation sites completely independently of the sequence context.

### 2.1.2. Challenges in Methylation Assignment

In contrast to the methods presented in the previous section, LC-MS/MS allows the detection of methylation sites, including their positions and simultaneous identification of the protein identities. The identification can be done in high-throughput on a proteome-wide scale. The numbers of methyl-groups present on a peptide can be readily derived from its mass-to-charge (m/z) ratio. Additionally, cases have been reported in which spectral information was

sufficient to distinguish symmetric and asymmetric arginine di-methylation (Brame et al. 2004), though this may only be possible in certain instances.

However, several challenges are associated with the global discovery of PTMs in general, and methylation in particular, by LC-MS/MS. First, compared to most other PTMs, the changes in the physico-chemical properties invoked by methylation are more subtle. While there is a change in the ion pairing abilities of methylated arginine its overall charge state remains positive in physiologically relevant solutions. Further, methylation is among the smallest PTMs, i.e. interaction surfaces are only changed to a limited extent. For these reasons, it has proven difficult to develop affinity reagents that distinguish proteins in the methylated from the unmethylated form without a bias to their sequence context. This in turn limits the levels of enrichment that can be achieved for methylated proteins.

As a general challenge in the discovery of PTMs by tandem mass spectrometry it should be noted that the false discovery rate (FDR) on the subgroup of modified peptides may be different to that of unmodified peptides for several reasons: The modification might lead to altered fragmentation characteristics, a different number of modified and unmodified peptides in the sample causes different prior probabilities of identification and the search space of modified and unmodified peptides may be different (Fu & Qian 2014). The discrepancy between the number of modified peptides in the search space and in the sample is especially severe when the abundance of modified peptides in the sample is low. Therefore, to avoid underestimating the number of false positives, throughout this study FDRs are calculated for methylated peptides separately from the global FDR and in cases when the numbers of identified methylated peptides are too low for the reliable calculation of an FDR, results are filtered on database search engine scores.

Furthermore the arginine methylated peptides produced by tryptic digestion may have characteristics that are sub-optimal for analysis by LC-MS/MS: While the effect of

methylation on the efficiency of tryptic cleavage has not been systematically investigated, it is generally agreed that missed cleavage sites are often caused by di- and to a lesser extent by mono-methylation (Yagoub et al. 2015; Gehrig et al. 2004), resulting in unusually long peptides. On the other hand very short peptides often originate from glycine-arginine-rich regions in which arginine methylation sites are found frequently.

Additionally, methylation may potentially occur on several different amino acids, complicating the site localisation (see "3.1.4. Methylation on residues other than lysine and arginine"). Moreover, the formation of methyl-esters is an artefact that can arise during sample preparation from the reaction with methanol (Jung et al. 2008). Therefore, methanol was avoided in the sample preparation for LC-MS/MS experiments throughout this study.

Finally, methylation is a PTM that poses a particular challenge to discovery by tandem mass spectrometry as the mass difference introduced by the addition of one, two or three $CH_2$ (replacement of H by $CH_3$) groups (14.02, 28.03 and 42.05 Da respectively) is isobaric to several amino acid substitutions (Ong et al. 2004). Therefore, a spectral database search engine may, for example, mistakenly place a methyl-group on a peptide and compensate for the mass shift by assigning the amino acid of the lower mass instead of the actually occurring 14.02 Da heavier partner at a nearby location. This would cause a false positive identification. The occurrence of false negatives can be explained by an analogous scenario. In the case of the Lys-Arg and Asn-Lys pairs, the mass difference between the partners differs in the second digit from the mass of a di- and mono-methylation respectively, so that false assignments can be avoided by the use of a highly accurate mass analyser (Fig 2.1 A). On the other hand, the mass difference in the Asp-Glu, Asn-Gln, Ser-Thr pairs and several pairs of aliphatic amino acids originates from a difference in the number of $CH_2$-groups and is therefore exactly the same as the one introduced by a methylation event (Fig 2.1 B).

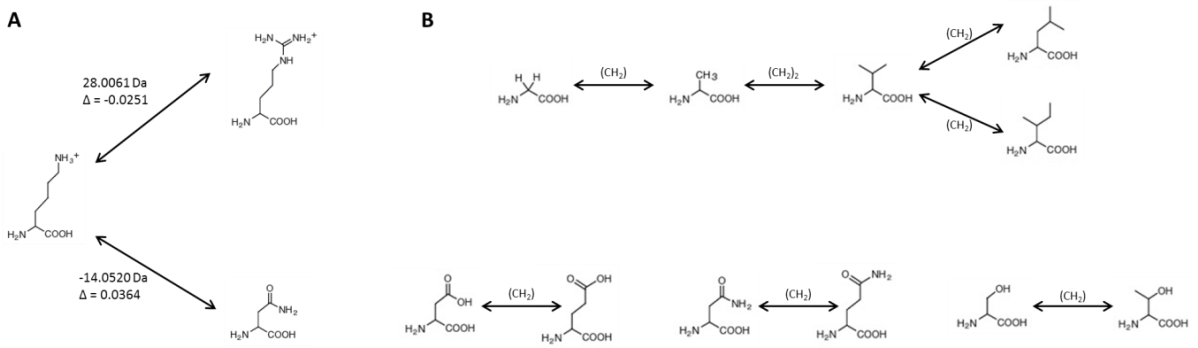



***Figure 2.1: Amino acid pairs with a mass difference isobaric to one or more methyl-groups.*** *A, The mass difference between lysine - arginine and lysine - aspartate residues differs in the second digit (Δ) from the mass of two and one methyl-groups respectively. B, The difference between several pairs of aliphatic amino acids, aspartate - glutamate, asparagine - glutamine, and between serine - threonine residues equals the mass of methyl-groups.*

An advantage in the analysis of lysine and arginine methylations is their relatively high chemical stability. Also no arginine demethylases are known in *S. cerevisiae* to date that may remove methyl-groups from proteins during sample preparation.

### 2.1.3. MethylSILAC

A strategy to counteract elevated false positive and false negative rates due to the properties of the methyl-group has been presented (Ong et al. 2004). Here cells are grown in both normal medium and medium in which methionine is replaced by a version with a stable isotope labelled methyl-group. The latter leads to the metabolic synthesis of heavy labelled SAM, the main, or sole cellular methyl-donor and therefore a transfer of heavy methyl-groups in methylation reactions. Upon mixing of the two cultures and bottom up MS analysis a characteristic pair of MS peaks can be observed due to the presence of one partner with a light and one with a heavy methyl-group. The presence or absence of this MS1 pair will improve the confidence in the presence or absence respectively of the methylation of a detected peptide in this heavy methyl SILAC (stable isotope labelling with amino acids in cell culture) strategy (in the following referred to as 'MethylSILAC'). It should be noted that a heavy-light pair can also be formed due to a methionine in the peptide sequence.

### 2.1.4. Proteome Wide Discovery Studies of Protein Arginine Methylation

While early identifications of protein arginine methylation sites were based on radioactive labelling and antibody-based detection on a protein-by-protein basis (see "2.1.1. Methods for the Identification of Protein Arginine Methylation") several proteome-wide studies based on LC-MS/MS appeared over the last years. The most important ones are summarized here (Table 2.1).

**Table 2.1:** Studies for the large scale discovery of methylation sites by LC-MS/MS in mammals.

| | MethylSILAC | Samples | Technique | # Methylation Sites | # Methylated Proteins |
|---|---|---|---|---|---|
| Boisvert, 2003 | n | HeLa | Protein IP | NA (*1) | NA (*1) |
| Ong, 2004 | y | HeLa | Protein IP | 59 | 31 |
| Uhlmann, 2012 | y | Jurkat; primary T-cells | Protein IP, SCX, IEF, HILIC | 249 | 131 |
| Bremang, 2013 | y | HeLa | Protein IP, SDS PAGE, IEF | 397 | 139 |
| Liu, 2013 | n | HEK293T | HP1β CD pull-down | 40 | 29 |
| Cao, 2013 | n (*2) | HeLa | Peptide IP | 552 | 413 |
| Moore, 2013 | y | 293T | 3xMBT domain pull-down | 26 | 24 |
| Guo, 2014 | n | HCT116; mouse tissues | Peptide IP | >1000 | >900 |
| Sylvestersen, 2014 | n | HEK293T | Peptide IP | 1027 | 494 |
| Geohegan, 2015 | y | Jurkat; primary T-cells | Peptide IP | 2502 | 1257 |

(*1) Only proteins enriched by IP are reported.
(*2) Light and heavy samples were analysed individually; heavy-light pairs are not a requirement for reporting sites.
CD: chromodomain; 3xMBT: three malignant brain tumor domains; IEF: Iso-electric focussing; HILIC: Hydrophilic interaction liquid chromatography; SCX: Strong cation exchange.

The first successful attempt at the identification of protein methylation by LC-MS/MS was reported in 2003 and was based on the enrichment of proteins from HeLa cells using antibodies for asymmetric and symmetric di-methyl arginine (Boisvert et al. 2003). While more than 200 proteins were enriched in this approach and several of them contained RG-rich regions, methylation sites were not mapped to the proteins. Subsequently, the MethylSILAC method for the confident mapping of methylation sites was developed, alongside which the identification of 59 methylation sites on 31 proteins from HeLa cells using enrichment with anti-methyl-lysine and anti-methyl-arginine antibodies was reported (Ong et al. 2004). A study which did not only make use of protein immunoprecipitation (IP), but also of sample fractionation without prior enrichment was published in 2012 (Uhlmann et al. 2012). A total of 249 arginine methylation sites on 131 proteins were found in Jurkat and primary T-cells and certain fractionation strategies outperformed the antibody-based enrichment. A further

attempt at the identification of methylated proteins via protein IP compared five anti-methyl lysine and six anti-methyl arginine antibodies, the most successful of which yielded 129 methylation sites from a nuclear fraction of HeLa cells (Bremang et al. 2013). In comparison, SDS PAGE and off-gel fractionation of nuclear extract without further enrichment enabled the detection of 210 and 87 sites respectively. In total, 397 sites on 138 proteins were detected.

Two studies reported the use of methyl-binding domains for the enrichment of lysine methylated proteins. The chromodomain of HP1β was employed to identify 40 methylation sites on 29 proteins from HEK293T cells (Liu et al. 2013). Subsequently, 26 sites on 24 proteins from 293T cells could be identified via a construct of three malignant brain tumour domain repeats (Moore et al. 2013).

The use of affinity purification on peptide level for the enrichment of lysine methylated peptides yielded 552 sites from 413 proteins from HeLa cells (Cao et al. 2013). Novel antibodies for the enrichment of arginine methylation sites on peptide level were introduced by Guo and lead to the identification of more than 1000 methylation sites on over 900 proteins in HCT116 cells and mouse tissues. The majority of these sites were mono-methylations (Guo et al. 2014). The same antibodies subsequently allowed the identification of 1027 arginine mono-methylation sites on 484 HEK293 proteins (Sylvestersen et al. 2014) and 2502 arginine mono-methylation sites on 1257 proteins of Jurkat and primary T-cells (Geoghegan et al. 2015).

Four out of the seven studies made use of the MethylSILAC strategy after its initial report (Table 2.1).

At the beginning of this study no proteome-wide identification experiments of protein methylation in *S. cerevisiae* were available. In due course, several such studies were reported

and are summarized in “2.9.9. Proteome-wide Discovery Studies of Protein Methylation in *S. cerevisiae*”.

## 2.2. Methods

### SDS PAGE and gel staining

Proteins (about 20 µg unless stated otherwise) were loaded onto Bis-Tris gels (4-12%: BioRad, #345-0124; 12%: 3450118) in SDS sample buffer (50 mM Tris(hydroxymethyl)aminomethane (Tris), 2% (w/v) SDS, 10% (v/v) glycerol, 0.02% (w/v) bromphenol blue, 1.25% (v/v) 2-mercaptoethanol, pH 6.8) and separated by electrophoresis in MOPS buffer (50 mM MOPS, 50 mM Tris, 0.1% (w/v) SDS, 1 mM EDTA, pH 7.7) at 100 to 170 V. Coomassie staining was performed with InstantBlue (Expedeon, ISB1L) according to the manufacturer`s instructions.

For silver staining, gels were incubated in 40% (v/v) ethanol and 10% (v/v) acetic acid for 1h, twice each in 30% (v/v) ethanol and water for 20 min, in 0.02% (w/v) $Na_2S_2O_3$ for 1 min and twice in water for 20 s. They were stained in the dark at 4 °C in 0.1% (w/v) $AgNO_3$ for 20 min, washed in water three times for 20 s and for an additional minute after transfer to a fresh tray. The gels were developed with 0.05% (v/v) formalin in 3% (w/v) $Na_2CO_3$ until sufficient staining was achieved.

### In-gel digestion

To perform in-gel digests, the regions of interest were cut from a gel and divided into pieces of about 1 $mm^3$ and transferred to a 1.5 ml tube. They were then washed in 200 µl 30% (v/v) ethanol overnight with an exchange of the solution after 2 h and then dehydrated with 200 µl acetonitrile for 5 min. After removal of acetonitrile, the gel pieces were incubated in 30 µl 10 mM dithiothreitol (DTT) for 30 min and 30 µl iodoacetamide for 30 min. They were dehydrated with 200 µl acetonitrile for 5 min, washed with 200 µl 100 mM ammonium bicarbonate for 10 min and dehydrated as above. They were then incubated with 30 µl of 2

ng/µl trypsin (Promega, V5111) in 50 mM ammonium bicarbonate for 10 min on ice, after which excess solution was removed, 5 µl 50 mM ammonium bicarbonate added and samples incubated at 37 °C overnight. If AspN (Promega, V1621), ArgC (Roche, 11370529001), chymotrypsin (Thermo, 90056), GluC (Promega, V1651)), elastase (Worthington, LS006365) or LysC (Thermo, 90051) were used instead of trypsin, digestions were carried out in the solutions and at the concentration and temperature recommended by the manufacturers. Peptides were extracted from the gel pieces by subsequent addition, 10 min of vortexing and collection of the supernatants of 50 µl each of 50 mM ammonium bicarbonate, 50% (v/v) acetonitrile with 5% (v/v) formic acid and 85% (v/v) acetonitrile with 5% (v/v) formic acid in water. The solutions were combined, dried in a vacuum centrifuge and pellets resuspended in 2% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water.

**Western blotting**

For western blot analysis SDS PAGE separated proteins were transferred to a PVDF membrane (Millipore, Immobilon-F) by wet blotting at 20 V overnight. The membrane was incubated with blocking buffer (SEA Block, 10782925) for 1 h at room temperature before a 1 h incubation with rabbit antibodies Me-R4-100 or D5A12 (both CST: 13522 and 8015), ICP0801 (Immunechem: ICP0801), SYM10 or ASYM24 (both Millipore: 07-412 and 07-414) and mouse anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody GA1R (Pierce, MA5-15738) each at a 1 : 1,000 dilution. The membrane was washed four times for 5 min in phosphate buffered saline (PBS) + 0.1% Tween-20 (PBST), incubated for 1h with fluorescently labelled anti-rabbit and anti-mouse secondary antibodies (Thermo, 10175663 and 10751195) at a 1 : 15,000 dilution, washed as above, rinsed in PBS and imaged on an Odyssey detection system (Li-cor).

**Yeast culture and MethylSILAC**

The following media were used for the growth of yeast cultures (Anon 1994): -Minimal Synthetic Defined (SD) medium: Yeast nitrogen base without amino acids (YNB-AA: Formedium, CYN0405: 5 g/l ammonium sulphate, 2 µg/l biotin, 400 µg/l Ca-panthotenate, 2 µg/l folic acid, 2000 µg/l inositol, 400 µg/l nicotinic acid, 200 µg/l p-aminobenzoic acid, 400 µg/l pyridoxine HCl, 200 µg/l riboflavin, 400 µg/l thiamine HCl, 500 µg/l boric acid, 40 µg/l copper sulphate, 100 µg/l potassium iodine, 200 µg/l ferric chloride, 400 µg/l manganese sulphate, 200 µg/l sodium molybdate, 400 µg/l zinc sulphate, 1g/l potassium phosphate, 0.5 g/l magnesium sulphate, 0.1 g/l sodium chloride, 0.1 g/l calcium chloride) with 20 g/l glucose.

-Synthetic Complete (SC) medium: YNB-AA with synthetic complete dropout mix (SC, Kaiser mixture: Formedium, DSCK1000: 18 mg/l adenine, 76 mg/l each of L-alanine, L-arginine HCl, L-asparagine, aspartic acid, L-cysteine, L-glutamine, L-glutamic acid, glycine, L-histidine, myo-inositol, L-isoleucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, uracil, L-valine, 380 mg/l L-leucine, 8 mg/l para-aminobenzoic acid) with 20 g/l glucose.

-Yeast Peptone Dextrose (YPD) medium: 10 g/l yeast extract (BD, 212750), 20 g/l peptone (BD, 211677), 20 g/l glucose.

Glucose was substituted with 20 g/l raffinose or 20 g/l galactose where indicated.

For MethylSILAC experiments *S. cerevisiae* BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) cells were inoculated into 5 ml SC medium of and shaken at 180 rpm and 30 $^{0}$C overnight. They were diluted into 600 ml of the same or the corresponding medium in which methionine was replaced with methionine-[methyl $^{13}CD_3$] (Sigma, 299154) to an $OD_{600}$ of 0.003. After growth to $OD_{600}$ of about 0.5 heavy and light cultures were harvested by centrifugation (5 min at 3,000 rpm).

**Jurkat cell culture**

To create a positive control sample for peptide immunoffinity enrichment, Jurkat cells, clone E6-1 (ATCC® TIB-152™) were grown in RPMI-1640 (AthenaES, 0422) with 10% (v/v) dialyzed FBS and either light methionine or where methionine had been substituted with methionine-[methyl $^{13}CD_3$] (Sigma, 299154). Cells were grown in tissue culture flasks for three passages and shaking at 80 rpm for an additional five passages.

**Protein immunoaffinity enrichment**

The methods for the purification of methylated proteins from nuclear extracts by IP were adapted from a previous study on HeLa S3 lysates (Bremang et al. 2013):

*S. cerevisiae* cells were grown with light and heavy methionine as described above. Unlabelled and labelled HeLa S3 cells were a kind gift of the Bonaldi lab. Yeast cells corresponding to 300 ml culture were washed in water and resuspended in 200 ml 0.1 mM Tris, pH 9.4 with 10 mM DTT and shaken at 60 rpm and 30 °C for 15 min. They were spun down and washed once in 1.2 M sorbitol in 20 mM HEPES pH 7.4, before incubation in the same buffer with 2.75 mg zymolase 20T per 1 g yeast (60 rpm, 30°C, 1 h). Spheroplasts were washed in 200 ml ice-cold 1.2 M sorbitol, 20 mM PIPES, 1 mM $MgCl_2$ (pH 6.8).

Yeast spheroplasts and HeLa S3 cells were resuspended in 4 ml cold lysis buffer (10 mM KCl, 1.5 mM $MgCl_2$, 10 mM HEPES-KOH pH 7.9, 0.2% NP40 with protease inhibitors). They were incubated for 10 min on ice, then lysed by 25 strokes in a Dounce homogenizer on ice. The samples were spun (3,500 g, 25 min 4 $^0$C) and the pellets (nuclear fraction) were washed with 1 ml PBS before resuspension in high-salt buffer (450 mM KCl, 2 mM $MgCl_2$, 20 mM HEPES-KOH pH 7.9, 20% glycerol, 0.2 mM EDTA, protease inhibitor tablets) and rotation at 4 $^0$C for 1 h. Following ultra-centrifugation for 1 h at 60,000 g and 4 $^0$C the supernatants were collected and samples adjusted to 150 mM NaCl, 2 mM $MgCl_2$, 20%

glycerol, 20 mM HEPES-KOH pH 7.9, 1 mM EDTA, 0.25% sodium deoxycholate, 1 mM DTT, 0.5 mM PMSF (IP buffer) and proteins were quantified by bicinchoninic acid (BCA) assay (Pierce, 23227F)(Smith et al. 1985) and visualized by SDS PAGE. Lysate of cells from light and heavy cultures were mixed in a 1:1 protein ratio. The samples were snap frozen in liquid nitrogen and stored at -80 $^{0}$C.

The lysates were centrifuged for 10 min at 16,000 g, pre-cleared twice by incubation for 1 h at 4 $^{0}$C with 40 µl of protein-G magnetic beads (Dynabeads, Invitrogen 100.04D) and incubated overnight at 4 $^{0}$C with 50 µg of anti-methylated lysine antibody (Acris, AP09328PU-N). They were mixed with 40 µl protein-G beads that had been pre-equilibrated with IP buffer and rotated for 3 h at 4 $^{0}$C. The beads were washed four times with ice-cold PBS before heating at 75 $^{0}$C for 5 min in SDS sample buffer, separation by SDS-PAGE and Coomassie or silver staining.

The protocol was performed on three samples. For one sample all solutions from spheroplasting to IP contained 1 mM of demethylase inhibitors 5-Carboxy-8-hydroxyquinoline (IOX1) and tranylcypromine (TCP) (Hopkinson et al. 2013; Lee et al. 2006). For a second sample, the antibody was omitted in the pull-down to serve as a negative control.

The proteins were in-gel digested and analysed by mass spectrometry using a QExactive mass spectrometer.

**3xMBT-domain based pull-down experiments**

Pull-down with the three MBT domain repeats of L3MBTL1 (3xMBT) was performed essentially as described (Moore et al. 2013). The GST-tagged 3xMBT construct in a pGEX6P1 plasmid and the corresponding D355N mutated version were a kind gift of the Gozani lab (Stanford University). The plasmids were transformed into OneShot TOP10 *E. coli* cells (Invitrogen, C4040-03) according to manufacturer`s instructions and plated on

lysogeny broth (LB) agar plates containing 100 µg/ml ampicillin (Bertani 1951). Cells were cultured and plasmids extracted according to the instructions of the Plasmid Maxi Kit (Qiagen, 12162). The plasmids were transformed into *E. coli* BL21 cells. Additionally, empty pGEX6P1 was transformed as a negative control. Bacteria from an overnight culture were diluted into 500 ml LB containing 100 µg/ml ampicillin and grown at 37 $^{0}$C for 2.5 h. Protein expression was induced by adding IPTG to a final concentration of 0.1 mM and cultures continued at 20 $^{0}$C overnight. Cells were harvested and resuspended in 10 ml lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.05% (v/v) NP40) supplemented with protease inhibitor tablets, PMSF at 0.5 mM and 0.25 mg/ml lysozyme. The samples were incubated on ice for 30 minutes and sonicated four times for 10 seconds using a micro-tip sonicator at 22% output. Cell debris was removed by spinning at 4° C for 15 minutes at 30,000 g.

Glutathione-sepharose 4B beads (GE, 17-0756-01) were washed three times with 10 ml lysis buffer and the cleared bacterial lysate was bound to 300 µl beads by rotation overnight at 4°C. The domain-bound beads were again washed three times with 10 ml lysis buffer.

Pellets of about 1 g *S. cerevisiae* BY4741 cells were resuspended in 150 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.05% (v/v) NP40, supplemented with protease inhibitors. Cells were lysed by four rounds of bead beating (30 s, 6,000 rpm) in a Precellys Dual bead beater and cellular debris removed by 10 min of centrifugation at 20,000 g. Lysates from light and heavy cultures were mixed in a 1:1 protein ratio. The pull-down was performed by mixing 10 mg protein from crude extracts with 50 µl bait-saturated glutathione sepharose 4B beads and rotation overnight at 4 °C. After washing three times in lysis buffer, proteins were eluted with 20 µl 10 mg/ml L-glutathione in 50 mM Tris pH 8.0 during a 4 h rotation at 4 °C. To remove the bait-protein, the eluate of one sample was dialyzed overnight at 4°C against a 10,000x volume of dissociation buffer (50 mM Tris pH 7.5, 750 mM NaCl) in a 0.1 ml Slide-A-Lyzer MINI Dialysis Device with 3.5 kDa molecular weight cut-off (Thermo Scientific, 69550).

The dialysed sample was incubated with GSH-beads and washed in dialysis buffer for 4 h at 4 °C. After centrifugation the supernatant was collected.

The eluates were separated on a SDS PAGE gel and each lane cut into pieces of about 1 cm. A separate piece was cut for the band originating from GST-3xMBT. The gel pieces were subjected to in-gel digestion with AspN, ArgC, chymotrypsin, GluC, elastase, LysC and trypsin according to manufacturers' instructions. The peptides from gel slices without major bands were combined into two pools before analysis using a QExactive mass spectrometer. Application of 600 µg nuclear extract from HEK293T to the pull-down was used as a positive control as this has been described to yield methylated proteins (Moore et al. 2013).

**HILIC fractionation**

Yeast cell pellets were resuspended in 10 ml lysis buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.2% (v/v) NP-40, 5 mM EDTA, 1 mM PMSF) and lysed by bead beating six times for 30 s at 6,000 rpm with 30 s on ice in between. Lysates were cleared by centrifugation and supernatants of cells from light and heavy cultures were mixed in a 1:1 labelled : unlabelled protein ratio.

SDS was added to 1% and DTT to 5 mM and samples were heated 5 min at 50 $^{0}$C while shaking at 350 rpm. The solution was allowed to cool to room temperature, before adding iodoacetamide to 30 mM and incubation in the dark at RT at 350 rpm for 40 min. Three volumes of an ice cold mix of acetone and ethanol (1:1) were added and samples centrifuged for 30 min at 4 $^{0}$C and 15,000 g. The supernatant was removed and samples were resuspended in 1 ml 8 M urea in 100 mM Tris pH 7.8. The protein concentration was estimated by BCA assay and volumes corresponding to 500 µg of light and heavy sample were mixed. The concentration of urea was reduced to 1 M by diluting in 50 mM Tris-HCl (pH 8) and digestion with 20 µg trypsin was carried out at 37 $^{0}$C overnight. The digests were acidified with trifluoroacetic acid (TFA) to a final concentration of 0.2% and peptides purified by solid

phase extraction on a SepPak C18 column. The eluates were dried in a vacuum centrifuge and the residual resuspended in 105 µl 95% (v/v) acetonitrile, 0.1% aqueous formic acid (FA) and cleared by centrifugation (5min at 5,000 rpm).

The supernatant (100 µl; about 1 mg peptides) was subjected to HPLC separation on a hydrophilic interaction liquid chromatography column (HILIC; SeQuant ZIC-HILIC, 3.5 µm, 200 Å, 15 x 4.6 mm; Merck, 1.50449.0001) with mobile phases of 0.1% (v/v) formic acid in water (A) and 95% (v/v) acetonitrile and 0.1% (v/v) formic acid in water (B) at a flow rate of 0.5 ml/min. The column was flushed at 90% B for 22.4 min, before ramping solvent B down to 40% over 50 min, remaining at 40% B for 12.5 min and equilibrating at 90% B for 30 min. The absorption at a wavelength of 214 nm was recorded and eluting fractions were collected at 1 min intervals from a run time of 22 to 90 min.

Each fraction was dried in a vacuum centrifuge, resuspended in 20 µl 2% (v/v) acetonitrile, 0.1% (v/v) formic acid in water and subjected analysis using a QExactive mass spectrometer.

**Peptide immunoaffinity enrichment**

Yeast cells were lysed by ten rounds of bead-beating for 30 s at 6,000 rpm in 20 mM Tris buffer, pH 8 and 8 M urea and debris was removed by centrifugation at 20,000 g for 15 min. The protein concentration was determined by BCA assay (Pierce, 23227F) and heavy and light samples mixed in a 1:1 protein ratio.

Jurkat cells were harvested by centrifugation and lysis performed by resuspending five pellets of about 1 x $10^8$ cells in 1 ml 8 M urea in 20 mM Tris buffer pH 8, shaking at room temperature for 15 minutes and sonication with a microtip with 3 bursts of 30 seconds at 20% output and 1 minute cooling on ice in between. The resulting lysate was cleared by centrifugation.

Cleared lysates corresponding to 25 mg protein per sample from *S. cerevisiae* or Jurkat cells were subjected to 30 min each of reduction (5 mM DTT), alkylation (15 mM iodoacetamide) and quenching (5 mM DTT) at room temperature. The urea concentration was reduced to 2 M by adding 20 mM Tris pH 8 and $CaCl_2$ was added to 1 mM. Digestion was carried out with trypsin (Promega, V5111) at a 1:100 mass-ratio and incubation at 37 $^0$C overnight. Samples were split into batches corresponding to 12.5 mg protein and TFA added to 0.1% (v/v). After an incubation of ten minutes at room temperature, insoluble material was removed by centrifugation (2,000 g, 5 min) and samples desalted by solid phase extraction (Sep-Pak C18 Plus, Waters) with stepwise elution at 10%, 20%, 30% and 40% (v/v) acetonitrile in 0.1% (v/v) aqueous TFA. The eluates were combined, frozen at -80 $^o$C and lyophilized. Aliquots of 80 µl protein A agarose bead slurry (Roche, 11719408001) were washed twice in immunoaffinity purification buffer (IAP; 50mM MOPS, 10mM $NaH_2PO_4$, 50 mM NaCl) and incubated with a mixture of 100 µl Me-R4-100 and 50 µl D5A12 anti-methyl antibodies (CST, 13522 and 8015), adjusted to a final volume to 500 µl with IAP buffer, then washed three times in IAP buffer. The peptide aliquots equivalent to 12.5 mg protein were resuspended in 1.3 ml of IAP buffer, cleared at 20,000 g for 5 min, added to the antibody-bound beads and rotated overnight at 4 $^o$C. The beads were washed three times in immunoaffinity purification buffer, once in water and peptides eluted in 100 µl 0.15% TFA. Peptides were purified by passing through a stage tip with a glass microfiber filter (Whatman, 1820-042) followed by desalting on a C18 filter (Supelco, 66883-U). The latter was washed with 0.1% TFA and peptides eluted in 50 µl 50% (v/v) acetonitrile, 0.1% (v/v) TFA in water. Samples corresponding to 25 mg protein starting material were combined, dried in a vacuum centrifuge, resuspended in 20 µl 2% (v/v) acetonitrile, 0.1% (v/v) formic acid in water and subjected to LC-MS/MS analysis using a QExactive and Orbitrap Elite mass spectrometer. The experiment was performed in triplicate for *S. cerevisiae* samples.

The immunoaffinity purification using a mix of both antibodies had been previously employed successfully on human cells (Sylvestersen et al. 2014). As RGG-motif containing peptides occurred with high frequency here, the experiment was repeated with 100 µl Me-R4-100 alone.

**Mass spectrometric methods**

QExactive

Peptides were separated by nano-UPLC (U3000; Thermo) using a 50 cm C18 column (nanoEASY; 75 µm I.D.) at 4 $^{o}$C and a flow rate of 250 nl/min and mobile phases of 0.1% (v/v) formic acid and 5% (v/v) dimethyl-sulfoxide (DMSO) in water (A) and 0.1% (v/v) formic acid and 5% (v/v) DMSO in acetonitrile (B). The percentage of solvent B was ramped from 2% to 5% within 5 min and up to 35% within 60 min. The analyte was electrosprayed into a QExactive hybrid mass spectrometer in positive ionisation mode and analysed with data-dependent acquisition, selecting the 15 most intense precursor peaks with a charge greater than 1 for fragmentation at a normalized collision energy of 28%. Precursor scans were performed at a resolution of 70,000 with an ion target of $3x\ 10^{6}$ and maximum accumulation time of 100 ms (scan range of 300 to 1,500 Th) and fragment scans at a resolution of 17,500 with an ion target of $1x\ 10^{5}$ and a maximum accumulation time of 128 ms (scan range of 200 to 2,000 Th). Precursor ions above an intensity threshold of $4.7x\ 10^{4}$ were selected for higher-energy collisional dissociation (HCD) with an isolation window of 1.6 Th. Dynamic exclusion of 27 s was enabled.

Orbitrap Elite

To further obtain spectra with high sequence coverage for the peptide immunoaffinity enrichment experiment, peptides were separated by nano-UPLC (Waters, NanoAcquity) on a 25 cm C18 column (75 µm I.D.) at 40 $^{o}$C and a flow rate of 250 nl/min and mobile phases of 0.1% (v/v) formic acid in water (A) and 0.1% (v/v) formic acid in acetonitrile (B). The

percentage of solvent B was ramped from 3 to 40% (v/v) over 2 h. The analyte was electrosprayed into an Orbitrap Elite hybrid mass spectrometer in positive ionisation mode and analysed with data-dependent acquisition, performing one collision-induced dissociation (CID) and one electron-transfer dissociation (ETD) based fragment scan on each of the five most intense precursors. The precursor scan was acquired in the Orbitrap on a mass range from 300 to 1,500 Th at a resolution of 120,000. CID was performed at a normalized collision energy of 35% and ETD by reaction with fluranthene with a reaction time of 100 ms and supplemental activation enabled. Fragment scans were done in the ion trap with an isolation window of 1.8 Th and precursor threshold of 1,000. Fragment ions were accumulated to an ion target of 50,000 or a maximum of 100 ms. Dynamic exclusion after one repeat was enabled for 10 s and precursors of unassigned charge state or charge = 1 were rejected from fragmentation. Lock mass correction on m/z = 445.1200 (polysiloxane) was enabled.

**MS data analysis**

The raw files obtained from MS analyses were converted to the .mgf format using MSConvert (Chambers et al. 2012), keeping the 200 most intense peaks.

The MS data from all experiments were searched with Mascot (v2.5, in-house version; Matrix Science) against a UniProt/SwissProt database limited to entries from *S. cerevisiae* with common contaminants added. The appropriate protease (usually trypsin) was selected, allowing up to 3 missed cleavages. A 10 ppm precursor- and 0.07 Da or 0.5 Da fragment-tolerance for MS2 data acquired in the orbitrap and the ion trap respectively were chosen. For scoring, b-, y- (inclusive water and ammonium loss) and a- ions were allowed. For MS2 spectra recorded in the orbitrap scoring based on immonium ions was additionally enabled. Data were searched with a light and a heavy parameter set with the modifications allowed for each experiment shown in table 2.2.

**Table 2.2:** Modification parameters used in Mascot searches

| | | | Protein IP | 3xMBT pull-down | HILIC fractionation | Peptide IP |
|---|---|---|---|---|---|---|
| light parameters | fixed | Carbamidomethyl (C) | x | x | x | x |
| | variable | Oxidation (M) | x | x | x | x |
| | | Mono-methyl (K) | x | x | x | |
| | | Di-methyl (K) | x | x | x | |
| | | Tri-methyl (K) | x | | x | |
| | | Mono-methyl (R) | | | x | x |
| | | Di-methyl (R) | | | x | x |
| heavy parameters | fixed | Carbamidomethyl (C) | x | x | x | x |
| | | Label:13CD3 (M) | x | x | x | x |
| | variable | Oxidation (M) | x | x | x | x |
| | | $^{13}CD_3$ Mono-methyl (K) | x | x | x | |
| | | $^{13}C_2D_6$ Di-methyl (K) | x | x | x | |
| | | $^{13}C_3D_9$ Tri-methyl (K) | x | | x | |
| | | $^{13}CD_3$ Mono-methyl (R) | | | x | x |
| | | $^{13}C_2D_6$ Di-methyl (R) | | | x | x |

The results from the peptide immunoaffinity purification were first searched allowing only arginine mono-methylation in a first round and arginine mono- and di-methylation in a separate search. The results from the light and heavy parameter sets were re-formatted to be compatible with the novel software MethylQuant, merged into one file and analysed with MethylQuant_0_90_64bit. MethylQuant is a software for the identification of methylated peptides from LC-MS/MS analysis of MethylSILAC labelled samples. It was kindly provided by V. Geoghegan (University of Oxford): For each sequenced putative methylated peptide, MethylQuant calculates the theoretical m/z values for the first 3 isotopomers of the light and heavy MethylSILAC pair (6 isotopomers in total). Within a defined retention time window (here +/- 0.22min), the point at which the required number of MethylSILAC isotopomers (here: 5, max: 6) are at a maximum summed intensity is found (max. overlap). From this point, MethylQuant finds the start and end elution times for the MethylSILAC pair up to a maximum retention time window (here +/- 1 min from max. overlap). An averaged MS spectrum for the MethylSILAC pair is then generated between the start and stop times. The summed intensity of the first 3 heavy isotopomers is divided by the summed intensity of the first 3 light isotopomers, giving a heavy-to-light (H/L) ratio of the MethylSILAC pair.

Methylated peptides with an H/L ratio of >0.1 and <10 were accepted, except for the experiments in section 2.3.5, where the number of peptide spectrum matches (PSMs) to methylated peptides was considered sufficiently high for using a quantile-based threshold. In this case the 5$^{th}$ and 95$^{th}$ percentile of the H/L ratios (determined at E < 0.05) were set as cut-offs. The false discovery rate (FDR) was estimated by $FDR = \frac{2 \times \#\ decoy\ PSMs}{\#\ target\ PSMs + \#\ decoy\ PSMs}$. The results were filtered at a Mascot E-value corresponding to a FDR < 0.05 and the presence of ions arising from fragmentation directly next to the methylated residue (one if the methylated arginine was C-terminal) were further required. Additionally, if the localisation percentage of the Mascot site analysis for the first hit was less than three times the value of the following hit, the site assignment is reported as ambiguous.

**Bioinformatics analysis**

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al. 2009) was used to evaluate the enrichment of gene ontology terms among the proteins found to be arginine methylated. The *S. cerevisiae* proteome was selected as a background. Terms from the GO Fat database were considered for the description of the dataset.

To visualize the sequence context in which arginine methylations occur, a consensus sequence motif using a +/- 10 amino acid sequence window around the methylation site was generated against all *S. cerevisiae* SwissProt sequence entries as a background using iceLogo (Colaert et al. 2009).

The position of methylation sites on proteins was annotated relative to protein regions and domains annotated in UniProt and visualized using DomainDraw (Fink & Hamilton 2007). The image was overlaid with information on predicted secondary structures and unstructured regions obtained from the Yeast Resource Center (YRC; http://depts.washington.edu/yeastrc/) (Malmström et al. 2007).

## 2.3. Results

### 2.3.1. Global Methylation Levels Assessed by Western Blotting

To determine the level of complexity of the protein arginine methylation landscape in *S. cerevisiae* and evaluate antibodies available for subsequent enrichment approaches, wild type (BY4741) cells were initially profiled by performing immunoblotting with antibodies that specifically recognise peptides containing arginine mono-methylation in an RGG-context (D5A12), arginine mono-methylation in a degenerate sequence context (Me-R4-100, ICP0801) (Guo et al. 2014), asymmetric (ASYM24) (Côté et al. 2003, p.68) and symmetric (Boisvert et al. 2002) di-methylation (see also Table A2.1 in the appendix). Distinct bands could be recognized in crude BY4741 lysate with antibodies D5A12, ASYM24 and Me-R4-100, while no bands were visible when probing with ICP0801 (Fig. 2.2). In contrast the signal for SYM10 appeared as a continuous smear over most of the mass range. The protein methylation profile of *S. cerevisiae* lysate was compared to cellular extracts prepared from Jurkat cells, a human T lymphocyte cell line and calf histones were used as a control. The signal obtained with Jurkat cell lysate exceeds the one for yeast lysates for each antibody except SYM10, indicating a greater complexity of the protein methylation landscape in humans compared to *S. cerevisiae*.

It is not clear why distinct signals can be observed on Jurkat lysate as opposed to *S. cerevisiae* lysate when probing with SYM10. It should however be noted that both the SYM10 and ASYM24 signal in yeast is unaltered in different MTase deletion backgrounds and therefore may be attributed to non-specific binding (see “4.3.1. Global Differences in Methylation Levels in Methyltransferase Deletion Strains”). Only a single band, which according to its molecular weight may originate from a histone, was visible in Jurkat lysate for ICP0801. This may therefore indicate a generally low sensitivity or strong sequence bias of this antibody.

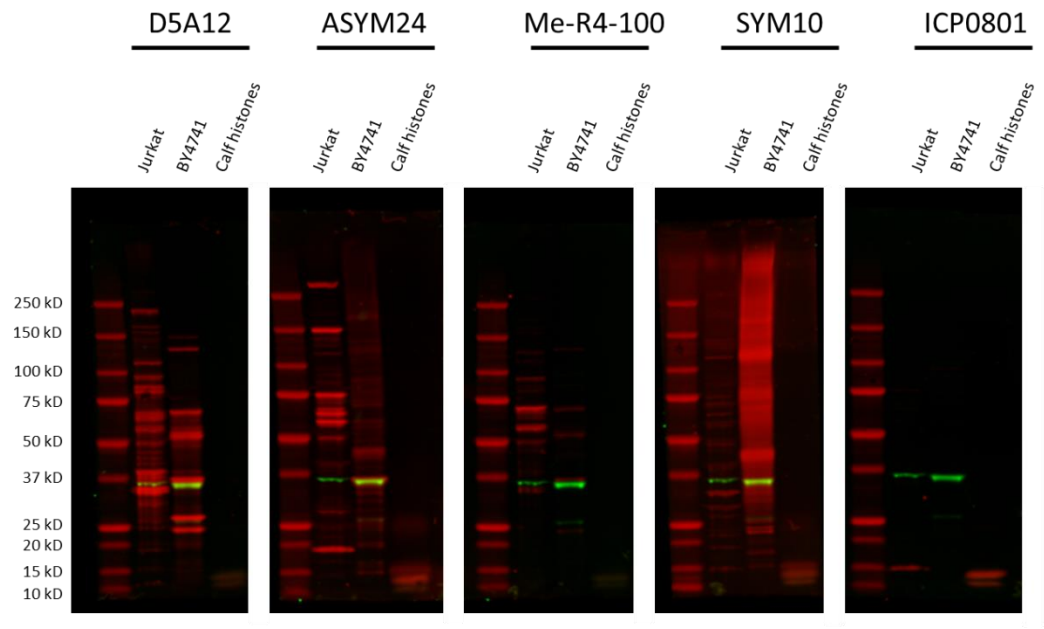



***Figure 2.2: Methylation landscape detectable by western blotting in Jurkat and S. cerevisiae BY4741 cells.*** *Cell extracts prepared from human Jurkat and S. cerevisiae BY4741 cells and purified calf histones were separated by SDS-PAGE and analysed by western blotting using the primary antibodies D5A12, recognizing arginine mono-methylation in an RGG context, ASYM24, recognizing asymmetric dimethyl arginine, Me-R4-100 and ICP0801, recognizing arginine mono-methylation in a degenerate sequence context and SYM10, recognizing symmetric dimethyl arginine (red). GAPDH was used as a loading control in each case (green).*

### 2.3.2. Influence of Growth Medium on Protein Methylation

As it had been observed that the composition of the medium used for *S. cerevisiae* growth can have a strong impact on arginine methylation levels (Lakowski et al. 2013), it was investigated by western blotting which medium would be appropriate for experiments for the global identification of protein methylation. As synthetic medium needs to be used for MethylSILAC labelling, it was particularly important to test if the use of synthetic media would reduce methylation compared to YPD medium which is routinely used as a rich growth medium (Anon 1994).

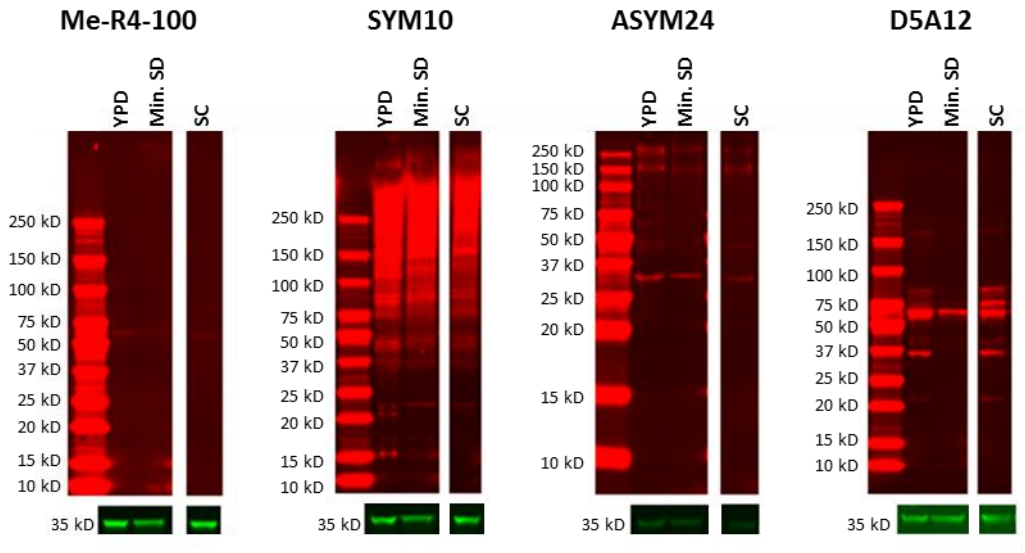



***Figure 2.3: Comparison of protein methylation levels upon growth in different culture media by western blotting.*** *Yeast cells were grown in YPD and minimal SD (Min. SD) and synthetic complete (SC) medium and crude lysate used for western blotting with antibodies Me-R4-100, SYM10, ASYM24 and D5A12 (red). GAPDH was used as a loading control (green).*

The western blotting results are shown in fig. 2.3. In agreement with the findings by Lakowski et al. (2013), a reduction of arginine mono-methylation levels by antibodies Me-R4-100 and D5A12 were observed for yeast cells grown in minimal SD compared to YPD medium. No obvious reduction of methylation levels occurred when using SC medium, so this medium was chosen for subsequent MethylSILAC experiments.

### 2.3.3. Protein Lysine Methylation Assessed by Protein Immunoaffinity Purification

To identify proteins methylated on lysine residues in *S. cerevisiae* a pull-down with an anti-methyl-lysine antibody was performed. The antibody used (Acris, AP09328PU-N) had yielded the highest number of methylation sites in a comparison of different antibodies on proteins from a human cell line (Bremang et al. 2013). The results obtained by separating proteins by SDS page and Coomassie staining are shown in fig. 2.4. Visually, no clear differences in the band pattern were observed between the eluates from pull-down samples where antibody was omitted and such with anti-methylated lysine antibody. Further, no differences were observed whether yeast lysate had or had not been treated with the

demethylase inhibitors TCP and IOX1 (Hopkinson et al. 2013; Lee et al. 2006). As the methylated proteins may have been immunoprecipitated below the detection limit of Coomassie staining, each gel lane containing eluate was cut into 13 pieces and subjected to in-gel digestion. The obtained digests were analysed by high-resolution mass spectrometry on MS and MS/MS level in data dependent acquisition mode with HCD fragmentation. A spectral database search identified 116 and 141 proteins for pull-downs on samples treated with and without demethylase inhibitors respectively and 121 proteins for the control in which the antibody had been omitted. The corresponding pull-down on HeLa S3 cells returned 167 protein identifications (p-value cut-off: 0.05; corresponding to an FDR < 6%). Fourteen proteins were found in the two pull-down samples (with and without demethylase inhibitors TCP and IOX1), but not in a control in which the antibody had been omitted. These therefore represent candidate proteins that might specifically interact with the antibody. Additionally, 52 proteins were found exclusively in the sample to which a demethylase inhibitor had been added. These may represent proteins that are recognized by the antibody via methylation sites, which however are removed by demethylases if no inhibitor is added. The protein sequence coverages were limited to a median of 8.3%.

To verify some of these candidates, it was tested whether methylated peptides had indeed been identified, utilizing the MethylSILAC strategy (Ong et al. 2004). A single peptide each was confidently identified in the pull-downs with and without demethylase inhibitors, which in both cases was STTTGHLIYK(me1), derived from Elongation factor 1-alpha. As the calculation of an FDR is not appropriate at very low numbers, confident identification here refers to E-values <0.05 ($5x10^{-6}$ and $6x10^{-4}$ respectively) and a fragment spectrum satisfactory to visual inspection, requiring a b- or y-ion series of at least four consecutive fragments and including peaks originating from fragmentation directly neighbouring the methylated residue (for at least one partner of the MethylSILAC pair) (Fig. 2.5).

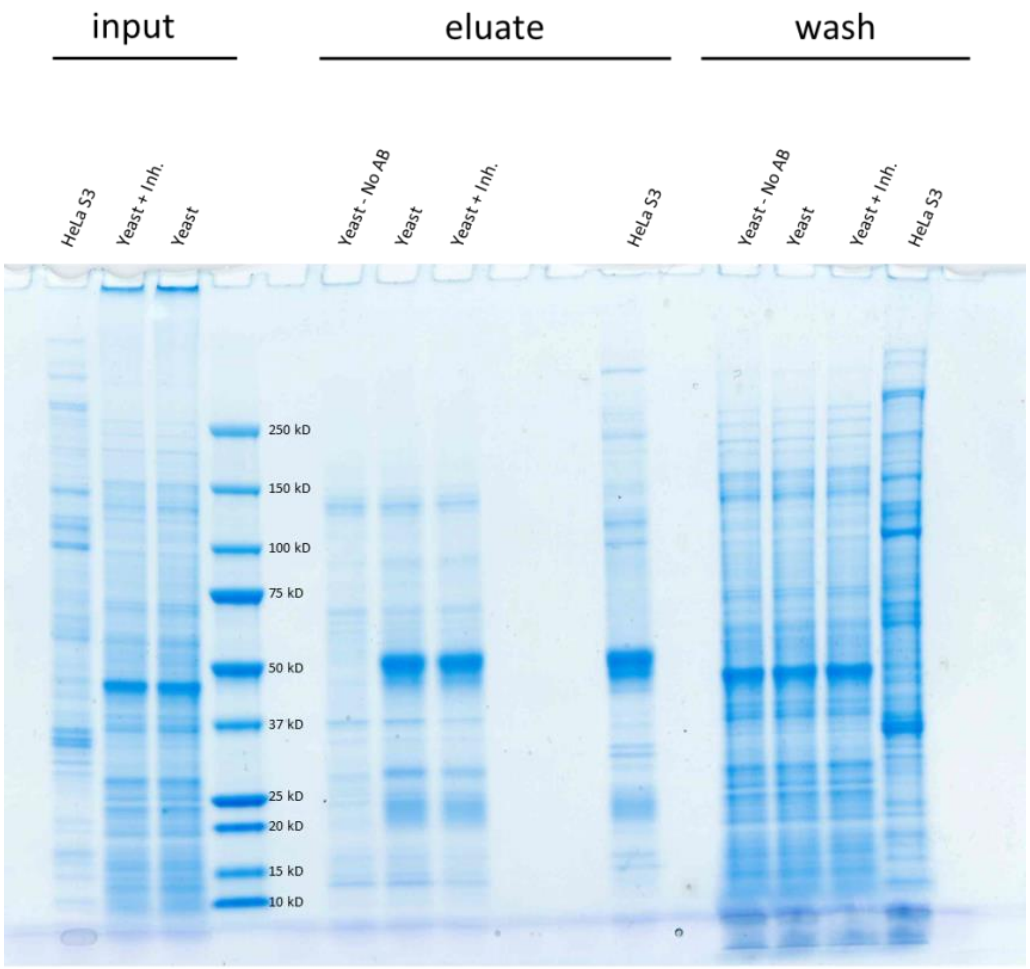



***Figure 2.4: SDS-PAGE gel of protein enrichment by protein immunoaffinity purification with an anti-methyl lysine antibody.*** *Lysates were obtained from HeLa S3 and S. cerevisiae cells and methylated proteins enriched through IP with an anti-methylated lysine antibody (Acris, AP09328PU-N). The demethylase inhibitors TCP and IOX1 were each added at 1 mM concentration to one sample of yeast lysate (Yeast + Inh.). Antibody was omitted during the pull-down for one sample of yeast lysate (Yeast – No AB). Inputs (1% of lysate), eluates and washes were resolved on the gel and detected by Coomassie staining.*

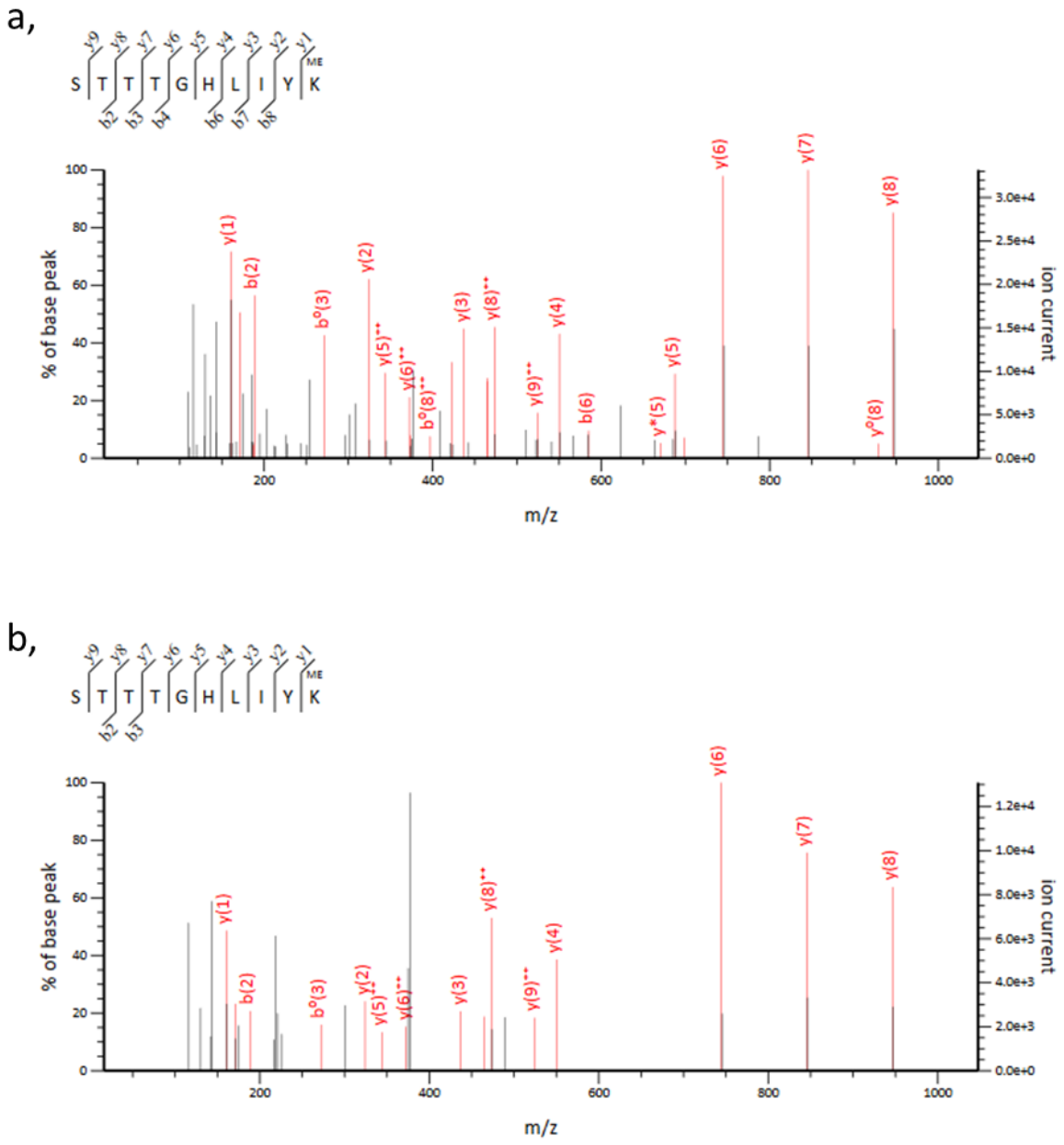



***Figure 2.5: MS/MS spectra of methylated peptides identified by protein IP.*** *Fragment spectra of* $_{21}$*STTTGHLIYK(me1)*$_{30}$ *(Elongation factor 1-alpha; Uniprot accession: P02994) from samples after protein pull-down with anti-methyl lysine antibody from yeast lysates without (a) and with demethylase inhibitors added (b) respectively. Only the MS/MS spectra of the light versions of the MethylSILAC pair are shown.*

### 2.3.4. Enrichment of Lysine Methylated Proteins via Protein Domain Interaction

While antibodies represent the most widely used reagent for proteome-wide enrichment of post-translationally modified proteins, protein domains binding specifically to a modification can provide an alternative. A successful application of this experimental strategy for the identification of protein lysine methylation employed an affinity reagent created from the three malignant brain tumour domain repeats (3xMBT) of L3MBTL1 (Moore et al. 2013). The MBT-domain was chosen as it had been shown to bind mono- and di-methyl lysine with high specificity, without a strong dependency on sequence context and because its binding is only affected to a limited extent by neighbouring modifications (Nady et al. 2012; Moore et al. 2013).

While there are no examples of MBT-domain containing proteins in *S. cerevisiae*, it was argued that if this domain recognizes lysine methylation independently of sequence context it may detect methylation sites on proteins regardless of the species of origin.

The structural basis for the interaction of the MBT domain with methylated lysines has been determined and it has been found that a substitution of aspartate D355 to asparagine in the binding pocket reduces the affinity to methylated lysine, while not affecting the domain structure (Li et al. 2007; Moore et al. 2013). This mutant version of 3xMBT can therefore be used as a negative control in purifications. Plasmids encoding the 3xMBT and the 3xMBT-D355N constructs for bacterial expression were kind gifts of the Gozani lab.

For the enrichment of proteins mono- and di- methylated on lysine an attempted pull-down was performed on MethylSILAC labelled *S. cerevisiae* lysate using 3xMBT as an affinity reagent, which had been shown to interact with mono- and di-methyl lysine in human cell lysates (Moore et al. 2013). The eluted proteins were separated by SDS PAGE. HEK293T nuclear extracts were used as a positive control as they had been shown to yield identifications of methylated proteins by this strategy (Moore et al. 2013). While a high

number of bands could be detected by silver-staining in the eluate of the 3xMBT and 3xMBT-D355N pull-down on HEK293T lysates, only a few are visible for yeast lysate.

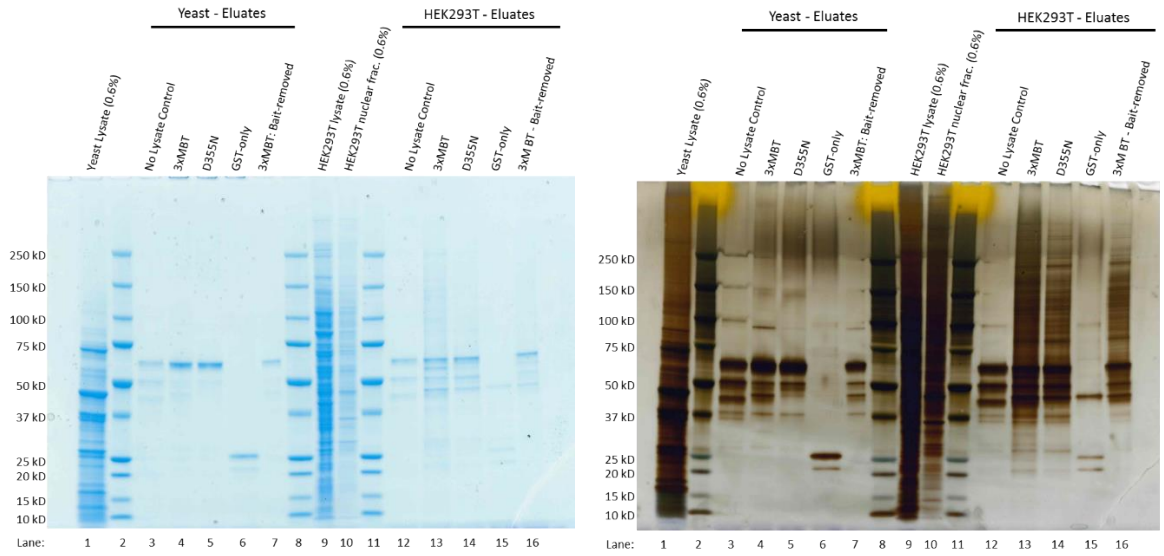



***Figure 2.6: Enrichment of methylated proteins via the 3xMBT bait construct.*** *Coomassie (left) and silver (right) stained SDS-PAGE gels of crude lysates from yeast (lane 1) and HEK293T cells (lane 9), nuclear lysate of HEK293T cells (lane 10) and eluates from pull-down experiments using either the wild type 3xMBT, a 3xMBT-D355N mutant or GST-only coupled to beads. 'No Lysate Control' represents an experiment in which no lysate was added to a pull-down with the 3xMBT construct. '3xMBT', 'D355N' and 'GST-only' indicate the bait used for the pull-down. 'Bait removed' represents eluate after a second pull-down for removal of the bait by re-binding to beads.*

It was noted in the original report of 3xMBT as a tool for the discovery of methylated proteins that methylated proteins are also found in the 3xMBT-D355N pull-downs, albeit in most cases in lower amounts (Moore et al. 2013), potentially explaining bands in the D355N pull-down (Fig. 2.6, lane 14). Additionally, proteins interacting with the MBT domains in a methylation-independent manner are also expected to be seen both with the wild type (wt) and mutant version of the construct.

A control in which the affinity-reagent was replaced with the GST-domain alone was included to visualize proteins binding unspecifically to the beads or GST. The eluates from this control are much cleaner than of 3xMBT-D355N for both experiments on budding yeast and

HEK293T samples. As it was observed that the use of GST-3xMBT results in more than one band after purification from *E. coli*, a further control in which yeast or HEK293T lysate was omitted from the pull-down was additionally included (Fig. 2.6, lanes 3 and 12). This serves to visualize strong interactors in the bacterial proteome and degradation products of the bait protein. Indeed, most bands observed in eluates from *S. cerevisiae* are also present in this control, indicating that the bands originate from the original purification of the bait. Eluates before and after high-salt elution and removal of the bait with GSH-beads show similarly strong bands for the affinity construct; this observation indicates that the re-binding of the construct was not successful (Fig. 2.6, lanes 7 and 16).

In contrast to the high number of bands in the eluates of HEK293T samples, only a single band of approximately 100 kDa is present in the yeast sample (Fig. 2.6, silver stain, lane 4), but clearly absent in the control without lysate added (Fig. 2.6, silver stain, lane 3). This band was identified to be EF2 by mass spectrometry. The much lower intensity of this band in the 3xMBT-D355N pull-down suggests that this interaction strongly depends on the aspartate at position D355 in 3xMBT, indicating a methylation-dependent binding (Fig. 2.6, silver stain, lane 4 vs. lane 5).

The gel lanes were cut into pieces and subjected to in-gel digestion and LC-MS/MS analysis. While the experiment was performed with several proteases, only tryptic digestion allowed the identification of methylation sites. After filtering by MethylSILAC H/L-ratio and Mascot E-value to achieve a FDR < 0.05, identified methylated peptides were discarded if the methylation site was not surrounded by fragment ions. The methylated peptides identified in yeast and HEK293T cells are listed in table 2.3.

**Table 2.3.** Methylation sites identified via 3xMBT-based pull-down.

a, Methylation sites identified in *S. cerevisiae.*

| Protein ID | Protein Name | Site | E-value* | Peptide | Protein MW (kD) |
|---|---|---|---|---|---|
| P0CX27 | Rpl42A (60S ribosomal protein L42-A) | K40 | 1.9E-05 | ASLFAQGK(me1) | 12.2 |
| P0CX27 | Rpl42A (60S ribosomal protein L42-A) | K40 | 6.4E-04 | ASLFAQGK(me1)R | 12.2 |
| Q12420 | Snu66 (66 kDa U4/U6.U5 small nuclear ribonucleoprotein component) | K337 | 3.1E-02 | TSSNNELK(me1) | 66.4 |
| P02994 | EF1A (Elongation factor 1-alpha) | K30 | 8.8E-6 / 2.2E-4 | STTTGHLIYK(me1) | 50.0 |
| P02994 | EF1A (Elongation factor 1-alpha) | K316 | 2.9E-3 / 4.3E-5 | NVSVK(me2)EIR | 50.0 |
| P02994 | EF1A (Elongation factor 1-alpha) | K390 | 9.6E-5 / 6.4E-5 | KLEDHPK(me1) | 50.0 |
| P32324 | EF2 (Elongation factor 2) | K613 | 9.8E-03 | DDFK(me1)AR | 93.3 |
| P32324 | EF2 (Elongation factor 2) | K613 | 3.8E-02 | DDFK(me2)AR | 93.3 |
| P46677 | Taf1 (Transcription initiation factor TFIID subunit 1) | K974 | 6.8E-03 | INNLEELEK(me2) | 120.7 |

b, Methylation sites identified in HEK293T cells.

| Protein ID | Protein Name | Site | E-value* | Peptide | Protein MW (kD) |
|---|---|---|---|---|---|
| O14647 | CHD2 (Chromodomain-helicase-DNA-binding protein 2) | K1403 | 1.4E-03 | EK(me2)QM(ox)SSR | 211.3 |
| Q04864 | REL (Proto-oncogene c-Rel) | K251 | 2.6E-02 | TPPYCK(me2) | 68.5 |
| Q5CZC0 | FSIP2 (Fibrous sheath-interacting protein 2) | K256 | 2.9E-03 | IEEEWK(me2) | 780.6 |
| P68104 | EF1A1 (Elongation factor 1-alpha 1) | K165 | 1.8E-2 / 1.4E-2 | MDSTEPPYSQK(me1)R | 50.1 |
| P68104 | EF1A1 (Elongation factor 1-alpha 1) | K165 | 5.8E-5 / 2.0E-4 | MDSTEPPYSQK(me2)R | 50.1 |
| P68104 | EF1A1 (Elongation factor 1-alpha 1) | K165 | 3.5E-03 | M(ox)DSTEPPYSQK(me2)R | 50.1 |
| P68104 | EF1A1 (Elongation factor 1-alpha 1) | K318 | 1.1E-3 / 8.2E-4 | NVSVK(me2)DVR | 50.1 |
| Q6NVY1 | HIBCH (3-hydroxyisobutyryl-CoA hydrolase) | K377 | 4.2E-02 | EVTEEDLNNHFK(me2) | 43.5 |
| Q96JW4 | S41A2 (Solute carrier family 41 member 2) | K5 | 3.0E-02 | M(ox)TNSK(me1)GRSITDK | 62.5 |
| Q96PQ7 | KLHL5 (Kelch-like protein 5) | K443 | 2.4E-02 | LIM(ox)EAMK(me2) | 84.5 |
| Q9UN79 | SOX13 (Transcription factor SOX-13) | K470 | 6.2E-03 | SM(ox)TNQEK(me2) | 69.2 |

*Both E-values are given if both partners of the heavy-light pair were identified.

Five proteins were found to be methylated in *S. cerevisiae* using enrichment via protein domain interaction. Three methylation sites (K30, K316 and K390) were detected on elongation factor 1-alpha (EF1A). The single methylation site (K613) on Elongation factor 2 was found in both mono- and di-methylated form. Mono-methylation of K40 on 60S ribosomal protein L42-A was detected on a peptide without and with a missed cleavage and all sites on EF1A were identified via both the heavy and the light peptide, providing additional confidence for the assignment. Nine methylation sites were mapped to eight proteins from HEK293T cells, with two methylation sites (K165 and K318) mapped to EF-1-alpha-1. One site (K165) was found both with mono- and di-methylation and with and without methionine oxidation, while the other site (K318) was identified in both heavy and light forms, rendering them high-confidence identifications.

### 2.3.5. Identification of Methylation Sites via Sample Fractionation by HILIC

In an attempt to identify methylation sites without the potential bias introduced by an antibody or another affinity reagent, LC-MS/MS analysis on a HILIC-fractionated protein digest was performed without prior enrichment (Fig. 2.7).

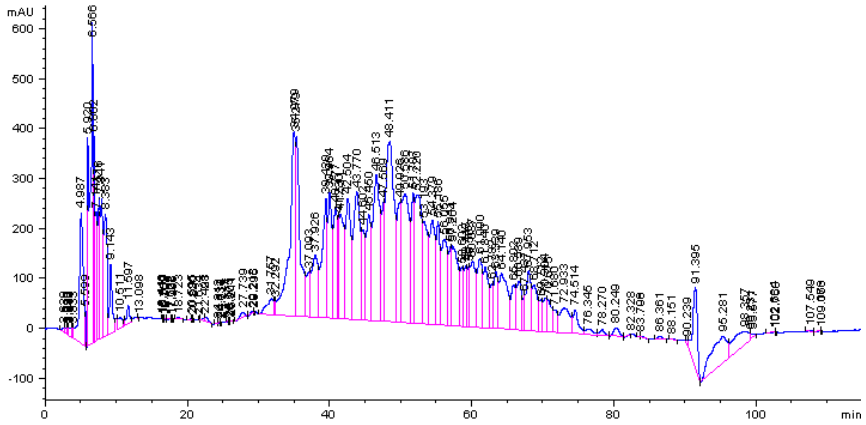



***Figure 2.7: HILIC chromatogram of Saccharomyces cerevisiae protein digest.*** *About 1 mg of proteins from crude lysate were subjected to tryptic digestion and separated by HILIC. UV absorption at 214 nm is shown. Fractions were collected at 1 min intervals from a run time of 22 to 90 min.*

To estimate which proportion of the *Saccharomyces cerevisiae* proteome was covered by HILIC mediated separation of peptides from crude extract and LC-MS/MS analysis of the fractions, the data were first searched with the Preview software (Protein Metrics) to determine frequent modifications and appropriate parameters for Mascot searches. Error tolerances of 10 ppm for precursor and 0.07 Da for fragment ions were found suitable based on the error distributions calculated by Preview. Methionine oxidation which had been found with a frequency of 7% and deamidation of asparagine and glutamine which had been found with 3% in all identified peptides were included as variable modifications in the subsequent Mascot search. Interestingly, di-methylation on lysine and arginine was also found among the most frequent modifications, with a frequency of 0.6%.

The distribution of sequence coverages obtained on the 5,553 proteins identified with these settings (E < 0.05, corresponding to a FDR < 2.5%) is shown in fig. 2.8. The median sequence coverage was 44% (1st quartile: 34%, 3rd quartile: 56%).

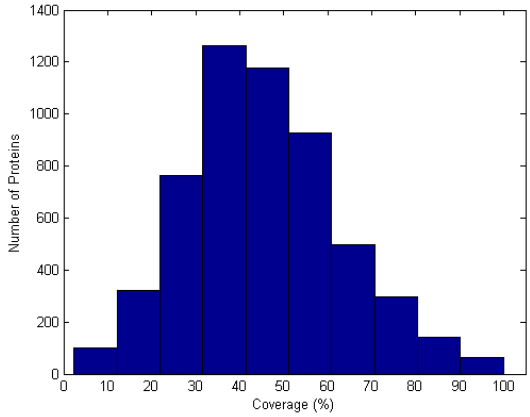



***Figure 2.8: Histogram of the number of proteins identified at different sequence coverages**. Tryptic peptides from yeast crude lysate were seperated into 68 fractions by HILIC and analysed by LC-MS/MS. Data are based on a total of 5,553 proteins identified at E < 0.05 by Mascot search.*

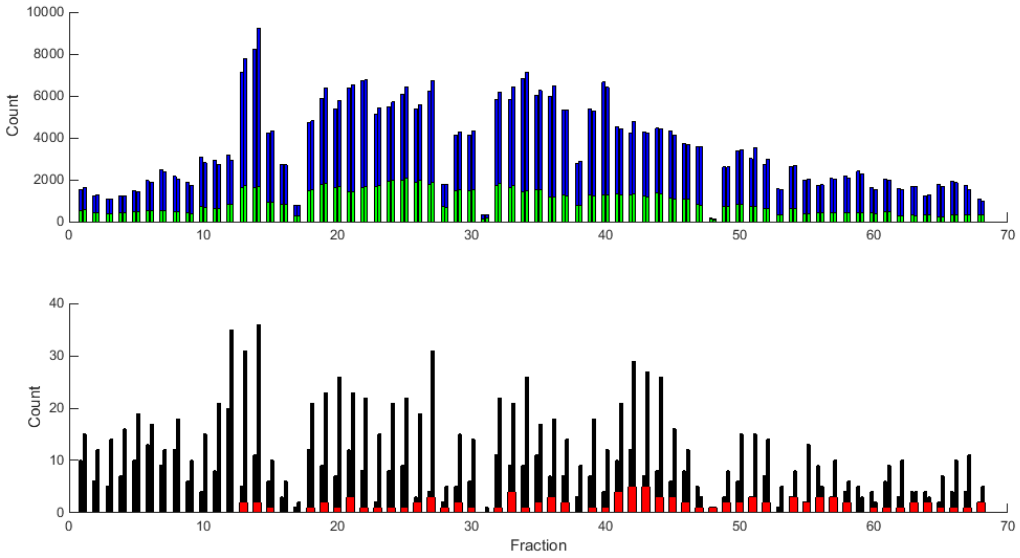



***Figure 2.9: Counts of identifications per HILIC fraction.** Top, Total number of peptide spectrum matches (blue) and peptides (green) at a Mascot E-value < 0.05. Bottom, Methylated-peptides at Mascot E-value < 0.05 (black) and methylated peptides at MethylQuant FDR < 0.05 (red). For total PSMs, peptides and methylated peptides the two bars per fraction represent results with heavy and light parameters.*

The MethylQuant software (developed by V. Geoghegan, University of Oxford) was employed to make use of the MethylSILAC strategy to increase the confidence of assigned methylation sites. Common $5^{th}$ and $95^{th}$ percentiles of the H/L-ratios of peptides from all fractions at E <0.05 were determined and set as cut-offs. The results were further filtered based on their E-values to achieve a FDR < 0.05 based on a target-decoy strategy and by manual inspection, requiring fragment ions directly surrounding the methylation site.

Fig. 2.9 depicts between about 2,000 to 8,000 peptide spectrum matches and about 200 to 500 peptides being identified at a Mascot E-value below 0.05 in each fraction. About ten peptides were identified as methylated per fraction according to Mascot search at E < 0.05, less than five remained after processing with MethylQuant and filtering to an FDR < 0.05. Note that most of these methylated peptides were detected in more than one fraction. Eight, four and two sites were found for lysine mono-, di- and tri-methylation respectively and two sites each for arginine mono- and di-methylation (Table 2.4). The corresponding fragment spectra are shown in fig. 2.10 and fig. A2.1 in the appendix.

The lysine methylation sites were distributed to seven proteins: Elongation factor 1-alpha (EF1A), Elongation factor 2 (EF2), Glyceraldehyde-3-phosphate dehydrogenase (G3P1, G3P2 or G3P3), histone H3, Pyruvate decarboxylase isozyme 1 (Pdc1), 60S ribosomal protein L42 (Rpl42A or Rpl42B) and V-type proton ATPase catalytic subunit A. In the case of Glyceraldehyde-3-phosphate dehydrogenase and 60S ribosomal protein L42 the data cannot distinguish if the methylated peptide is derived from one or several isoforms. The four arginine methylation sites are found on four individual proteins: Guanine nucleotide exchange factor (Lte1), Nucleolar protein 3 (Npl3), Nuclear localisation sequence-binding protein (Nsr1) and Nuclear migration protein (Num1).

**Table 2.4.** Methylation sites identified via HILIC fractionation

a, Lysine methylation sites found by HILIC-fractionation and LC-MS/MS

| Protein Accession | Protein Name | PTM | Methylated Residue | Modified Peptide |
|---|---|---|---|---|
| P02994.1 | EF1A | Kme1 | 30 | STTTGHLIYK(me1) |
| P02994.1 | EF1A | Kme3 | 79 | GITIDIALWK(me3)FETPK |
| P02994.1 | EF1A | Kme1 | 316 | NVSVK(me1)EIR |
| P02994.1 | EF1A | Kme2 | 316 | NVSVK(me2)EIR |
| P02994.1 | EF1A | Kme1 | 390 | KLEDHPK(me1) |
| P32324.1 | EF2 | Kme2 | 613 | DDFK(me2)AR |
| P00360.3; P00358.3; | G3P1; G3P2; | | | |
| P00359.3 | G3P3 | Kme1 | 225 | VLPELQGK(me1) |
| P61830.2 | H3 | Kme1 | 80 | EIAQDFK(me1)TDLR |
| P61830.2 | H3 | Kme2 | 80 | EIAQDFK(me2)TDLR |
| P61830.2 | H3 | Kme3 | 80 | EIAQDFK(me3)TDLR |
| P06169.7 | Pdc1 | Kme1 | 520 | VATTGEWDK(me1) |
| P0CX28.1; P0CX27.1 | Rpl42A; Rpl42B | Kme1 | 40 | ASLFAQGK(me1)R |
| P0CX28.1; P0CX27.1 | Rpl42A; Rpl42B | Kme1 | 55 | QSGFGGQTK(me1)PVFHK |
| P17255.3 | VatA | Kme2 | 207 | K(me2)SDFTLYHTWPVR |

b, Arginine methylation sites found by HILIC-fractionation and LC-MS/MS

| Protein Accession | Protein Name | PTM | Methylated Residue | Modified Peptide |
|---|---|---|---|---|
| P07866.2 | Lte1 | Rme1 | 649 | ESSPLNDSR(me1) |
| Q01560.1 | Npl3 | Rme2 | 363 | GGYSR(me2)GGYGGPR |
| P27476.1 | Nsr1 | Rme2 | 357 | LDFSSPRPNNDGGRGGSR(me2) |
| Q00402.2 | Num1 | Rme1 | 175 | LTNLSMECR(me1) |

The methylation site on histone H3 was found in mono-, di- and tri-methylated form, while position 316 on EF1A is present in mono- and di-methylated state. Additionally, EF1A was found methylated at four different lysines, while Rpl44a/b is methylated at K40 and K55.

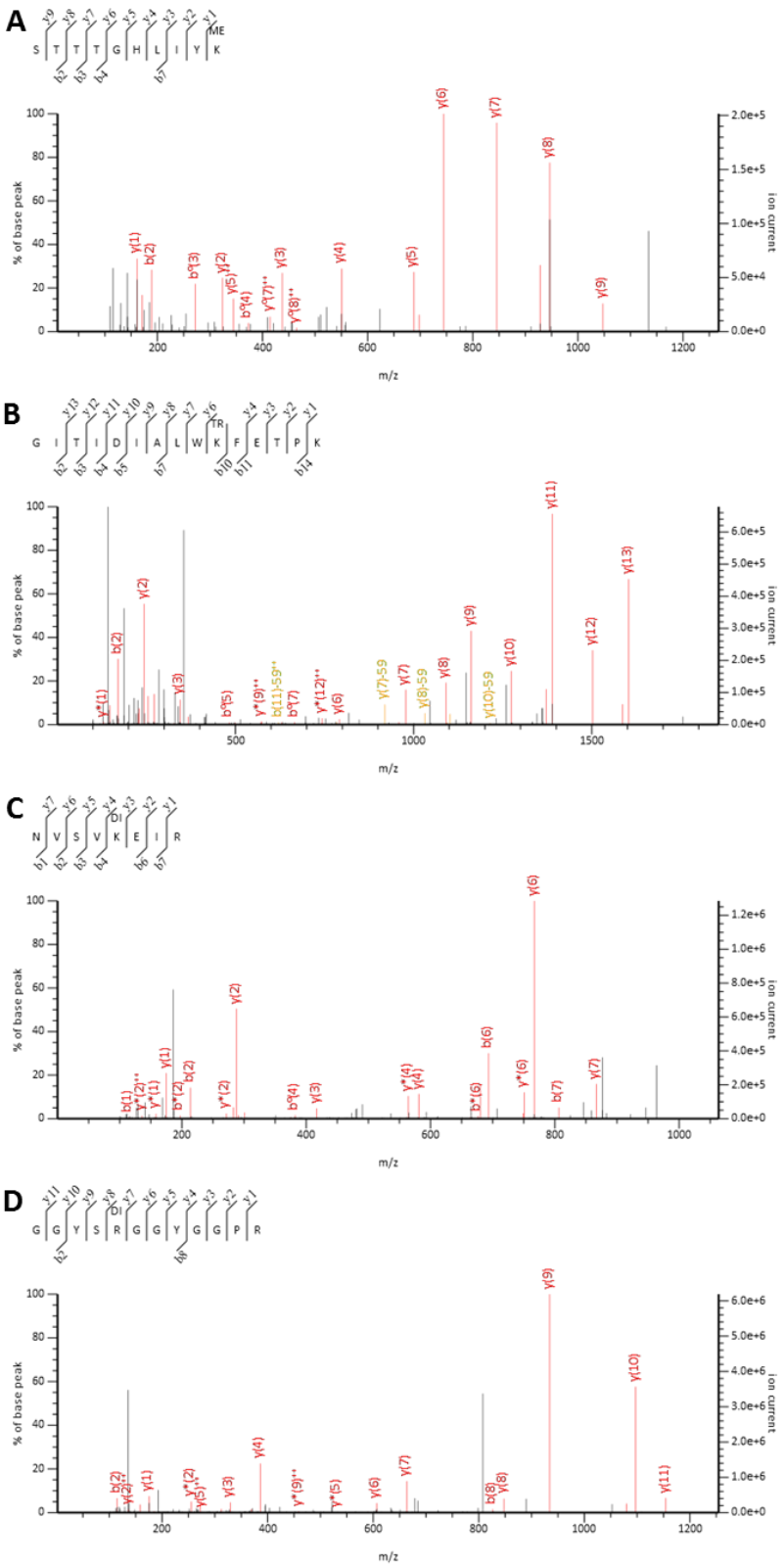



***Figure 2.10: Fragment spectra of methylated peptides identified by HILIC fractionation and LC-MS/MS.*** *a, Peptide $_{21}$STTTGHLIYK(me1)$_{30}$ of protein EF1A b, $_{70}$GITIDIALWK(me3)FETPK$_{84}$ of EF1A c, $_{312}$NVSVK(me2)EIR$_{319}$ of EF1A d, $_{359}$GGYSR(me2)GGYGGPR$_{370}$ of Npl3. * Neutral loss of ammonia; $^{o}$ Neutral loss of water; '-59' indicates neutral loss of trimethyl amine.*

### 2.3.6. Identification of Arginine Methylation Sites via Peptide Immunoaffinity Purification

Alternatively to affinity purification of methylated proteins, the enrichment step can be performed at the peptide level after proteolytic digestion (Guo et al. 2014). The advantage of this strategy is that the unmethylated peptides of a protein bearing a methylation site are not present during LC-MS/MS analysis, therefore rendering the sample less complex.

#### *2.3.6.1. MethylSILAC Facilitates Improved Identification of Methylation Sites*

During the time of this study, a set of antibodies targeting arginine mono-methylation became commercially available and were successfully used for affinity purification of methylated peptides from human cells and mouse tissues (Guo et al. 2014).

An equivalent strategy to comprehensively survey protein arginine methylation in *S. cerevisiae* was applied here. The identification of at least one feature of the heavy/light SILAC pair and a heavy to light peak intensity ratio of >0.1 and <10 were required (Fig. 2.11 A, B). The corresponding experiment was performed on peptides from Jurkat cell lysates as a positive control (Fig. 2.11 C, D; Fig. 2.12). As shown in fig. 2.11 E, the false discovery rate ($FDR = \frac{2 \times \#\ decoy\ PSMs}{\#\ target\ PSMs + \#\ decoy\ PSMs}$) calculated only on PSMs to peptides containing a methyl-group is consistently higher than on all PSMs at the same E-value cut-off after the Mascot search. It was next considered if the use of the MethylSILAC strategy reduces the number of false positives, and indeed the FDR at a given E-value after applying MethylQuant and imposing a threshold on the heavy to light ratio was consistently lower than after the database search alone (Fig. 2.11 F). This indicates an increase in specificity, but might however come at a loss of sensitivity (i.e. a lower number of true positive identifications at the given E-value cut-off). However, the number of PSMs at a given FDR was consistently higher with than without MethylSILAC (Fig. 2.11 G), which confirms that overall the sensitivity at a given FDR is increased when using the MethylSILAC strategy.

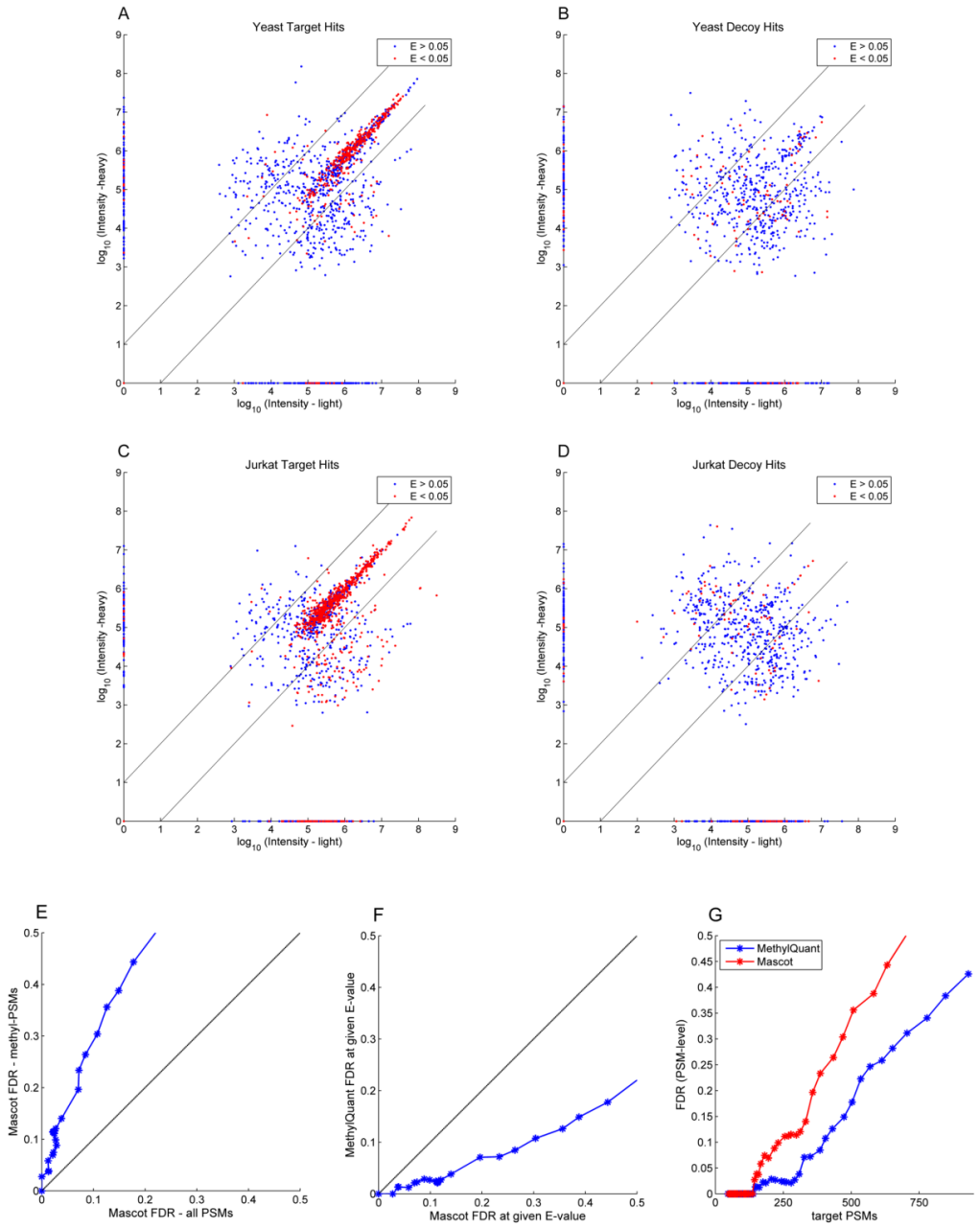



***Figure 2.11: Confidence assessment of protein arginine methylation mapping.*** *(A–D) Light versus heavy intensities (log10 transformed; peaks with an intensity of 0 are replaced by 0 on the log-scale) of feature pairs identified as arginine methylated by Mascot search. Identifications with a Mascot E-value < 0.05 are shown in red and > 0.05 in blue. An arbitrary upper threshold of E < 0.3 was imposed in the second case. Feature pairs were detected using MethylQuant. The heavy to light ratio thresholds of 0.1 and 10 applied to these data are shown as black lines. Data obtained for S. cerevisiae (A, B), and Jurkat cells (C, D) are shown for Mascot target (A, C) and decoy (B, D) hits. (E) FDRs at different E-value thresholds after Mascot search for all peptide spectrum matches (PSMs) versus matches to methylated peptides. (F) FDR after Mascot search versus FDR at the same E-value cut-offs after processing with MethylQuant and filtering to a heavy-to-light ratio of > 0.1 and < 10. (G) FDR at PSM level at different numbers of identified target PSMs for processing with Mascot alone (red) or via the MethylQuant strategy (blue). Plots E–G are based on S. cerevisiae samples. (Plank et al. 2015)*

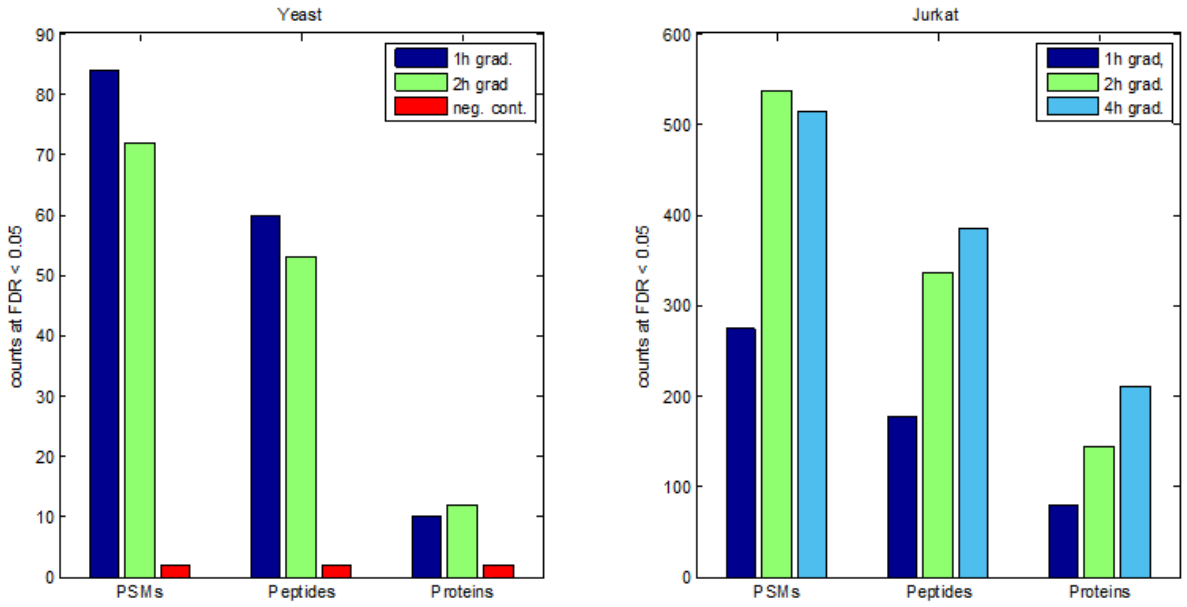



***Figure 2.12: Bar chart of peptide-spectrum matches (PSMs), peptides and proteins harbouring methylation sites.*** *Data were obtained by database search and filtering to a heavy-light ratio of the precursors of >0.1 and <10 and FDR <0.05, but without manual filtering and are shown for yeast and Jurkat samples and different gradient length in LC-MS/MS were explored. The negative control was obtained by omitting antibodies from the pull-down setup. (Plank et al. 2015)*

#### *2.3.6.2. Identification of Protein Arginine Methylation Sites*

After filtering based on the MS1 intensity ratio and MS2 spectral matching, 62 matches to peptides of the *S. cerevisiae* proteome remained. The list comprises 41 unambiguous methylation sites, several of which are either novel or provide evidence for mono-methylation of known di-methylation sites (Table 2.5, Table 2.6) (Low et al. 2013; Erce et al. 2013; Hart-Smith et al. 2012; Green et al. 2002; Hsieh et al. 2007). The sites were mapped to 13 proteins, for six of which methylation had not been reported previously, while the remaining seven provide confirmation of previously described methylated proteins. Fragment spectra for each novel site are shown in fig. 2.11 and fig. A2.2 in the appendix. These spectra show near-complete coverage of at least either the b- or y-ion series and the methylation site is directly surrounded by fragment ions, resulting in confident site localisation. The identification of several hundred methylated peptides from Jurkat lysate indicated that the lower number of sites in *S. cerevisiae* was not due to technical difficulties (Fig. 2.12). Three to five methylation sites each were unambiguously mapped to Nop1, Nsr1, Sbp1, Dbp2, Ded1, and Psp2 and 12

to Npl3. Several of these sites were covered by more than one peptide and often two methylation sites were found on the same peptide as the sites are in close proximity. An additional methylated peptide matched to Nop1 in three different positions (R16, R34, R52), making it an ambiguous assignment. A further MS/MS spectrum for Psp2 is ambiguous in localizing a methylation to one of the five unambiguous sites (R551) or to an additional site (R553) (Table 2.5). Twelve unambiguously mapped methylation sites were further located on Npl3, several with MS/MS spectra matching to different peptides or modification states of the peptide. Furthermore, four methylation sites were detected on Sbp1 and five were mapped to the double-stranded RNA binding protein Dbp2. Three methylation sites were mapped to Ded1, which is like Dbp2, an ATP-dependent RNA helicase. In the case of Nab2, one methylation site could be mapped to R222, while the second methylation in this peptide might be located on R226 or R229. Co-fragmentation of a mixture of peptides with methyl-groups localized at different residues is also possible.

Note that the fragment ions surrounding R222 provide evidence against di-methylation of this residue and therefore suggest that the additional 14 Da difference in the precursor mass is attributed to methylation of one of the remaining arginines. One methylation site each was detected on Gar1, Rpl4, Elp4, Lge1 and Lsm4. Database search identifications at FDR < 0.05 for both the heavy and light version of the methylated peptides originating from Lsm4 and Lge1 were detected. For the Elp4 peptide the complete y-ion series could be matched to the spectrum (Fig. 2.13) and the methylation of Gar1 had been previously reported (Xu et al. 2003). When searching for di- additionally to mono-methylation, 14 peptides indicated the occurrence of di-methylation at sites reported above and at two additional sites on Psp2 (R443, R447) (Table 2.7). However, in many cases, the confidence of site localisation was low, potentially due to co-fragmentation of peptides with different distributions of methyl-groups.

**Table 2.5:** Protein arginine methylation sites in *S. cerevisiae* identified by tandem mass spectrometry (Plank et al. 2015)

| Accession | Protein Name | Prot. Prev. Known | Methylated Residue | Methyl. Site Prev. Known | Study | Modified Peptide |
|---|---|---|---|---|---|---|
| P50109.2 | Psp2 | n | 419 | | | YNGNHNNNNGNFR(me1)GSNR |
| | | | 425 | | | YR(me1)GGPNGSSYK |
| | | | 575 | | | GGR(me1)GSSGNYSNYNNR |
| | | | 538 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR |
| | | | 551 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR |
| | | | 551 or 553 | | | GNYNSSGMNGGSRGR(me1) or GNYNSSGMNGGSR(me1)GR |
| P24783.1 | Dbp2 | n | 43 | | | NSYNDRPQGGNYR(me1)GGFGGR |
| | | | 509 | | | YGGGR(me1)GGR(me1)GGYGR |
| | | | 512 | | | YGGGR(me1)GGR(me1)GGYGR |
| | | | 518 | | | R(me1)GGYGGGR |
| | | | 525 | | | GGYGGGR(me1)GGYGGNR |
| P06634.2 | Ded1 | y | 51 | y(me2) | Erce, 2013 | NNSSNYNNNNGGYNGGR(me1)GGGSFFSNNR |
| | | | 62 | y(me2) | Erce, 2013 | R(me1)GGYGNGGFFGGNNGGSR |
| | | | 578 | y | Erce, 2013 | DNSFR(me1)GGSGWGSDSK |
| P15646.1 | Nop1 | y | 70 | y(me1 and me2) | Low, 2013 | GGSR(me1)GGR(me1)GGAAGGAR |
| | | | 73 | y (me1 and me2) | Low, 2013 | GGR(me1)GGAAGGAR |
| | | | | | | GGSR(me1)GGR(me1)GGAAGGAR |
| | | | | | | GGR(me1)GGAAGGAR(me1) |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK |
| | | | 81 | y(me2) | Low, 2013 | GGAAGGAR(me1)GGAK |
| | | | | | | GGR(me1)GGAAGGAR(me1) |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK |
| | | | 16 or 34 or 52 [(b)] | all three: y(me2as - UniProt by Similarity) | | GGSR(me1)GGFGGR |
| P28007.1 | Gar1 | y | 147 | y(me2as- UniProt probable) | | SGAPGGR(me1)GGASMGR |
| P27476.1 | Nsr1 | y | 353 | y(me2as- UniProt probable) | | LDFSSPRPNNDGGR(me1) |
| | | | 379 | y(me2as- UniProt probable) | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 382 | y(me2as- probable) | UniProt | GGR(me1)GGFRPSGSGANTAPLGR |
| | | | 386 | | | GGAR(me1)GGR(me1)GGFRPSGSGANTAPLGR<br>GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |
| | | | | n | | GGRGGFR(me1)PSGSGANTAPLGR<br>GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |
| P49626.2 / P10664.4 | Rpl4B/Rpl4A | n | 95 | | | SGQGAFGNMCR(me1)GGR |
| Q01560.1 | Nop3 / Npl3 | y | 307 | y(me1 and me2) | Hart-Smith, 2012 | GGFSR(me1)GGFGGPR<br>GGFSR(me1)GGFGGPR(me1)<br>GGFSR(me1)GGFGGPR(me1)GGFGGPR |
| | | | 314 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGFGGPR<br>GGFSR(me1)GGFGGPR(me1)<br>GGFSR(me1)GGFGGPR(me1)GGFGGPR |
| | | | 321 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) |
| | | | 329 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) |
| | | | 337 | y(me1 and me2) | Hart-Smith, 2012 | GGYGGYSR(me1)GGYGGSR |
| | | | 344 | n | | GGYGGSR(me1)GGYDSPR<br>GGYGGSR(me1)GGYDSPR(me1) |
| | | | 351 | y(me1) | Hart-Smith, 2012 | GGYDSPR(me1)GGYDSPR<br>GGYGGSR(me1)GGYDSPR(me1)<br>GGYDSPR(me1)GGYDSPR(me1) |
| | | | 358 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1)GGYSR<br>GGYDSPR(me1)GGYDSPR(me1)<br>GGYDSPR(me1)GGYSR(me1)<br>GGYDSPR(me1)GGYSR(me1)GGYGGPR [(c)] |
| | | | 363 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPRGGYSR(me1)<br>GGYSR(me1)GGYGGPR<br>GGYDSPR(me1)GGYSR(me1)<br>GGYDSPR(me1)GGYSR(me1)GGYGGPR [(c)] |
| | | | 377 | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR |
| | | | 384 | y(me1 and me2) | Hart-Smith, 2012 | GSYGGSR(me1)GGYDGPR<br>GSYGGSR(me1)GGYDGPRGDYGPPR |
| | | | 391 | y (me2) | Hart-Smith, 2012 | GGYDGPR(me1)GDYGPPR |
| | | | 314 or 321 [(b)] | both: y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1) |
| | | | 351 or 358 [(b)] | 351: y(me1); 358: y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1) |
| | | | 377 or 384 [(b)] | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR or NDYGPPRGSYGGSR(me1) |

| | | | | | | |
|---|---|---|---|---|---|---|
| P32505.1 | Nab2 | y | 222 | (region: 180-256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) |
| | | | 226 or 229 | (region: 180-256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) |
| P10080.2 | Ssbp1 / Sbp1 | y | 125 | y(me1 and me2) | Hsieh, 2007 [(a)] | TPGQM(ox)QR(me1)<br>TPGQMQR(me1) |
| | | | 145 | y(me2) | Hsieh, 2007 [(a)] | GGFR(me1)GGYR |
| | | | 165 | y(me2) | Hsieh, 2007 [(a)] | GGAR(me1)GGFNGQK<br>GGAR(me1)GGFNGQKR<br>GR(me1)GGAR(me1)GGFNGQKR |
| P40070.1 | Lsm4 | n | 119 | | | DSNNNR(me1)GNYNR |
| Q02796.1 | Lge1 | n | 39 | | | GGAGGSYYR(me1)GGNASYGAR |
| Q02884.1 | Elp4 | n | 13 | | | GEILNDR(me1)GSGLR |
| P06169.1 | Pdc1 | n | 161 | | | TTYVTQR(me1)PVYLGLPANLVDLNVPAK [(d)] |

[(a)] Sites of methylation are not given in this study, but could be partially derived from the peptides and number of methyl-groups; these sites were found by expressing Sbp1 and Hmt1 in *E.coli.*

[(b)] The identified peptide matches to more than one location in the protein.

[(c)] Only one of the sites was surrounded by fragment ions.

[(d)] This identification resulted from a separate IP with antibody Me-R4-100 only.

**Table 2.6:** Protein arginine methylation sites in yeast identified by immunoaffinity purification with Me-R4-100 antibody (Plank et al., 2015)

| Accession | Protein Name | Prot. Prev. Known | Methylated Residue | Methyl. Site Prev. Known | Study | Modified Peptide | Mascot E-value | Charge | MC | Exp m/z | Calc m/z | Error (ppm) | H/L ratio | conf. meth. pos. (%) | sequence context (if methylated R is C-terminal) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P06634.2 | Ded1 | y | 51 | y(me2) | Erce, 2013 | NNSSNYNNNNGGYNGGR(+14.02)GGGSFFSNNR | 4.00E-08 | 3 | 1 | 951.7418 | 951.7404 | 1.4 | 0.48 | 96.8(R17), 3.2(R27) | |
| | | | 62 | y(me2) | Erce, 2013 | R(+14.02)GGYGNGGFFGGNNGGSR | 8.90E-07 | 3 | 1 | 582.2681 | 582.2681 | -0.1 | 0.40 | 100(R1) | |
| P15646.1 | Nop1 | y | 73 | y (me1 and me2) | Low, 2013 | GGR(+14.02)GGAAGGAR(+14.02)GGAK | 1.50E-05 | 2 | 2 | 614.3368 | 614.3369 | -0.2 | 0.26 | | |
| | | | 81 | y(me2) | Low, 2013 | GGR(+14.02)GGAAGGAR(+14.02)GGAK | 1.50E-05 | 2 | 2 | 614.3368 | 614.3369 | -0.2 | 0.26 | | |
| P06169.1 | Pdc1 | n | 161 | | | TTYVTQR(+14.02)PVYLGLPANLVDLNVPAK | 1.90E-05 | 3 | 0 | 919.5082 | 919.5164 | -8.9 | 0.15 | | |
| P10080.2 | Ssbp1 / Sbp1 | y | 125 | y(me1 and me2) | Hsieh, 2007 (a) | TPGQMQR(+14.02) | 0.0003 | 2 | 0 | 416.2101 | 416.2107 | -1.3 | 1.09 | | GQMQR*GGFRG RGG |
| Q01560.1 | Npl3/Nop3 | y | 307 | y(me1 and me2) | Hart-Smith, 2012 | GGFSR(+14.02)GGFGGPR(+14.02) | 0.00037 | 3 | 1 | 393.8719 | 393.8721 | -0.4 | 0.43 | | GGFSR*GGFGGPRGGFGGPR*GGYGGYSRG |
| | | | 321 | y(me1 and me2); y(me1 and me2) | Hart-Smith, 2012 | GGFSR(+14.02)GGFGGPR(+14.02) | 0.00037 | 3 | 1 | 393.8719 | 393.8721 | -0.4 | 0.43 | | GGFSR*GGFGGPRGGFGGPR*GGYGGYSRG |
| | | | 363 | y(me1 and me2) | Hart-Smith, 2012 | GGYSR(+14.02)GGYGGPR | 0.00078 | 3 | 1 | 399.8636 | 399.8635 | 0.3 | 0.15 | 100(R5) | |
| P50109.2 | Psp2 | n | 538 | | | NNGR(+14.02)GNYNSSGMNGGSR(+14.02)GR | 0.0019 | 3 | 2 | 661.6347 | 661.6367 | -3.1 | 0.52 | 81.9(R4,R17), 10.1(R4,R19) | |
| | | | 551 | | | NNGR(+14.02)GNYNSSGMNGGSR(+14.02)GR | 0.0019 | 3 | 2 | 661.6347 | 661.6367 | -3.1 | 0.52 | 81.9(R4,R17), 10.1(R4,R19) | |

(a) Sites of methylation are not given in this study, but could be partially derived from the peptides and number of methyl groups; these sites were found by co-expression of Sbp1 and Hmt1 in *E.coli.*

MC: Missed Cleavages, Exp m/z: experimental m/z; Calc m/z: calculated m/z; conf. meth. pos. (%): percentage of confidence in localization of methyl-group for the position indicated in brackets as calculated by Mascot.

A repetition of the IP with antibody Me-R4-100 (CST: 8711) alone resulted in the discovery of one additional site on protein Pdc1 (R161) (Table 2.6).

**Table 2.7:** Methylation sites from database search for arginine mono- and di-methylation (Plank et al. 2015).

| Accession | Protein Name | Prot. Prev. Known | Methylated Residue | Methyl. Site Prev. Known | Study | Modified Peptide | Mascot E-value | Charge | Exp m/z | H/L ratio | conf. meth. pos. (%) | site localization |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P27476 | Nsr1 | y | R379me2, R382me1 | y(me2as- UniProt probable), y(me2as- UniProt probable) | | GGAR(me2)GGR(me1)GGFRPSGSGANTAPLGR | 3.7E-04 | 4 | 564.5520 | 0.58 | 11.3 | ambiguous |
| | | | R382me1, R386me2 | y(me2as- UniProt probable), n | | GGARGGR(me1)GGFR(me2)PSGSGANTAPLGR | 6.3E-03 | 4 | 567.8200 | 0.58 | 10.2 | ambiguous |
| | | | R382me2, R386me1 | y(me2as- UniProt probable), n | | GGARGGR(me2)GGFR(me1)PSGSGANTAPLGR | 8.1E-03 | 4 | 564.5521 | 0.58 | 8.7 | ambiguous |
| | | | R382me2 | y(me2as- UniProt probable) | | GGARGGR(me2)GGFRPSGSGANTAPLGR | 2.8E-04 | 4 | 561.0486 | 0.66 | 34.8 | ambiguous |
| | | | R386me2 | n | | GGARGGRGGFR(me2)PSGSGANTAPLGR | 3.3E-04 | 4 | 561.0497 | 0.56 | 26.4 | ambiguous |
| Q01560 | Nop3/ Npl3 | y | R314me1, R321me2 | y(me1 and me2), y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGFGGPR(me2) | 8.3E-03 | 3 | 439.8984 | 0.44 | 98.3 | |
| | | | R321me1, R329me2 | y(me1 and me2), y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me2) | 5.2E-04 | 3 | 496.2441 | 0.42 | 99.9 | |
| | | | R307me1, R314me2 | y(me1 and me2), y(me1 and me2) | Hart-Smith, 2012 | GGFSR(me1)GGFGGPR(me2) | 1.4E-04 | 3 | 398.5439 | 0.39 | 100 | |
| | | | R337me2, R344me1 | y(me1 and me2), n | Hart-Smith, 2012 | GGYGGYSR(me2)GGYGGSR(me1)GGYDSPR | 3.8E-03 | 3 | 746.7000 | 0.55 | 59.1 | ambiguous |
| | | | R358me1, R363me2 | y(me1 and me2), y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1)GGYSR(me2) | 8.4E-04 | 3 | 438.5459 | 0.46 | 98.5 | |
| | | | R358me2, R363me1 | y(me1 and me2), y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me2)GGYSR(me1)GGYGGPR | 7.7E-03 | 3 | 653.3139 | 0.60 | 85.8 | |
| P50109 | Psp2 | n | R443me2, 447me2 | n | | GGHNNRGNR(me2)GGYR(me2)GGSSYNNNNNNTNDNNNNNNNSSSNNNNGSR | 4.5E-04 | 5 | 954.4217 | 0.74 | 18.6 | ambiguous |
| | | | R443me2, R447me1 | n | | GNR(me2)GGYR(me1)GGSSYNNNNNNTNDNNNNNNNSSSNNNNGSR | 5.5E-09 | 4 | 1026.6800 | 0.82 | 13.2 | ambiguous |
| P32505 | Nab2 | y | R222me1, R226me2 | (region: 180-256) | Green, 2002 | GGR(me1)GGNR(me2)GGR | 1.0E-02 | 2 | 493.2736 | 0.35 | 97 | |
| P15646 | Nop1 | y | R73me2, R81me1 | y (me1 and me2), y(me2) | Low, 2013 | GGR(me2)GGAAGGAR(me1)GGAK | 5.3E-05 | 3 | 414.5660 | 0.71 | 100 | |
| | | | R70me2, R73me1, R81me2 | y (me1 and me2), y (me1 and me2), y(me2) | Low, 2013 | GGSR(me2)GGR(me1)GGAAGGAR(me2)GGAK | 1.2E-02 | 4 | 407.4783 | 0.36 | 60.8 | |

Exp m/z: experimental m/z; conf. meth. pos. (%): percentage of confidence in localization of methyl-group for the position indicated in brackets as calculated by Mascot
A site is considered ambiguous if the confidence in localization of the methyl-group is less than three times the confidence of the peptide with rank 2.

### *2.3.6.3. Gene Ontologies of Methylated Proteins*

The program DAVID was used to elucidate the gene ontologies (GOs) for which arginine methylated proteins are enriched (Huang et al. 2009). This was performed separately for the arginine methylation sites identified in this study and for all currently known sites including the ones presented here. The results are shown in table 2.8 and table A2.2 in the appendix, respectively. The GO terms most significantly (Benjamini-Hochberg corrected $p < 0.05$) enriched in both cases are 'RNA binding' and 'RNA processing'. 'RNA transport', 'RNA localisation' and 'RNA export from nucleus' are among the top categories for the list derived from all known methylation sites, while both lists emphasize RNA metabolism of both mRNA and rRNA. The cellular compartments that showed most significant enrichment of

methylated proteins were ‘nucleolus’ for the sites reported in this study and ‘ribonucleoprotein complex’ for the complete list of methylated proteins. These findings are in agreement with previous reports describing the association between arginine methylation and RNA-related functions (Liu & Dreyfuss 1995; Bedford & Clarke 2009).

#### *2.3.6.4. Sequence Context of Arginine Methylation Sites*

To determine whether arginine methylation in *S. cerevisiae* is associated with a defined sequence motif, the residue frequencies surrounding the methylation sites were evaluated. For the sites identified in this study, a clear increase in the frequency of glycine residues is perceptible, especially at the positions directly following the methylated arginine. The sequence context is similar when adding previously identified sites (including di-methyl arginines) (Fig. 2.14). Indeed, 31 of the 41 sites were in an RGG context, while nine were followed by a single glycine. R386 in Nsr1 forms an exception and is followed by proline. A spectrum formed by the combination of a CID- and ETD-based fragmentation on the same precursor resulted in complete sequence coverage of the corresponding peptide adding evidence against the possibility of a mislocalisation of the methylation (Fig. 2.15).

As one of the antibodies in the pull-down mix (D5A12) had been raised against mono-methylated arginine in an RGG context, it was important to ensure that the strong prevalence of RGG-motif peptides in the results was not solely explained by the use of this antibody. Therefore the experiment was separately performed with an antibody raised against arginine methylated peptides without sequence bias alone (Me-R4-100). In this experiment, eight sites were in an RGG, two in an RG and one in an RP context (Table 2.5).

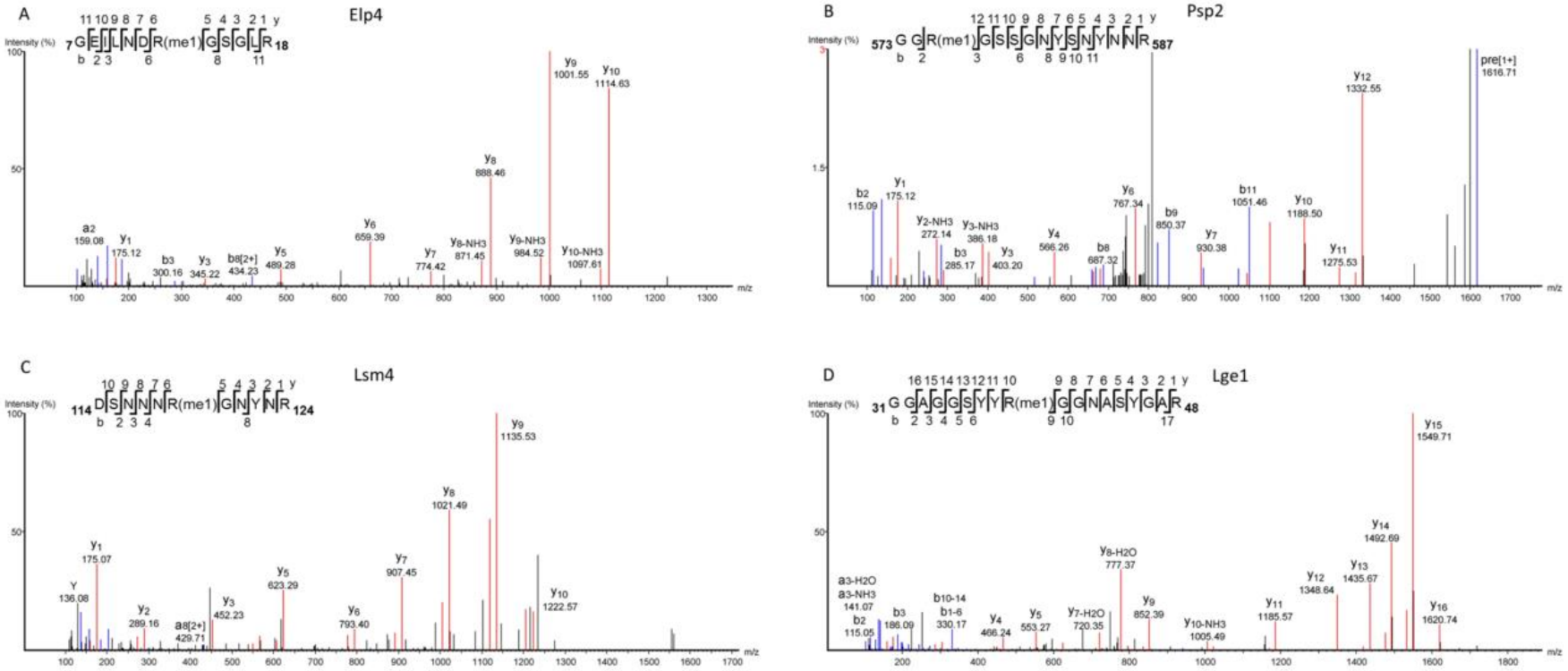



***Figure 2.13: Peptide MS/MS spectra demonstrating protein arginine methylation.*** *MS/MS spectrum of peptide $_{7}$GEILNDR(me1)GSGLR$_{18}$ derived from protein Elp4 (A), $_{573}$GGR(me1)GSSGNYSNYNNR$_{587}$ from Psp2 (B), $_{114}$DSNNNR(me1)GNYNR$_{124}$ from Lsm4 (C) and $_{31}$GGAGGSYYR(me1)GGNASYGAR$_{48}$ from Lge1 (D). (Plank et al. 2015)*

**Table 2.8:** GO_fat terms enriched among proteins found methylated in this study (Plank et al., 2015).

| Category | Term | Count | % | P-Value | Genes | List Total | Pop Hits | Pop Total | Fold Enrichment | Bonferroni | Benjamini | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GOTERM_MF_FAT | GO:0003723~RNA binding | 10 | 11.63 | 4.4E-08 | NOP1, SBP1, GAR1, RPL4A, NAB2, DBP2, LSM4, NSR1, NPL3, DED1 | 11 | 504 | 4190 | 7.6 | 1.9E-06 | 1.9E-06 | 4.1E-05 |
| GOTERM_BP_FAT | GO:0006396~RNA processing | 9 | 10.47 | 1.8E-06 | ELP4, NOP1, GAR1, NAB2, DBP2, LSM4, NSR1, NPL3, DED1 | 12 | 515 | 4870 | 7.1 | 2.5E-04 | 2.5E-04 | 2.2E-03 |
| GOTERM_CC_FAT | GO:0005730~nucleolus | 6 | 6.98 | 3.2E-05 | NOP1, SBP1, GAR1, LSM4, NSR1, NPL3 | 9 | 269 | 4595 | 11.4 | 9.3E-04 | 9.3E-04 | 2.7E-02 |
| GOTERM_BP_FAT | GO:0034470~ncRNA processing | 7 | 8.14 | 3.5E-05 | ELP4, NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 335 | 4870 | 8.5 | 4.7E-03 | 2.3E-03 | 4.1E-02 |
| GOTERM_BP_FAT | GO:0034660~ncRNA metabolic process | 7 | 8.14 | 8.7E-05 | ELP4, NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 393 | 4870 | 7.2 | 1.2E-02 | 3.9E-03 | 1.0E-01 |
| GOTERM_BP_FAT | GO:0006364~rRNA processing | 6 | 6.98 | 9.9E-05 | NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 239 | 4870 | 10.2 | 1.3E-02 | 3.3E-03 | 1.2E-01 |
| GOTERM_BP_FAT | GO:0016072~rRNA metabolic process | 6 | 6.98 | 1.2E-04 | NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 248 | 4870 | 9.8 | 1.6E-02 | 3.2E-03 | 1.4E-01 |
| GOTERM_MF_FAT | GO:0003727~single-stranded RNA binding | 3 | 3.49 | 2.8E-04 | NAB2, NPL3, DED1 | 11 | 11 | 4190 | 103.9 | 1.2E-02 | 6.0E-03 | 2.6E-01 |
| GOTERM_BP_FAT | GO:0042254~ribosome biogenesis | 6 | 6.98 | 6.1E-04 | NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 351 | 4870 | 6.9 | 7.9E-02 | 1.4E-02 | 7.1E-01 |
| GOTERM_CC_FAT | GO:0031981~nuclear lumen | 6 | 6.98 | 9.5E-04 | NOP1, SBP1, GAR1, LSM4, NSR1, NPL3 | 9 | 545 | 4595 | 5.6 | 2.7E-02 | 1.4E-02 | 8.0E-01 |
| GOTERM_BP_FAT | GO:0022613~ribonucleoprotein complex biogenesis | 6 | 6.98 | 1.1E-03 | NOP1, GAR1, DBP2, LSM4, NSR1, NPL3 | 12 | 397 | 4870 | 6.1 | 1.3E-01 | 2.0E-02 | 1.3E+00 |
| GOTERM_CC_FAT | GO:0005732~small nucleolar ribonucleoprotein complex | 3 | 3.49 | 1.1E-03 | NOP1, GAR1, LSM4 | 9 | 30 | 4595 | 51.1 | 3.2E-02 | 1.1E-02 | 9.5E-01 |
| GOTERM_CC_FAT | GO:0030529~ribonucleoprotein complex | 6 | 6.98 | 1.4E-03 | NOP1, SBP1, GAR1, RPL4A, LSM4, NPL3 | 9 | 590 | 4595 | 5.2 | 3.9E-02 | 9.9E-03 | 1.2E+00 |
| GOTERM_CC_FAT | GO:0043232~intracellular non-membrane-bounded organelle | 7 | 8.14 | 3.7E-03 | NOP1, SBP1, GAR1, RPL4A, LSM4, NSR1, NPL3 | 9 | 1120 | 4595 | 3.2 | 1.0E-01 | 2.1E-02 | 3.0E+00 |
| GOTERM_CC_FAT | GO:0043228~non-membrane-bounded organelle | 7 | 8.14 | 3.7E-03 | NOP1, SBP1, GAR1, RPL4A, LSM4, NSR1, NPL3 | 9 | 1120 | 4595 | 3.2 | 1.0E-01 | 2.1E-02 | 3.0E+00 |
| GOTERM_CC_FAT | GO:0070013~intracellular organelle lumen | 6 | 6.98 | 5.1E-03 | NOP1, SBP1, GAR1, LSM4, NSR1, NPL3 | 9 | 783 | 4595 | 3.9 | 1.4E-01 | 2.4E-02 | 4.2E+00 |
| GOTERM_CC_FAT | GO:0043233~organelle lumen | 6 | 6.98 | 5.1E-03 | NOP1, SBP1, GAR1, LSM4, NSR1, NPL3 | 9 | 783 | 4595 | 3.9 | 1.4E-01 | 2.4E-02 | 4.2E+00 |
| GOTERM_CC_FAT | GO:0031974~membrane-enclosed lumen | 6 | 6.98 | 6.5E-03 | NOP1, SBP1, GAR1, LSM4, NSR1, NPL3 | 9 | 827 | 4595 | 3.7 | 1.7E-01 | 2.7E-02 | 5.3E+00 |
| GOTERM_BP_FAT | GO:0010608~posttranscriptional regulation of gene expression | 4 | 4.65 | 7.9E-03 | SBP1, RPL4A, NAB2, DBP2 | 12 | 192 | 4870 | 8.5 | 6.6E-01 | 1.2E-01 | 8.9E+00 |
| GOTERM_MF_FAT | GO:0070717~poly-purine tract binding | 2 | 2.33 | 1.4E-02 | NAB2, NPL3 | 11 | 6 | 4190 | 127.0 | 4.6E-01 | 1.9E-01 | 1.2E+01 |
| GOTERM_MF_FAT | GO:0008143~poly(A) RNA binding | 2 | 2.33 | 1.4E-02 | NAB2, NPL3 | 11 | 6 | 4190 | 127.0 | 4.6E-01 | 1.9E-01 | 1.2E+01 |
| GOTERM_BP_FAT | GO:0016071~mRNA metabolic process | 4 | 4.65 | 2.0E-02 | NAB2, DBP2, LSM4, NPL3 | 12 | 270 | 4870 | 6.0 | 9.3E-01 | 2.6E-01 | 2.1E+01 |
| GOTERM_BP_FAT | GO:0000377~RNA splicing, via transesterification reactions with bulged adenosine as nucleophile | 3 | 3.49 | 2.1E-02 | LSM4, NPL3, DED1 | 12 | 102 | 4870 | 11.9 | 9.4E-01 | 2.5E-01 | 2.2E+01 |
| GOTERM_MF_FAT | GO:0042802~identical protein binding | 5 | 5.81 | 2.1E-02 | ELP4, GAR1, NAB2, DBP2, LSM4 | 11 | 487 | 4190 | 3.9 | 6.0E-01 | 2.1E-01 | 1.8E+01 |
| GOTERM_BP_FAT | GO:0000375~RNA splicing, via transesterification reactions | 3 | 3.49 | 2.4E-02 | LSM4, NPL3, DED1 | 12 | 109 | 4870 | 11.2 | 9.6E-01 | 2.6E-01 | 2.5E+01 |
| GOTERM_BP_FAT | GO:0009451~RNA modification | 3 | 3.49 | 2.8E-02 | ELP4, NOP1, GAR1 | 12 | 119 | 4870 | 10.2 | 9.8E-01 | 2.8E-01 | 2.8E+01 |
| GOTERM_BP_FAT | GO:0008380~RNA splicing | 3 | 3.49 | 3.9E-02 | LSM4, NPL3, DED1 | 12 | 142 | 4870 | 8.6 | 1.0E+00 | 3.4E-01 | 3.7E+01 |
| GOTERM_CC_FAT | GO:0000932~cytoplasmic mRNA processing body | 2 | 2.33 | 4.1E-02 | SBP1, LSM4 | 9 | 24 | 4595 | 42.5 | 7.0E-01 | 1.4E-01 | 3.0E+01 |
| GOTERM_BP_FAT | GO:0006417~regulation of translation | 3 | 3.49 | 6.0E-02 | SBP1, RPL4A, DBP2 | 12 | 180 | 4870 | 6.8 | 1.0E+00 | 4.5E-01 | 5.2E+01 |
| GOTERM_BP_FAT | GO:0032268~regulation of cellular protein metabolic process | 3 | 3.49 | 7.3E-02 | SBP1, RPL4A, DBP2 | 12 | 201 | 4870 | 6.1 | 1.0E+00 | 4.9E-01 | 5.9E+01 |
| GOTERM_BP_FAT | GO:0006397~mRNA processing | 3 | 3.49 | 7.5E-02 | NAB2, LSM4, NPL3 | 12 | 204 | 4870 | 6.0 | 1.0E+00 | 4.8E-01 | 6.0E+01 |
| GOTERM_MF_FAT | GO:0004004~ATP-dependent RNA helicase activity | 2 | 2.33 | 8.1E-02 | DBP2, DED1 | 11 | 35 | 4190 | 21.8 | 9.7E-01 | 5.1E-01 | 5.4E+01 |
| GOTERM_MF_FAT | GO:0008186~RNA-dependent ATPase activity | 2 | 2.33 | 8.1E-02 | DBP2, DED1 | 11 | 35 | 4190 | 21.8 | 9.7E-01 | 5.1E-01 | 5.4E+01 |
| GOTERM_MF_FAT | GO:0003724~RNA helicase activity | 2 | 2.33 | 9.6E-02 | DBP2, DED1 | 11 | 42 | 4190 | 18.1 | 9.9E-01 | 5.1E-01 | 6.1E+01 |

For proteins containing multiple methylation sites, these cluster in specific regions (Fig. 2.16) (Fink & Hamilton 2007). Including previously described sites, arginine methylation stretches across a wide region from R8 to R81 in Nop1. The methylated arginines were evenly spaced at every fourth residue for most of this sequence. In Nsr1, seven sites were localized in the sequence from R353 to R386 with the distance of four amino acids being apparent in the pairs R362 with R366, R375 with R379 and R382 with R386. In Npl3, the region containing 19 methylated arginines spanned from R284 to R404 and in Sbp1 four methylated residues appear at positions R125, R145, R161 and R165. Two proteins showed clustering of some of the methylation sites: In Dbp2 one site was found near the N-terminus (R43), while the four others were close to the C-terminus (R509, R512, R518, R525; protein length: 546). The second is Ded1, where the detected methylation sites were R51, R62 and R578 at a protein length of 604. For Psp2 the sites were distributed in two regions: 419, 425 and 538, 551, 575.

This clustering together with the strong enrichment of methylation on proteins involved in functions related to RNA metabolism posed the question whether the methylation sites themselves are frequently found in specific protein motifs or domains, especially RNA-binding domains. For ten of the 13 proteins at least a subset of their methylation sites was located in an RGG-box (defined here as at least two RGG sequences, each separated by a maximum of nine amino acids): Psp2, Ded1, Npl3, Lge1, Nop1, Gar1, Nsr1, Nab2, Sbp1, Dbp2. Nop1 contains one extended RGG-box (residues 8-83) and three methylation sites confirmed in this study (R70, R73 and R81) fell into this sequence. All seventeen residues annotated as probable di-methyl-arginines in UniProt were localized within this region. Similarly for Gar1 the methylation site found here was within an RGG-box and all probable di-methyl arginines are found in one of the two RGG-boxes. In Nsr1, R353, R379 and R382, identified as methylated, as well as probable di-methylation sites R362, R366, R375 and R382, are part of the RGG-box from 53 to 384.

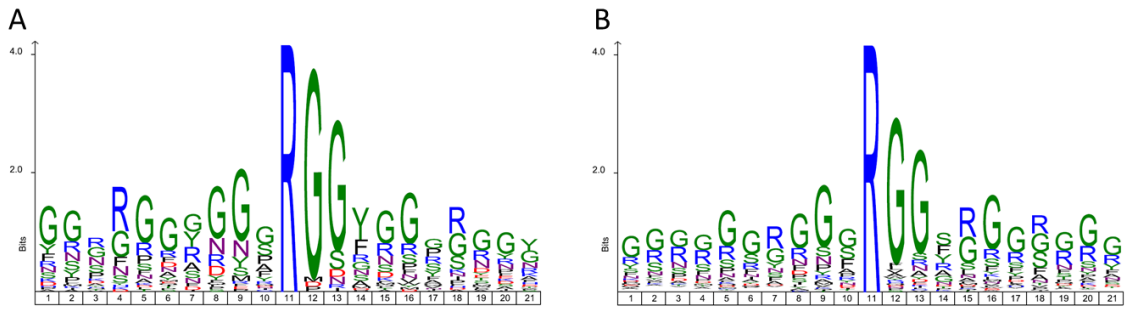



***Figure 2.14: Arginine methylation sequence motifs.*** *Consensus motifs for the sequence of ten amino acids upstream to ten downstream of the methylated arginine shown for methylation sites identified by peptide IP (**A**) and all known arginine methylation sites in S. cerevisiae (**B**). Sequence logos were generated using IceLogo against the background of the S. cerevisiae proteome listed in SwissProt.* (Plank et al. 2015)

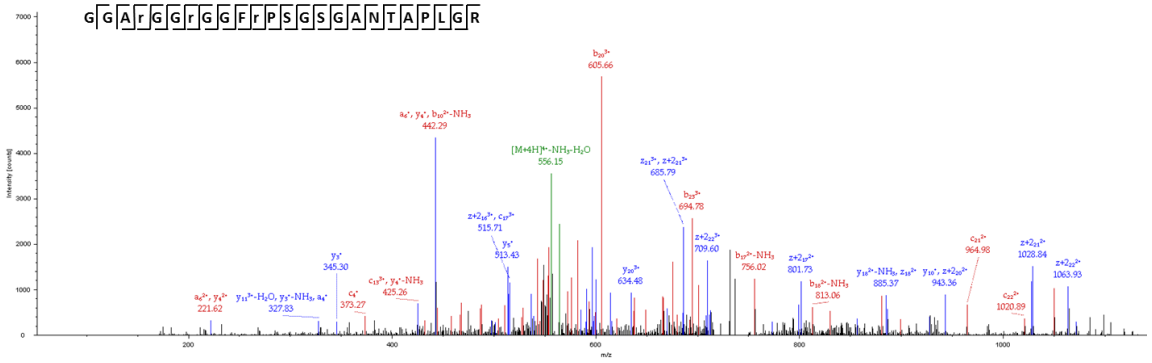



***Figure 2.15: Combination of CID and ETD MS/MS spectra demonstrating protein arginine methylation of peptide*** $_{376}$***GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR***$_{399}$. *Mono-methylated arginines are indicated by lowercase 'r'.* (Plank et al. 2015)

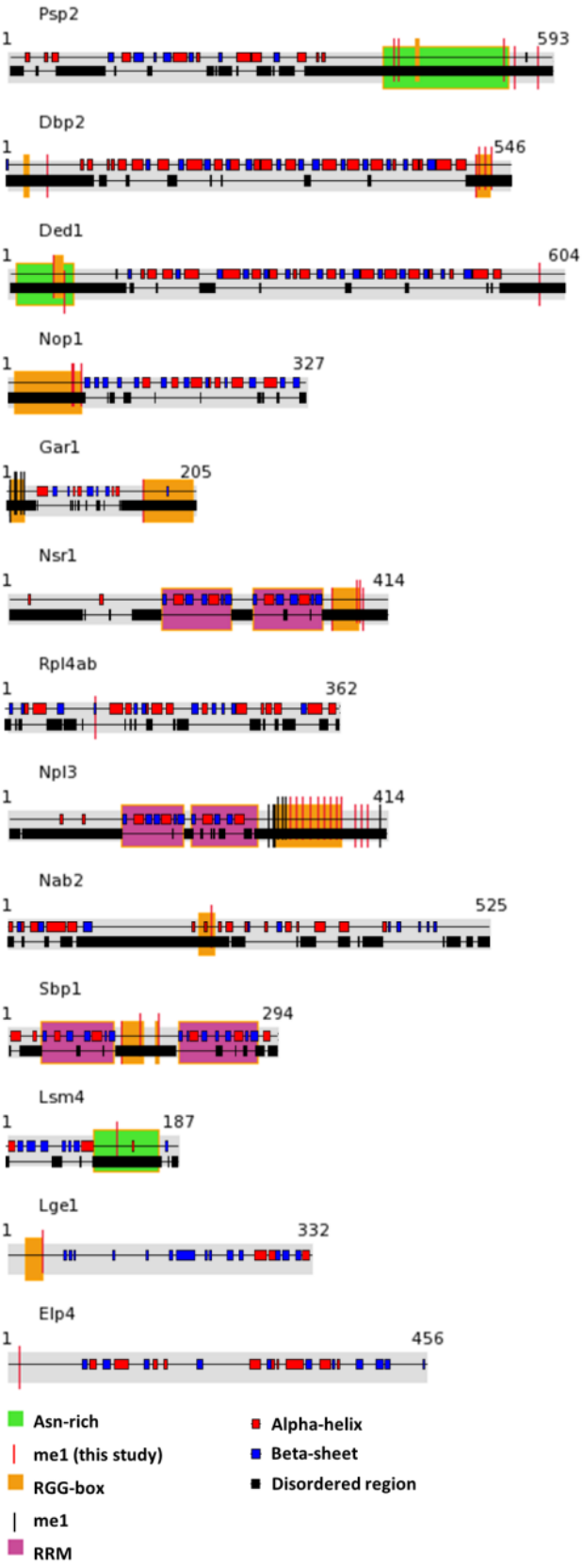



***Figure 2.16: Schematic view of the domain organisation and secondary structures of arginine methylated proteins identified by peptide IP.*** *RGG-boxes, Asn-rich motifs, and RRM domains are shown as boxes as indicated in the legend, and unambiguous arginine mono-methylation sites as vertical lines in red when found in this study and black when described previously (see also Table 2.5). An RGG box was defined as at least two RGG sequences, each separated by a maximum of nine amino acids. In the foreground secondary structure elements are displayed on the top row with alpha-helices as red and beta sheets as blue boxes. Disordered regions are shown as black boxes on the bottom row. This information is not available for Lge1 and Elp4. Annotated methylation sites from this and previous studies are at positions: Psp2: 419, 425, 538, 551, 575; Dbp2: 43, 509, 512, 518, 525; Ded1: 51, 62, 578; Nop1: 70, 73, 81; Gar1: 4, 8, 11, 15, 19, 147; Nsr1: 353, 379, 382, 386; Rpl4ab: 95; Npl3: 284, 288, 290, 294, 298, 302, 307, 314, 321, 329, 337, 344, 351, 358, 363, 377, 384, 391, 404; Nab2: 222; Sbp1: 125, 145, 165; Lsm4: 119; Lge1: 39; Elp4: 13. The protein domains were generated using the drawing tool available at http://domaindraw.imb.uq.edu.au/ and secondary structure depictions are from http://www.yeastrc.org/ (Plank et al. 2015).*

In Nab2 the only unambiguously identified site R222 was part of this protein`s RNA-binding RGG-box (209-228; annotated by similarity in UniProt) as are R125 and R145 (RGG-box: 125-151) and R161 and R165 (RGG-box: 161-167) in Sbp1. For Dbp2 four out of the five identified sites (R509, R512, R518, R525) were located in the RNA-binding RGG-box. The remaining site is distant, closer to the N-terminus (R43). Eighteen of the known arginine methylation sites in Npl3, including all observed in this study (R284, R288, R290, R294, R298, R307, R314, R321, R329, R337, R344, R358, R363, R377, R384, R302, R351, R391) were within the glycine-arginine-rich region from position 280 to 398, which contains several RGG sequences. Only one site, located directly C-terminal of this region, was found on an arginine not followed by glycine.

Additionally, in Psp2, all five identified sites are localized within (R419, R425, R538) or in close vicinity (R551, R575) to the region designated as asparagine-rich in UniProt. Notably, only one of these sites is in an RGG context, while in all other cases the arginine is followed by only one glycine. In Ded1 two sites (R61, R62) and the only site identified in Lsm4 (R119) are also located in an asparagine-rich region (Fig. 2.16). As no overlap between methylation sites and annotated, structured protein domains could be discovered, the localisation of the sites was next compared with secondary structure predictions and it was found that, with the exception of R95 in Rpl4, all sites are located within extended unstructured regions (no secondary structure information was available for Lge1 and Elp4) (Fig. 2.16). Finally, Ded1 and Dbp2 both contain a DEAD box and Nop1, Npl3 and Sbp1 two RRM-domains each, but no methylation sites were found in these locations.

## 2.9. Discussion

Four experimental approaches were employed for the identification of protein methylation sites in *S. cerevisiae*. An antibody and the 3xMBT domain were used as affinity reagents for the enrichment of lysine methylated proteins and IP on peptide level was performed for the discovery of arginine methylation sites. HILIC fractionation without prior enrichment allowed the mass spectrometric detection of both lysine and arginine methylation sites.

### 2.9.1. Global Arginine Methylation Levels Assessed by Western Blotting

To obtain an overview of the prevalence of protein arginine methylation in *S. cerevisiae* compared to a human cell line in which extensive efforts for the discovery of protein methylation had been undertaken (Uhlmann et al. 2012; Geoghegan et al. 2015), western blotting was performed with different anti-methyl-arginine antibodies. The number of bands detected by western blotting with an RGG biased antibody (D5A12) is higher than when using the antibodies raised without bias to the sequence context of the methylated arginine (Me-R4-100 and ICP0801) (Fig. 2.2). Also note that the five bands detected with Me-R4-100 coincide in molecular weight with major bands on the D5A12 blot and therefore are likely to target the same proteins. This suggests that mono-methyl arginines in *S. cerevisiae* are mainly expected on RGG rich proteins, that other methylation sites only exist on proteins of lower abundance or in lower methylation stoichiometry or that the Me-R4-100 antibody performs less well in recognizing proteins with methylation sites in a different sequence context. The number of bands found is similar to previously reported results from western blotting using a different antibody and to autoradiography after [methyl-$^{3}$H]SAM labelling, even though the latter is expected to also detect methylation on amino acids other than arginine (Low et al. 2013; Xu et al. 2003). While the detected bands in this assay may be due to unspecific binding, the disappearance of these bands in a deletion strain of PRMT Hmt1 strongly indicates that they are indeed due to the detection of arginine methylation. In contrast, the

signal detected with antibodies for symmetrically and asymmetrically di-methylated arginines was unaltered in single gene deletion strains of all known PRMTs in *S. cerevisiae* (see "4.3.1. Global Differences in Methylation Levels in Methyltransferase Deletion Strains"). It is therefore not clear if these signals truly originate from the detection of methylation sites.

Western blotting also showed that the absolute numbers of arginine methylated proteins in *S. cerevisiae* is expected to be lower than in humans. Notably, the number of bands observed on Jurkat lysate is much lower than the number of known arginine methylation sites identified in this cell line by mass spectrometry (Table 2.1). Together these findings indicate the limited use of western blotting for assessing the complexity of the arginine methylation landscape.

### 2.9.2. MethylSILAC

To overcome one of the main challenges associated with the identification of methylation sites by LC-MS/MS, i.e. the misassignment of peptide sequences due to amino acid pairs with a mass difference isobaric to the methyl-group, the MethylSILAC strategy described previously was used (Ong et al. 2004). In this method the presence of a heavy-light pair serves as additional evidence for the presence of methylation.

In all experiments reported here, an approach was chosen in which the light and heavy sample are mixed, as is the case in most publications reporting the use of MethylSILAC to date (Uhlmann et al. 2012; Bremang et al. 2013; Ong et al. 2004), as opposed to analysing the heavy and light sample individually. Mixing the samples has the advantages of reduced instrument time, simpler data analysis and, as the samples are digested in the same tube, more similar peptide characteristics (e.g. missed cleavages and methionine oxidation) can be expected. If cultures are mixed, the protein extraction is also expected to be more similar. A disadvantage of mixing labelled samples is that all methylated peptides will be halved in their concentration compared to unmixed samples, while the concentrations of all other peptides (except if containing methionine) are not reduced. Not mixing samples might therefore be a

viable option if the LC is highly reproducible so that runs can be aligned *in silico*. The more complex the sample processing is (e.g. involving pull-downs, etc.) the higher is the necessity for sample mixing to obtain similar intensities of both partners of a MethylSILAC pair compared to the background.

The MethylQuant software in combination with custom built pre- and post-processing algorithms was used to determine identifications of methylated peptides that originated from feature pairs with a similar intensity of the light and heavy partner.

### 2.9.3. Influence of Growth Medium on Protein Methylation

To test if the use of a synthetic growth medium for MethylSILAC labelling would lead to a reduction of methylation levels compared to growth in YPD, western blotting with a panel of anti-methyl antibodies was employed. While there was a clear reduction of arginine mono-methylation levels upon growth in SD medium, no apparent change was observed when using SC medium. Therefore, the latter was the medium of choice for subsequent experiments for the identification of protein methylation sites.

During the time of this writing, further changes to growth conditions were reported that can have an impact on global methylation levels. In particular, stationary phase instead of log-phase growth was found by amino acid analysis to dramatically increase the levels of symmetric di-methyl arginine (Lakowski et al. 2015). It will be intriguing to test if this finding will be reflected in identifications at peptide level.

### 2.9.4. Protein Lysine Methylation Assessed by Protein Immunoaffinity Purification

IP with modification-specific antibodies is a standard method for the large scale identification of PTMs. In a previous study on human cells, five anti-methyl lysine and six anti-methyl arginine antibodies had been employed in the discovery of protein methylation sites in a human cell line, which together led to the identification of 308 methylation sites on 99

proteins (Bremang et al. 2013). With respect to the number of identified sites, this approach outperformed identification strategies based on in-gel and off-gel fractionation of proteins without prior enrichment in the same study.

A similar experiment for the discovery of lysine methylation sites in yeast with an antibody specified to target pan-methyl lysine, i.e. recognize mono-, di- and tri-methylated lysine, was performed here. A pull-down in which the antibody was omitted served as a negative control. As lysine methylation has been found to be reversed by demethylases (Shi et al. 2004, p.1), broad-specificity demethylase inhibitors were included in one sample to test if this leads to a higher number of identified methylation sites by protecting methyl-groups from demethylation during sample preparation.

By mass spectrometry 116 and 141 proteins respectively from pull-downs on samples with and without demethylase inhibitors added were identified, fourteen of which were found in both samples, but not the negative control. One lysine methylated peptide (STTTGHLIYK(me1)) was identified, which is considered a high confidence identification as it was found in both samples with highly significant assignments to the fragment spectra (Fig. 2.5). The peptide originates from Elongation factor 1-alpha and was also found by other experimental approaches (see “2.7. Enrichment via protein domain interaction” and “2.8. Sample fractionation by HILIC”). As no further methylation sites were identified, it is not possible to conclude from these results if the addition of demethylase inhibitors to lysates is beneficial for the purification of lysine methylated proteins.

It is possible that further methylation sites that could not be identified via MS/MS analysis are located on proteins immunoprecipitated with this approach, especially on the fourteen proteins that were found in both pull-down samples, but not in the negative control. The only study reporting the identification of methylation sites in *S. cerevisiae* by protein-IP to date identified 90 proteins after pull-down with an anti-methyl arginine antibody (Low et

al. 2013). However, it did not report the assignment of methylation sites directly from this pull-down.

### 2.9.5. Enrichment of Lysine Methylated Proteins via Protein Domain Interaction

Protein domains that naturally function to bind post-translationally modified amino acids can provide an alternative affinity reagent to antibodies. The three malignant brain tumour domain repeats (3xMBT) of the L3MBTL1 protein had been chosen as a tool for the detection of lysine mono- and di-methylation in a recent study as this domain had been found to be highly specific, while not strongly dependent on sequence context (Moore et al. 2013; Nady et al. 2012). While there are no examples of MBT-domain containing proteins in *S. cerevisiae*, it can be argued that, if this domain recognizes lysine methylation independently of sequence context it may detect methylation sites on proteins regardless of the species of origin.

The strategy for identifying proteins mono- and di-methylated on lysine presented by Moore (2013) is based on determining quantitative differences between protein levels in a pull down with the 3xMBT construct and a mutated version, 3xMBT-D355, which shows reduced binding to methylated proteins. A disadvantage of this strategy is that interaction partners of methylated proteins cannot be distinguished from proteins directly binding to the construct. Additionally, as the priority here was to identify methylation sites rather than methylated proteins, it was chosen not to utilize an analysis strategy based on quantitative differences. Instead 3xMBT was used for the enrichment of methylated proteins and mapping of methylation sites was attempted by bottom up proteomics. The disadvantage of this method is that, unlike for protein identification, it is not sufficient to identify the peptides of a protein which are most suitable for LC-MS/MS detection, but that the peptide carrying the methylation site needs to be identified.

A comparison of eluates from pull-downs with 3xMBT on yeast lysate to a control in which no lysate had been applied indicated that only one band visible by silver-staining did not

originate from the bacterial sample from which the construct had been purified (Fig. 2.6). This band was absent in the pull-down with 3xMBT-D355N, indicating methylation-dependent binding. Further proteins may have been enriched, but may not be visible by silver staining. In contrast, the eluates from HEK293T nuclear lysate yielded a high number of bands with both 3xMBT and 3xMBT-D355N that were absent when omitting the lysate. The complexity of eluates using 3xMBT and 3xMBT-D355N was similar, albeit the protein abundance appeared lower in the 3xMBT-D355N sample. As most of the bands were absent in a sample in which GST alone had been used instead of the affinity reagent, these proteins bind 3xMBT specifically. Their presence in the 3xMBT-D355N sample may be due to residual binding of methylated proteins to this construct or to methylation-independent interactions with the MBT domains.

**Table 2.9:** Methylated proteins identified in *S. cerevisiae* by 3xMBT domain based enrichment and their functions according to UniProt.

| Protein ID | Protein Name | Function (UniProt) |
|---|---|---|
| P0CX27 | Rpl42A | Component of the large ribosomal subunit |
| Q12420 | Snu66 | Component of the U4/U6.U5 tri-snRNP particle, one of the building blocks of the spliceosome. Required for pre-mRNA splicing. |
| P02994 | EF1A | GTP-binding component of the eukaryotic elongation factor 1 complex (eEF1). In its active GTP-bound form, binds to and delivers aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis. In the presence of a correct codon-anticodon match between the aminoacyl-tRNA and the A-site codon of the ribosome-bound mRNA, the ribosome acts as a GTPase activator and the GTP is hydrolyzed. The inactive GDP-bound form leaves the ribosome and must be recycled by its guanine nucleotide exchange factor (GEF) (eEF1B subcomplex) before binding another molecule of aminoacyl-tRNA. Required for nuclear export of aminoacyl-tRNAs. May also be involved in translational quality control by targeting cotranslationally damaged proteins to the proteasome. Also exhibits actin filament-binding and -bundling activities and is involved in cytoskeleton organization. |
| P32324 | EF2 | Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome. |
| P46677 | Taf1 | Functions as a component of the DNA-binding general transcription factor complex TFIID. Binding of TFIID to a promoter (with or without TATA element) is the initial step in pre-initiation complex (PIC) formation. TFIID plays a key role in the regulation of gene expression by RNA polymerase II through different activities such as transcription activator interaction, core promoter recognition and selectivity, TFIIA and TFIIB interaction, chromatin modification (histone acetylation by TAF1), facilitation of DNA opening and initiation of transcription. |

Gel lanes were cut into slices and subjected to in-gel digestion and LC-MS/MS which allowed to map seven methylation sites onto five yeast proteins (Table 2.3 a). Lysine methylation on

three of these proteins had been reported previously (Webb et al. 2008; Couttas et al. 2012; Cavallius et al. 1993).

As shown in table 2.9, one methylated protein is a component of the large ribosomal subunit and two are translation elongation factors involved in the delivery of aminoacyl-tRNA to the ribosome and the translocation step. Interestingly, protein methylation on various residues had been described on translation factors in both prokaryotes and eukaryotes and 16 of the 23 *S. cerevisiae* protein methyltransferases with known function catalyse the methylation of proteins involved in translation, suggesting an important role of the modification in this context (Polevoda & Sherman 2007; Couttas et al. 2012; Clarke 2013). For example, EF-Tu, the bacterial homologue of EF1A is lysine-methylated in E. coli (L'Italien & Laursen 1979). Elongation factor 1-alpha 1 was also among the eight proteins that were found lysine methylated in HEK293T cells with this method (Table 2.3 b). The remaining two methylated proteins from yeast are a component of the spliceosome (Snu66, Q12420) and a member of the general transcription factor complex TFIID (TAF1, P46677).

The original publication of 3xMBT as an affinity reagent reports 313 proteins enriched above 2-fold in the 3xMBT vs. 3xMBT-D355N pull-down from nucleoplasmic extract of 293T cells (Moore et al. 2013). However, in an attempt to identify methylation sites using the MethylSILAC approach, only 26 were detected. In the experiment described here, 1003 proteins were detected in the eluate from the pull-down on HEK293T nuclear lysate. However, as it was not attempted to identify methylated proteins by a quantitative comparison of proteins enriched in the 3xMBT vs. 3xMBT-D355N, it is not clear how many interact specifically with the methyl binding site of 3xMBT.

Together, these results indicate that, while the 3xMBT-construct provides a useful tool for the discovery of lysine mono- and di-methylated proteins, mapping of methylation sites constitutes a severe additional challenge beyond the identification of methylated proteins.

### 2.9.6. Identification of Methylation Sites via Sample Fractionation by HILIC

To avoid the bias introduced by an affinity reagent, LC-MS/MS based discovery of methylation sites was also performed after peptide fractionation by HILIC without previous enrichment. In an earlier study on human samples, HILIC had been shown to outperform other types of separation in terms of the number of methylation sites identified (Uhlmann et al. 2012).

Considering that the *S. cerevisiae* genome contains about 6,700 protein coding genes (uniprot.org, organism ID: 559292, 06/2015), the achieved identification of 5,553 proteins with a median sequence coverage of 44% represents a considerable portion of the proteome. While a further increase in sample pre-fractionation might lead to still higher sequence coverages, a theoretical upper bound is imposed by the length of peptides generated by tryptic cleavage and their suitability for LC-MS/MS. This limitation could be addressed by combining different proteases. The high number of identified proteins together with good sequence coverage results in a good chance of covering many of the potentially methylated sites in the proteome. The probability of an identification of a methylation event may however be lower as methylations may be sub-stoichiometric and the database matching of a methylated peptide is more error-prone than of the average unmodified peptide. The latter is for reasons detailed in ‘2.1.2. Challenges in Methylation Assignment’.

By offline HILIC fractionation and LC-MS/MS analysis employing the MethylQuant strategy eleven lysine and four arginine methylation sites on seven and four proteins respectively were detected.

The same translation elongation factors (EF1A and EF2) and ribosomal protein (Rpl42) were identified among the lysine methylated proteins as via 3xMBT-based enrichment. Additionally, two metabolic enzymes of glycolysis and alcoholic fermentation were found lysine methylated, as well as a V-type ATPase and histone H3. The arginine methylated

proteins included two proteins involved in RNA processing (Npl3 and Nsr1), as well as a guanine nucleotide exchange factor (Lte1) and a protein involved in the control of nuclear migration (Num1). See table 2.10 for an extended functional description of these proteins. All methylation sites on EF1A found in this experiment had been described previously in the methylation states found here, except for K316, which had been found in the di- and tri-methylated, but not mono-methylated version (Cavallius et al. 1993; Couttas et al. 2012; Wang et al. 2015). Similarly, the methylation of 60S ribosomal protein L42 at position 40 had been known (Webb et al. 2008). The only methylation site detected on histones was on lysine 80 (counting from the initiator methionine), which was found mono-, di- and tri-methylated, as had been reported previously (Ng et al. 2002; van Leeuwen et al. 2002; Lacoste et al. 2002). EF2 had been found to be tri-methylated on lysine 15 and di-methylated on lysine 509, while the di-methylation of lysine 613, which was detected in this experiment had not been reported (Couttas et al. 2012). Methylation of Glyceraldehyde-3-phosphate dehydrogenase, Pyruvate decarboxylase and V-type proton ATPase catalytic subunit A had not been reported previously.

**Table 2.10:** Functions of methylated proteins detected by HILIC fractionation.

a, Lysine methylated proteins found by HILIC-fractionation and functions according to UniProt

| Protein Accession | Protein Name | Function (UniProt) |
|---|---|---|
| P02994.1 | EF1A | GTP-binding component of the eukaryotic elongation factor 1 complex (eEF1). In its active GTP-bound form, binds to and delivers aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis. In the presence of a correct codon-anticodon match between the aminoacyl-tRNA and the A-site codon of the ribosome-bound mRNA, the ribosome acts as a GTPase activator and the GTP is hydrolyzed. The inactive GDP-bound form leaves the ribosome and must be recycled by its guanine nucleotide exchange factor (GEF) (eEF1B subcomplex) before binding another molecule of aminoacyl-tRNA. Required for nuclear export of aminoacyl-tRNAs. May also be involved in translational quality control by targeting cotranslationally damaged proteins to the proteasome. Also exhibits actin filament-binding and -bundling activities and is involved in cytoskeleton organization. |
| P32324.1 | EF2 | Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome. |
| P00360.3; P00358.3; P00359.3 | G3P1; G3P2; G3P3 | Catalytic activity: D-glyceraldehyde 3-phosphate + phosphate + NAD+ => 3-phospho-D-glyceroyl phosphate + NADH |
| P61830.2 | H3 | Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. DNA accessibility is regulated via a complex set of post-translational modifications of histones, also called histone code, and nucleosome remodeling. Component of the UAF (upstream activation factor) complex which interacts with the upstream element of the RNA polymerase I promoter and forms a stable preinitiation complex. Together with SPT15/TBP, UAF seems to stimulate basal transcription to a fully activated level. |
| P06169.7 | Pdc1 | Major of three pyruvate decarboxylases (PDC1, PDC5, PDC6) implicated in the nonoxidative conversion of pyruvate to acetaldehyde and carbon dioxide during alcoholic fermentation. Most of the produced acetaldehyde is subsequently reduced to ethanol, but some is required for cytosolic acetyl-CoA production for biosynthetic pathways. The enzyme is also one of five 2-oxo acid decarboxylases (PDC1, PDC5, PDC6, ARO10, and THI3) able to decarboxylate more complex 2-oxo acids (alpha-ketoacids) than pyruvate, which seem mainly involved in amino acid catabolism. Here the enzyme catalyzes the decarboxylation of amino acids, which, in a first step, have been transaminated to the corresponding 2-oxo acids. In a third step, the resulting aldehydes are reduced to alcohols, collectively referred to as fusel oils or alcohols. Its preferred substrates are the transaminated amino acids valine, isoleucine, phenylalanine, and tryptophan, whereas leucine is no substrate. In a side-reaction the carbanionic intermediate (or active aldehyde) generated by decarboxylation or by activation of an aldehyde can react with an aldehyde via condensation (or carboligation) yielding a 2-hydroxy ketone, collectively called acyloins. |
| P0CX28.1; P0CX27.1 | Rpl42A; Rpl42B | Component of the large ribosomal subunit |
| P17255.3 | VatA | Catalytic subunit of the peripheral V1 complex of vacuolar ATPase. V-ATPase (vacuolar ATPase) is responsible for acidifying a variety of intracellular compartments in eukaryotic cells. It is an electrogenic proton pump that generates a proton motive force of 180 mV, inside positive and acidic, in the vacuolar membrane vesicles. It may participate in maintenance of cytoplasmic $Ca^{2+}$ homeostasis. This is a catalytic subunit. |

b, Arginine methylated proteins found by HILIC-fractionation and functions according to UniProt

| Protein Accession | Protein Name | Function (UniProt) |
|---|---|---|
| P07866.2 | Lte1 | GDP-GTP exchange factor for TEM1, a Ras-like protein, component of the mitotic exit network (MEN). Activation of TEM1 by LTE1 in the bud ultimately leads to activation of CDC15 followed by the release of CDC14 from the nucleolus, which then inactivates cyclin-dependent kinases (CDKs) activity by several mechanism. Required for TEM1 localization to the bud cortex during mitotic exit. Fine-tunes the timing of the mitotic exit and couples this event with cytokinesis.Involved in proprotein-processing like proalpha factor-processing in the secretory pathway. |
| Q01560.1 | Npl3 | Required for pre-rRNA processing and nuclear import as well as mitochondrial protein targeting. Binds to poly(A). |
| P27476.1 | Nsr1 | Involved in pre-rRNA processing. Specifically binds nuclear localization sequences. Candidate for a receptor at the nucleus that may be involved in both RNA and protein transport. Binds telomeric sequences of the type (TG[1-3])n in vitro. |
| Q00402.2 | Num1 | Controls nuclear migration. NUM1 specifically controls the interaction of the bud neck cytoskeleton with the pre-divisional G2 nucleus. Functions in dynein-anchoring. During late anaphase forms dynein-interacting cortical microtubule capture sites at both cellular poles. This leads to dynein-dependent sliding of the microtubules in the bud. |

### 2.9.7. Identification of Arginine Methylation Sites via Peptide Immunoaffinity Purification

Finally, peptide immunoaffinity enrichment was used to reveal 41 arginine mono-methylation sites on 14 proteins, 7 of which have not yet been described. Two of the proteins not previously reported to be methylated (Psp2 and Dbp2) contained 5 unambiguously localized methylation sites each, which showed clustering in an Asn-rich region or RGG-box respectively. The remaining five proteins without previously identified methylation sites contained one or two methylation sites each. The obtained results strengthen the association between arginine methylation and RNA processing. They indicate that the previously observed prevalence of arginine methylation sites in RG-rich sequences extends to *S. cerevisiae*, however methylation sites in different sequence contexts are also presented.

The same experiment was performed with antibodies for di-methylated arginines (symmetric and asymmetric; antibodies: CST: 13222 and 13522), but no methylation sites could be discovered with the same criteria. The higher success of the peptide immunoaffinity purification approach for mono- as compared to di-methylated peptides is consistent with a previous study and may potentially be due to a higher quality of available antibodies targeting

the former modification (Guo et al. 2014). While the fact that not all previously known sites were found in this study indicates the restriction of the antibody based approach, the obtained data provide a considerable increase in the number of unambiguous mono-methylation sites.

An independent study employing peptide immunoprecipitation with a different anti-mono-methyl arginine antibody (Immunechem: ICP0801) was published subsequently and reported the identification of five methylated proteins. Four of these proteins overlap with proteins found methylated here (see '2.9.9. Proteome-wide Discovery Studies of Protein Methylation in *S. cerevisiae*').

#### *2.9.7.1. Arginine Methylated Proteins are Involved in mRNA Processing*

Arginine methylated proteins detected in this experiment and their functions according to UniProt are shown in table 2.11. Several of the proteins found to be methylated are involved in mRNA-related processes.

**Table 2.11:** Functions of arginine methylated proteins identified via peptide immuno-affinity purification according to UniProt.

| Accession | Protein Name | Function (UniProt) |
|---|---|---|
| P50109.2 | Psp2 | DNA polymerase alpha mutation suppressor. Suppressor of group II intron splicing defects of a mutation in MRS2. May play a role in mitochondrial mRNA splicing. |
| P24783.1 | Dbp2 | ATP-dependent RNA helicase involved nonsense-mediated mRNA decay and ribosome biogenesis through rRNA processing. |
| P06634.2 | Ded1 | ATP-binding RNA helicase involved in translation initiation. Remodels RNA in response to ADP and ATP concentrations by facilitating disruption, but also formation of RNA duplexes. Has weak ATP-dependent affinity for dsRNA, but strong ATP-dependent affinity for ssRNA. Acts as a virus host factor involved in the replication of the MBV and the L-A viruses by promoting the negative-strand RNA synthesis. May be involved in recognition of the preinitiation complex and DNA binding of the RNA polymerase III and play a role in mRNA splicing. |
| P15646.1 | Nop1 | S-adenosyl-L-methionine-dependent methyltransferase that has the ability to methylate both RNAs and proteins. Involved in pre-rRNA processing by catalyzing the site-specific 2'-hydroxyl methylation of ribose moieties in pre-ribosomal RNA. Site specificity is provided by a guide RNA that base pairs with the substrate. Methylation occurs at a characteristic distance from the sequence involved in base pairing with the guide RNA. Involved in the biogenesis of the 18S rRNA. Also acts as a protein methyltransferase by mediating methylation of 'Gln-105' of histone H2A (H2AQ105me), a modification that impairs binding of the FACT complex and is specifically present at 35S ribosomal DNA locus. |
| P28007.1 | Gar1 | Required for ribosome biogenesis. Part of a complex which catalyzes pseudouridylation of rRNA. This involves the isomerization of uridine such that the ribose is subsequently attached to C5, instead of the normal N1. Pseudouridine ("psi") residues may serve to stabilize the conformation of rRNAs. Essential for growth. |
| P27476.1 | Nsr1 | Involved in pre-rRNA processing. Specifically binds nuclear localization sequences. Candidate for a receptor at the nucleus that may be involved in both RNA and protein transport. Binds telomeric sequences of the type (TG[1-3])n in vitro. |
| P49626.2 / P10664.4 | Rpl4B/Rpl4A | Participates in the regulation of the accumulation of its own mRNA. |
| Q01560.1 | Nop3 / Npl3 | Required for pre-rRNA processing and nuclear import as well as mitochondrial protein targeting. Binds to poly(A). |
| P32505.1 | Nab2 | This essential protein binds to polyadenylated RNA and single-stranded DNA. It may be involved not only in RNA processing but also in transcription regulation. Believed to associate directly with nascent RNA polymerase II transcripts and remain associated during subsequent nuclear RNA processing reactions. |

| | | |
|---|---|---|
| P10080.2 | Ssbp1 / Sbp1 | Functions in the transition of mRNAs from translation to an mRNP complex destined for decapping. High-copy-number suppressor of decapping defects. Overexpression suppresses decapping defects in both DCP1-2 and DCP2-7 mutations. Acts to promote translational repression of mRNA in conjunction with DHH1 and subsequent mRNA localization to P bodies. Promotes translational repression of mRNA during glucose deprivation. |
| P40070.1 | Lsm4 | Component of LSm protein complexes, which are involved in RNA processing and may function in a chaperone-like manner. Component of the cytoplasmic LSM1-LSM7 complex which is involved in mRNA degradation by activating the decapping step. Component of the nuclear LSM2-LSM8 complex, which is involved in splicing of nuclear mRNAs. LSM2-LSM8 associates with multiple spliceosome snRNP complexes containing the U6 snRNA (U4/U6 snRNP, U4/U6.U5 snRNP, and free U6 snRNP). It binds directly to the U6 snRNA and plays a role in the biogenesis and stability of the U6 snRNP and U4/U6 snRNP complexes. It probably also is involved degradation of nuclear pre-mRNA by targeting them for decapping. LSM4 binds specifically to the 3'-terminal U-tract of U6 snRNA. LSM2-LSM8 probably is involved in processing of pre-tRNAs, pre-rRNAs and U3 snoRNA. LSM4, probably in a complex that contains LSM2-LSM7 but not LSM1 or LSM8, associates with the precursor of the RNA component of RNase P (pre-P RNA) and may be involved in maturing pre-P RNA. LSM4 is required for processing of pre-tRNAs, pre-rRNAs and U3 snoRNA. |
| Q02796.1 | Lge1 | Required to activate transcription of PDR3, a gene involved in retrograde regulation of multidrug resistance, a phenomenon that takes place in cells that have lost their mitochondrial genome. Also required for ubiquitination of histone H2B to form H2BK123ub1. H2BK123ub1 gives a specific tag for epigenetic transcriptional activation, telomeric silencing, and is also a prerequisite for H3K4me and H3K79me formation. Its precise role in H2BK123ub1 formation however is unclear and is independent of retrograde regulation of multidrug resistance function. |
| Q02884.1 | Elp4 | Acts as component of the RNA polymerase II elongator complex, which is a major histone acetyltransferase component of the RNA polymerase II (RNAPII) holoenzyme and is involved in transcriptional elongation. Association with elongating RNAPII requires a hyperphosphorylated state of the RNAPII C-terminal domain (CTD). Elongator binds to both naked and nucleosomal DNA, can acetylate both core and nucleosomal histones, and is involved in chromatin remodeling. It acetylates histones H3, preferentially at 'Lys-14', and H4, preferentially at 'Lys-8'. It functions as a gamma-toxin target (TOT); disruption of the complex confers resistance to Kluyveromyces lactis toxin zymocin (pGKL1 killer toxin). May also be involved in sensitiviy to Pichia inositovora toxin. May be involved in tRNA modification. ELP4 is required for the complex integrity and the complex HAT activity but is not required for the association of the complex with nascent RNA transcript. Is required for an early step in synthesis of 5-methoxycarbonylmethyl (mcm5) and 5-carbamoylmethyl (ncm5) groups present on uridines at the wobble position in tRNA. |
| P06169.1 | Pdc1 | Major of three pyruvate decarboxylases (PDC1, PDC5, PDC6) implicated in the nonoxidative conversion of pyruvate to acetaldehyde and carbon dioxide during alcoholic fermentation. Most of the produced acetaldehyde is subsequently reduced to ethanol, but some is required for cytosolic acetyl-CoA production for biosynthetic pathways. The enzyme is also one of five 2-oxo acid decarboxylases (PDC1, PDC5, PDC6, ARO10, and THI3) able to decarboxylate more complex 2-oxo acids (alpha-ketoacids) than pyruvate, which seem mainly involved in amino acid catabolism. Here the enzyme catalyzes the decarboxylation of amino acids, which, in a first step, have been transaminated to the corresponding 2-oxo acids. In a third step, the resulting aldehydes are reduced to alcohols, collectively referred to as fusel oils or alcohols. Its preferred substrates are the transaminated amino acids valine, isoleucine, phenylalanine, and tryptophan, whereas leucine is no substrate. In a side-reaction the carbanionic intermediate (or active aldehyde) generated by decarboxylation or by activation of an aldehyde can react with an aldehyde via condensation (or carboligation) yielding a 2-hydroxy ketone, collectively called acyloins. |

For the hnRNP protein, Npl3, twelve methylation sites were found in this study and six additional sites had been previously identified by mass spectrometry on the purified protein (Hart-Smith et al. 2012). A stretch of about 200 amino acids towards the C-terminus of this protein are rich in the RGG-motif and can be interpreted as consecutive and (partially) overlapping RGG-boxes. A close spacing of arginine residues is unfavourable for mass spectrometric detection as it leads to short peptides upon tryptic digestion. Potentially, the detection of the sites reported here was made possible by methyl-groups causing partial

miscleavages. The majority of the arginines in the glycine-arginine rich region are known methylation sites (Hart-Smith et al. 2012). Many tyrosine residues are interspersed between the RGG sequences, which is consistent with the finding that RGG-boxes are often rich in aromatic residues (Kiledjian & Dreyfuss 1992). Arginine methylation promotes the nuclear export of Npl3 (Shen et al. 1998), which is in contrast to findings in mammals, where arginine methylation of hnRNPs appears to facilitate nuclear localisation (Passos et al. 2006). In the cytoplasm the RGG domain of Npl3 binds to Sky1, a kinase targeting Npl3, in unmethylated, but not in methylated form (Lukasiewicz et al. 2007) which in turn leads to Npl3 nuclear import (Gilbert et al. 2001).

For Nab2, another hnRNP protein, methylation had been previously localized to the region of amino acids 180-256 (Green, 2002). The data obtained here confidently assign one methylation site to R222, while a second is either located on R226 or R229 or both, all of which are within an RGG-box. Nab2 is an essential protein, binding to poly-A RNA and single-stranded DNA. Both the RGG-box and $C_3H$-motifs further C-terminal have been found to bind RNA *in vitro* (Anderson et al. 1993, p.2). Npl3 and Nab2, as well as Yra1 which had been found to be methylated in previous studies (Yu et al. 2004) together play an important role in mRNP assembly and RNA export (reviewed by Tutucci & Stutz (2011)) and the role of RNA methylation in this process is reflected by the enrichment of GO-term 'RNA export from nucleus' for arginine methylated proteins found in this and previous studies.

Four methylation sites on the single-stranded nucleic acid binding protein Sbp1 that had been described after overexpressing this protein together with the major *S. cerevisiae* PRMT, Hmt1, in *E. coli* (Hsieh et al. 2007) were validated here. As this study is based on deducing methylation sites from MALDI MS1 spectra, the results presented here are experimentally complementary and confirm that the modification exists under wild type conditions in yeast. Sbp1 is an RNA binding protein and is involved in the formation of RNP complexes for RNA

de-capping. All known methylation sites on this protein lie within a glycine-arginine rich region and are followed by one or two glycines and the majority is located within RGG-boxes.

For another RGG-box containing protein, Dbp2, arginine methylation has not been described previously. All five methylation sites that were mapped to this protein lie within RGG sequences, one close to the N-terminus and the remaining clustering in the C-terminal RGG-box. Dbp2 is a double-stranded RNA binding protein of the DEAD-box family and an ATP-dependent RNA helicase. It is required for the formation of nuclear mRNPs which involves the binding of mRNA to Yra1, Nab2 and Mex67 (Erce et al. 2013). Dbp2 has also been linked to nonsense-mediated decay and rRNA processing (Bond et al. 2001).

Three methylation sites were mapped to Ded1, which like Dbp2 is a DEAD-box ATP-dependent RNA helicase and a paralogue of Dbp1 and which is required for translation initiation (Iost et al. 1999). Ded1 has been previously described to be methylated after co-expression with Hmt1 in *E. coli* (Erce et al. 2013). Here it was found that this modification also occurs in wild type budding yeast and that two sites previously reported as di-methylated also exist in the mono-methylated state. An additional previously reported site is R87.

The only ribosomal protein found to be arginine mono-methylated in this study, Rpl4A/Rpl4B, carries a methylation on R95. As the amino acid sequences of Rpl4A and Rpl4B are identical, both proteins are expected to be methylated. Other ribosomal proteins, Rps2, Rps3 and Rpl12A/Rpl12B, have been reported to be arginine methylated in *S. cerevisiae* previously (Lipson et al. 2010). The homologues of Rps2 are modified by the zinc-finger containing PRMTs Rmt3 and PRMT3 in *S. pombe* and human cells respectively. While there is no Rmt3 homologue in *S. cerevisiae*, this methylation is catalysed by Hmt1. The functional analogy suggests a biological function of this arginine methylation (Lipson et al. 2010). Rps3 is methylated by Sfm1 and Rpl12A/Rpl12B is modified on the N5-position of

R67 by Rmt2 (Young et al. 2012; Chern et al. 2002) indicating that a variety of methyltransferases act on ribosomal proteins.

#### *2.9.7.2. Arginine Methylation and rRNA Processing*

Three of the methylated proteins found in this study are involved in rRNA processing:

Nop1 is a homologue of fibrillarin, a mammalian RGG-box protein. It is a glutamine MTase and as part of the small subunit processome complex is involved in pre-18S rRNA processing (Schimmang et al. 1989; Tollervey et al. 1991). The three unambiguous methylation sites found on Nop1 had been previously reported by protein-IP with anti-methyl antibodies. In this, as in the current study the peptide GGSR(me1)GGFGGR was identified which can be mapped to three potential positions in the protein. As all three of these sites are within a glycine-arginine rich region, methylation of all three sites is likely.

A second methylated protein involved in 18S pre-rRNA processing is Gar1, a member of the H/ACA small nucleolar ribonucleoprotein (snoRNP) pseudouridylase complex. The methylation site identified in this study is one of 18 arginines in the two RGG-box regions of the protein annotated as probable di-methylation sites in UniProt. Here it was shown that this site can be found in mono-methylated form and it is suspected that most other sites were missed as they are expected to be located in very short peptides after tryptic digestion, even when assuming a missed cleavage at the methylation site. It has been suggested that methylation of Gar1 by Hmt1 is blocked by its association with RNA as *in vitro* methylation occurred to a higher extent in cell extracts treated with RNase than untreated ones (Frankel & Clarke 1999).

On a third protein involved in rRNA processing, Nsr1, which is structurally similar to mammalian nucleolin, four methylation sites that cluster in an RGG box were identified (Lee et al. 1992; Kondo & Inouye 1992). One of these sites is not in an RGG context, but instead followed by proline. A consensus motif in which methylated arginine precedes proline has

been reported (Uhlmann et al. 2012; Guo et al. 2014). The position of two further methyl-groups on a peptide with three arginines could not be unambiguously determined. Nsr1 binds nuclear localisation signals and plays a role in targeting proteins to the nucleolus (Lee et al. 1991; Yan & Mélèse 1993).

Four of these arginine methylated, RGG-box proteins, Sbp1, Npl3, Ded1, Nsr1 have been found to interact with translation initiation factor eIF4G, as has a sixth protein, Scd6, which contains a glycine-arginine rich region and constituted a marginal hit in this experiment. There is evidence that this interaction leads to inhibition of translation, e.g. upon an environmental stress, but may render the RNAs in a state poised for re-entry to translation (Hilliker et al. 2011; Rajyaguru & Parker 2012; Rajyaguru et al. 2012). It is tempting to speculate that this process is regulated by a PTM, for which arginine methylation seems a likely candidate. It is also noteworthy, that this would imply different roles for the RGG-box proteins during the life-time of the RNA.

#### *2.9.7.3. Further Sites of Arginine Methylation*

Psp2 had not previously been found to be methylated. Five methylation sites were confidently located on this protein, while for a sixth one the localisation is ambiguous. Psp2 contains an asparagine-rich region and all identified sites localize to this region or its close vicinity. Only one of these sites appears in an RGG context, while in all other cases the arginine is followed by only one glycine. Psp2 is a cytoplasmic protein that may play a role in mitochondrial mRNA splicing and is a high-copy suppressor of group II intron-splicing defects (Formosa & Nittis 1998; Waldherr et al. 1993).

One methylation site (R13) was found on Elp4 which had not previously been known to be methylated. This protein is a component of the elongator complex of RNA polymerase II transcription and required for the histone acetyltransferase activity of the complex, but not for

its association with the nascent RNA. Secondly, Elp4 has been linked to tRNA modification (Huang et al. 2005; Krogan & Greenblatt 2001; Winkler et al. 2002).

Little knowledge exists regarding the functions of Lge1 and methylation of this protein had not been previously reported. One methylation was mapped to the RGG sequence at R39 which is the third of three subsequent occurrences of this motif. Lge1 has been linked to transcriptional activation and H2BK123 ubiquitinylation, but its mechanism of action is unclear (Zhang et al. 2005, p.1; Hwang et al. 2003). It is also unknown if the protein forms interactions with DNA or RNA.

Arginine methylation of Sm proteins and Lsm4 (Like Sm) is well known in humans even though in the form of symmetric di-methylation (Brahms et al. 2000). This modification facilitates the interaction with the tudor-domain containing protein SMN (Brahms et al. 2001; Friesen et al. 2001). Arginine mono-methylation of an *S. cerevisiae* Lsm protein (Lsm4) is reported for the first time here. Lsm proteins form different complexes with distinct functions: the nuclear Lsm2-Lsm8 complex is part of the spliceosomal U6 snRNP and potentially also involved in the maturation of tRNA, snoRNA and rRNA while the cytoplasmic Lsm1-Lsm7 complex functions in mRNA decay (Kufel et al. 2002; Kufel et al. 2003; Beggs 2005; Salgado-Garrido et al. 1999; Bouveret et al. 2000). There is also evidence for a complex of Lsm2 to Lsm7, which may be involved in the maturation of pre-P RNA, the precursor of the RNA component of RNase P (Fernandez et al. 2004). While it is unclear if in *S. cerevisiae* the residue also exists in symmetrically di-methylated form, the homology to the human protein suggests that this modification is catalysed by a type II rather than type I MTase. This is noteworthy, as according to current knowledge the type I PRMT Hmt1 catalyses the vast majority of protein arginine methylation in *S. cerevisiae* (Gary et al. 1996). Also the sequence context is different from most of the methylation sites reported here, in that the methylated arginine is followed by a single glycine and no RGG motifs are found in its vicinity. In

humans, the Hmt1 homologue PRMT1 and PRMT3, which are both type I PRMTs, transfer methyl-groups to arginines in glycine-arginine rich (GAR) regions (Najbauer et al. 1993) (Tang et al. 1998), while the Hsl7 homologue PRMT5 has known target sites within and outside of GAR regions. As Sm proteins are among the targets of PRMT5, Hsl7 is a likely candidate for methylating Lsm4 (Boisvert et al. 2002; Boisvert et al. 2005).

Recently, it was reported that *Arabidopsis* PRMT5 (AtPRMT5) dissociates from chromatin where it normally methylates H4R3 upon salt stress and instead leads to increased levels of LSM4 symmetric di-methylation. This indicates that LSM4 methylation by AtPRMT5 is a regulated process (Z. Zhang et al. 2011). AtPRMT5 and its *Drosophila* homologue Dart5 have been implicated in pre-mRNA splicing (Deng et al. 2010; Sanchez et al. 2010). The C-terminal region of human Lsm4, in which arginine methylation occurs is RG-rich with three GRG, two GRGRGR sequences. In contrast, apart from the methylation site, no further RG or GR sequences are found in *S. cerevisiae* Lsm4, instead the methylation is located in a region with strong asparagine-bias (Fig. 4.1). This finding is striking in the light of Psp2 and Ded1 also being methylated in asparagine-rich regions. While Ded1 methylation appears in an RGG context, the methylated arginine in Lsm4 and several of the Psp2 methylation sites are followed by only one glycine. This poses the question whether the asparagine-rich sequence context might promote methylation of an otherwise weaker methylation sites or form a recognition site for a PRMT not dependent on the RGG motif or glycine-arginine rich regions.

#### *2.9.7.4. Methylation Sites are Located in RGG-boxes and Unstructured Regions*

The results of this study are in good agreement with previous findings since several previously known methylation sites were confirmed and most methylation sites found in an RGG-context. For ten of the identified proteins, Psp2, Lge1, Nop1, Gar1, Nsr1, Nab2, Sbp1, Dbp2, Ded1 and Npl3, some or all of the methylation sites are within an RGG-box, which is defined here as at least two RGG sequences, each separated by a maximum of nine amino

acids. Note however that the exact criteria for defining an RGG-box vary between studies (Kiledjian & Dreyfuss 1992; Burd & Dreyfuss 1994; Corley & Gready 2008). RGG-boxes are a highly conserved feature and have been found in several RNA-binding proteins (Burd & Dreyfuss 1994). Interactions of the RGG box with DNA and RNA have been described (Hanakahi et al. 1999; Mears & Rice 1996).

The functional importance of arginine methylation of an RGG-box has however been most clearly demonstrated in the context of protein–protein interaction: Sky1, a kinase involved in regulating proteins involved in mRNA transport, phosphorylates Npl3 (Siebel et al. 1999), which in turn promotes mRNA dissociation and re-import of Npl3 into the nucleus after RNA export (Gilbert et al. 2001). The RGG domain of Npl3 binds to Sky1 in unmethylated, but not in methylated form (Lukasiewicz et al. 2007).

The UniProt database lists eight entries as containing an RGG-box in *S. cerevisiae*. Five of these, Nop1, Gar1, Nsr1, Sbp1, Nab2 are known methylated proteins and this was confirmed by the results of this study. Dbp2 was found to be methylated in this study for the first time. The remaining two proteins, Ski2 and Pub1 have not yet been described to be methylated. Npl3, Psp2, Lge1 and Ded1 lack the RGG-box annotation in UniProt, but fall into this group according to the definition employed here.

The RGG-box was first described as a motif after mapping the RNA binding region of human hnRNP U to a stretch of 26 amino acids containing a repeat of three RGG sequences in RGGDFRGGAPGDRGG and demonstrating its conservation in different species (Kiledjian & Dreyfuss 1992). It has since been found in several RNA-binding proteins and its definition has been relaxed on several occasions to e.g. only requiring two RGG sequences and to allow increased space between them (Burd & Dreyfuss 1994; Corley & Gready 2008). Often aromatic amino acids are found in this motif and it has been suggested that they are involved in stacking with RNA bases (Kiledjian & Dreyfuss 1992). Hanakahi et al. described the

binding of the C-terminal RGG repeats of nucleolin to G-quartets in rDNA, where the recognition of the G4 structure is independent of its sequence context and the RNA-binding domains of this protein also interact with this structure (Hanakahi et al. 1999). Zanotti et al. (2006) showed the interaction of an RGG-box in FMRP to the G-quartet of the Sc1 RNA, however noted that this interaction is not sufficient for binding and suggested that some sequence context-specificity of either the RNA or the protein is required for binding. Mears & Rice (1996) described the interaction of the viral protein ICP27 with RNA via its RGG box. Several RNA-binding sequences containing repetitions of RGG or YGG have been found subsequently (Castello et al. 2012).

In a proteome-wide study in human cells, 43% of the detected arginine methylation sites were in an RG-context and proline at position -1, +1 and +4 or isoleucine or alanine at +1 were also frequently present (Uhlmann et al. 2012). One methylation site immediately followed by proline (R386 in Nsr1) was observed here, while eight sites had a proline at the -1 position, all of which were found in Npl3.

A pronounced preference for localisation of sites within unstructured regions suggests that this may be a general requirement for methylation to occur. The absence of secondary structure elements is consistent with the prevalence of glycines in the vicinity of methylated arginines and the finding that the only site found not to be followed by glycine is instead followed by proline. It is also consistent with the well-established occurrence of arginine methylation in histone tails (Byvoet et al. 1972). Additionally, sequences with strong asparagine-bias, which were observed for three of the methylated proteins, may be a further way to ensure an unfolded nature. These findings suggest that methylation sites are recognized by their primary structure rather than their fold by the corresponding PRMT or co-factor.

### 2.9.8. Comparison of Experimental Strategies

**Table 2.12:** Comparison of enrichment and fractionation approaches for the identification of protein methylation sites.

| Method | Advantages | Disadvantages |
|---|---|---|
| Protein IP | -Sensitive through enrichment | -Dependent on antibody quality<br>-Potential sequence bias<br>-Expensive reagents |
| 3xMBT | -Sensitive through enrichment<br>-Cheap reagents | -Only recognizes proteins that are similar to *in vivo* binding partners |
| HILIC | -No affinity bias<br>-Cheap reagents | -Limited sensitivity as only fractionation without enrichment<br>-Requires much instrument time |
| Peptide IP | -Sensitive through enrichment<br>-Absence of non-methylated peptides | -Dependent on antibody quality<br>-Potential sequence bias<br>-Expensive reagents |

A comparison between the different enrichment and fractionation methods used in this study is shown in table 2.12. While the IPs are expected to allow the detection of low abundance methylation sites through high-affinity enrichment, they are limited by potential sequence biases and the general quality of the antibodies used. Additionally, batch-to-batch variation is a concern with respect to reproducibility, particularly for polyclonal antibodies and the cost of the experiments is high due to the need to purchase antibodies. Using the 3xMBT domain as an affinity reagent removes this cost, while the sensitivity is still expected to be high due to enrichment. However, sequence bias is now introduced by the specificity of the protein domain.

In contrast, an orthogonal off-line chromatographic fractionation like HILIC is not limited by biased enrichment, however offers lower sensitivity than enrichment methods. There are no expenses for affinity reagents, but the requirement of mass spectrometry instrument time is high.

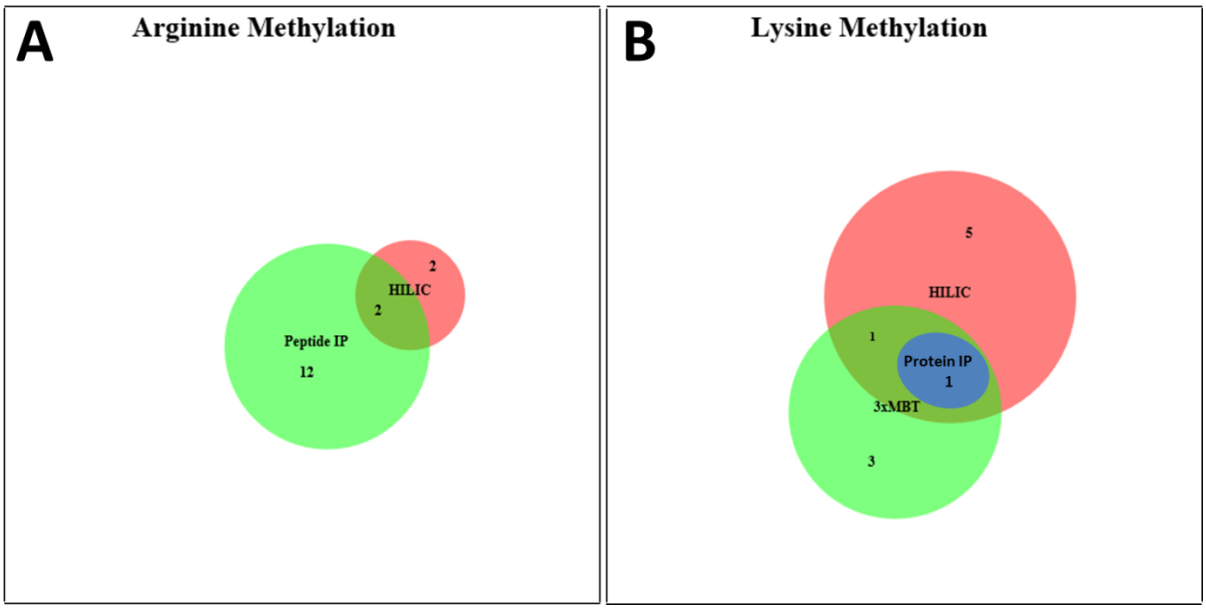



***Figure 2.17: Overlap of methylated proteins identified with different experimental approaches in this study.*** *Venn diagrams of arginine methylated proteins found by peptide immunoaffinity enrichment, and HILIC fractionation (A) and lysine methylated proteins identified via protein immunoaffinity enrichment, 3xMBT-based pull-down and HILIC fractionation. Diagrams were created using BioVenn* (Hulsen et al. 2008).

Complete lists of the arginine and lysine methylation sites that were found in different experiments of this study are shown in tables 2.13 and 2.14, respectively. The overlap of methylated proteins between experiments is depicted in fig. 2.17. The clearly most successful approach with 41 arginine methylation sites identified on 14 proteins was immunoaffinity enrichment at the peptide level. The fact that this experiment was successful with antibodies for arginine mono-methylation, while not with antibodies for di-methylation suggests that it is highly dependent on antibody quality. Out of the four arginine methylated proteins identified via HILIC fractionation, two were also identified by protein immunoaffinity enrichment (Fig. 2.17 A) which is potentially due to a high abundance of these proteins. While there is some overlap between experiments for lysine methylated proteins (Fig. 2.17 B), the majority of the proteins was found with only one approach. An exception is lysine methylation on translation initiation factor 1-alpha which was discovered after Protein IP, 3xMBT pull-down and HILIC fractionation.

**Table 2.13:** Protein arginine methylation sites in yeast identified by all experiments in this study.

| Accession | Protein Name | Prot. Prev. Known | Methylated Residue | Methyl. Site Prev. Known | Study | Modified Peptide | Experiment |
|---|---|---|---|---|---|---|---|
| P50109.2 | Psp2 | n | 419 | | | YNGNHNNNNGNFR(me1)GSNR | Peptide IP |
| | | | 425 | | | YR(me1)GGPNGSSYK | Peptide IP |
| | | | 575 | | | GGR(me1)GSSGNYSNYNNR | Peptide IP |
| | | | 538 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR | Peptide IP |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR | Peptide IP |
| | | | 551 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR | Peptide IP |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR | Peptide IP |
| | | | 551 or 553 | | | GNYNSSGMNGGSRGR(me1) or GNYNSSGMNGGSR(me1)GR | Peptide IP |
| P24783.1 | Dbp2 | n | 43 | | | NSYNDRPQGGNYR(me1)GGFGGR | Peptide IP |
| | | | 509 | | | YGGGR(me1)GGR(me1)GGYGR | Peptide IP |
| | | | 512 | | | YGGGR(me1)GGR(me1)GGYGR | Peptide IP |
| | | | 518 | | | R(me1)GGYGGGR | Peptide IP |
| | | | 525 | | | GGYGGGR(me1)GGYGGNR | Peptide IP |
| P06634.2 | Ded1 | y | 51 | y(me2) | Erce, 2013 | NNSSNYNNNNGGYNGGR(me1)GGGSFFSNNR | Peptide IP |
| | | | 62 | y(me2) | Erce, 2013 | R(me1)GGYGNGGFFGGNNGGSR | Peptide IP |
| | | | 578 | y | Erce, 2013 | DNSFR(me1)GGSGWGSDSK | Peptide IP |
| P15646.1 | Nop1 | y | 70 | y(me1 and me2) | Low, 2013 | GGSR(me1)GGR(me1)GGAAGGAR | Peptide IP |
| | | | 73 | y (me1 and me2) | Low, 2013 | GGR(me1)GGAAGGAR | Peptide IP |
| | | | | | | GGSR(me1)GGR(me1)GGAAGGAR | Peptide IP |
| | | | | | | GGR(me1)GGAAGGAR(me1) | Peptide IP |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK | Peptide IP |
| | | | 81 | y(me2) | Low, 2013 | GGAAGGAR(me1)GGAK | Peptide IP |
| | | | | | | GGR(me1)GGAAGGAR(me1) | Peptide IP |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK | Peptide IP |
| | | | 16 or 34 or 52 [b] | all three: y(me2as - UniProt by Similarity) | | GGSR(me1)GGFGGR | Peptide IP |
| P28007.1 | Gar1 | y | 147 | y(me2as- UniProt probable) | | SGAPGGR(me1)GGASMGR | Peptide IP |
| P27476.1 | Nsr1 | y | 353 | y(me2as- UniProt probable) | | LDFSSPRPNNDGGR(me1) | Peptide IP |
| | | | 357 | n | | LDFSSPRPNNDGGRGGSR(me2) | HILIC |
| | | | 379 | y(me2as- UniProt probable) | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR | Peptide IP |
| | | | 382 | y(me2as- UniProt probable) | | GGR(me1)GGFRPSGSGANTAPLGR | Peptide IP |
| | | | | | | GGAR(me1)GGR(me1)GGFRPSGSGANTAPLGR | Peptide IP |
| | | | | | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR | Peptide IP |
| | | | 386 | n | | GGRGGFR(me1)PSGSGANTAPLGR | Peptide IP |
| | | | | | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR | Peptide IP |
| P49626.2 / P10664.4 | Rpl4B/Rpl4A | n | 95 | | | SGQGAFGNMCR(me1)GGR | Peptide IP |
| Q01560.1 | Nop3 / Npl3 | y | 307 | y(me1 and me2) | Hart-Smith, 2012 | GGFSR(me1)GGFGGPR | Peptide IP |
| | | | | | | GGFSR(me1)GGFGGPR(me1) | Peptide IP |
| | | | | | | GGFSR(me1)GGFGGPR(me1)GGFGGPR | Peptide IP |
| | | | 314 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGFGGPR | Peptide IP |
| | | | | | | GGFSR(me1)GGFGGPR(me1) | Peptide IP |
| | | | | | | GGFSR(me1)GGFGGPR(me1)GGFGGPR | Peptide IP |
| | | | 321 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) | Peptide IP |
| | | | 329 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) | Peptide IP |
| | | | 337 | y(me1 and me2) | Hart-Smith, 2012 | GGYGGYSR(me1)GGYGGSR | Peptide IP |
| | | | 344 | n | | GGYGGSR(me1)GGYDSPR | Peptide IP |
| | | | | | | GGYGGSR(me1)GGYDSPR(me1) | Peptide IP |
| | | | 351 | y(me1) | Hart-Smith, 2012 | GGYDSPR(me1)GGYDSPR | Peptide IP |
| | | | | | | GGYGGSR(me1)GGYDSPR(me1) | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYDSPR(me1) | Peptide IP |
| | | | 358 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1)GGYSR | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYDSPR(me1) | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYSR(me1) | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYSR(me1)GGYGGPR [c] | Peptide IP |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | 363 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPRGGYSR(me1) | Peptide IP |
| | | | | | | GGYSR(me1)GGYGGPR | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYSR(me1) | Peptide IP |
| | | | | | | GGYDSPR(me1)GGYSR(me1)GGYGGPR [(c)] | Peptide IP |
| | | | | | | GGYSR(me2)GGYGGPR | HILIC |
| | | | 377 | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR | Peptide IP |
| | | | 384 | y(me1 and me2) | Hart-Smith, 2012 | GSYGGSR(me1)GGYDGPR | Peptide IP |
| | | | | | | GSYGGSR(me1)GGYDGPRGDYGPPR | Peptide IP |
| | | | 391 | y (me2) | Hart-Smith, 2012 | GGYDGPR(me1)GDYGPPR | Peptide IP |
| | | | 314 or 321 [(b)] | both: y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1) | Peptide IP |
| | | | 351 or 358 [(b)] | 351: y(me1); 358: y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1) | Peptide IP |
| | | | 377 or 384 [(b)] | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR or NDYGPPRGSYGGSR(me1) | Peptide IP |
| P32505.1 | Nab2 | y | 222 | (region: 180-256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) | Peptide IP |
| | | | 226 or 229 | (region: 180-256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) | Peptide IP |
| P10080.2 | Ssbp1 / Sbp1 | y | 125 | y(me1 and me2) | Hsieh, 2007 [(a)] | TPGQM(ox)QR(me1) | Peptide IP |
| | | | | | | TPGQMQR(me1) | Peptide IP |
| | | | 145 | y(me2) | Hsieh, 2007 [(a)] | GGFR(me1)GGYR | Peptide IP |
| | | | 165 | y(me2) | Hsieh, 2007 [(a)] | GGAR(me1)GGFNGQK | Peptide IP |
| | | | | | | GGAR(me1)GGFNGQKR | Peptide IP |
| | | | | | | GR(me1)GGAR(me1)GGFNGQKR | Peptide IP |
| P40070.1 | Lsm4 | n | 119 | | | DSNNNR(me1)GNYNR | Peptide IP |
| Q02796.1 | Lge1 | n | 39 | | | GGAGGSYYR(me1)GGNASYGAR | Peptide IP |
| Q02884.1 | Elp4 | n | 13 | | | GEILNDR(me1)GSGLR | Peptide IP |
| P06169.1 | Pdc1 | n | 161 | | | TTYVTQR(me1)PVYLGLPANLVDLNVPAK | Peptide IP |
| P07866.2 | Lte1 | n | 649 | | | ESSPLNDSR(me1) | HILIC |
| Q00402.2 | Num1 | n | 175 | | | LTNLSMECR(me1) | HILIC |

[(a)] Sites of methylation are not given in this study, but could be partially derived from the peptides and number of methyl groups; these sites were found by expressing Sbp1 and Hmt1 in *E.coli.*

[(b)] The identified peptide matches to more than one location in the protein.

[(c)] Only one of the sites was surrounded by fragment ions.

While the overlap of lysine methylated proteins between HILIC fractionation and antibody- and 3xMBT-based enrichment (Fig. 2.17 B) is reflected by an overlap of the methylation sites and types (K30me1, K316me2 and K390me1 on EF1A and K40me1 on Rpl42A/B; Table 2.14), the arginine methylations found by HILIC fractionation are complementary to those found by peptide immunoaffinity enrichment: R363 on Npl3 was found in di-methylated form in the former, while identified mono-methylated in the latter. Di-methylation of R357 of Nsr1 was detected after HILIC fractionation, whereas mono-methylation of four other sites of this protein had been found by peptide IP.

Note that one protein, Pdc1, carries both an identified arginine and lysine methylation site. These sites are ~360 residues apart and it remains to be determined if the co-occurrence is due to random chance or if the events are related.

**Table 2.14:** Protein lysine methylation sites in yeast identified by all experiments in this study.

| Accession | Protein Name | Prot. Prev. Known | Methylated Residue | Methyl. Site Prev. Known | Study | Modified Peptide | Experiment |
|---|---|---|---|---|---|---|---|
| P02994.1 | EF1A | y | 30 | y | Couttas, 2012 | STTTGHLIYK(me1) | HILIC; Protein IP; 3xMBT |
| | | | 79 | y | Cavallius, 1993; Couttas, 2012 | GITIDIALWK(me3)FETPK | HILIC |
| | | | 316 | y | Cavallius, 1993; Couttas, 2012 | NVSVK(me1)EIR | HILIC |
| | | | 316 | y | Cavallius, 1993; Couttas, 2012 | NVSVK(me2)EIR | HILIC; 3xMBT |
| | | | 390 | y | Cavallius, 1993; Couttas, 2012 | KLEDHPK(me1) | HILIC; 3xMBT |
| P32324.1 | EF2 | y | 613 | n | | DDFK(me1)AR | 3xMBT |
| | | | | | | DDFK(me2)AR | HILIC; 3xMBT |
| P00360.3/ P00358.3/ P00359.3 | G3P1/ G3P2/ G3P3 | n | 225 | | | VLPELQGK(me1) | HILIC |
| P61830.2 | H3 | y | 80 | y | Ng, 2002 | EIAQDFK(me1)TDLR | HILIC |
| | | | 80 | y | van Leeuwen, 2002; Lacoste, 2002 | EIAQDFK(me2)TDLR | HILIC |
| | | | 80 | y | van Leeuwen, 2002; Lacoste, 2002 | EIAQDFK(me3)TDLR | HILIC |
| P06169.7 | Pdc1 | n | 520 | | | VATTGEWDK(me1) | HILIC |
| P0CX28.1/ P0CX27.1 | Rpl42A/ Rpl42B | y | 40 | y | Webb, 2008; Couttas, 2012 | ASLFAQGK(me1) | 3xMBT |
| | | | 40 | y | Webb, 2008; Couttas, 2012 | ASLFAQGK(me1)R | HILIC; 3xMBT |
| | | | 55 | y | Webb, 2008 | QSGFGGQTK(me1)PVFHK | HILIC |
| P17255.3 | VatA | n | 207 | | | K(me2)SDFTLYHTWPVR | HILIC |
| Q12420 | Snu66 | n | 337 | | | TSSNNELK(me1) | 3xMBT |
| P46677 | TAF1 | n | 974 | | | INNLEELEK(me2) | 3xMBT |

As described in "2.9.7. Identification of Arginine Methylation Sites by Peptide Immunoaffinity Purification", the arginine methylation sites detected by peptide IP show a bias for localisation in an RG and RGG context and disordered regions and a functional enrichment for proteins involved in RNA processing. One arginine methylation site detected by HILIC fractionation overlaps with a site identified by peptide IP (R363 on Npl3) and is in an RGG context, while a second is an additional site (R357 on Nsr1) on a protein that had been found methylated and is followed by glycine. Both sites are located in unstructured regions and the corresponding proteins are involved in RNA processing. Two additional sites (R649 on Lte1 and R175 on Num1) identified after HILIC fractionation are neither located in an RG or RGG context nor on RNA-binding proteins. While no experimental or predicted structure is available for Lte1, the site on Num1 is located in a region that is neither annotated as an extended unstructured region nor as a protein domain. These findings suggest that while

the biases detected by peptide IP are likely to be mainly of a biological nature, they may be less pronounced when using an unbiased experimental method. Overall however, the numbers of sites detected by HILIC fractionation are too low to draw meaningful conclusions.

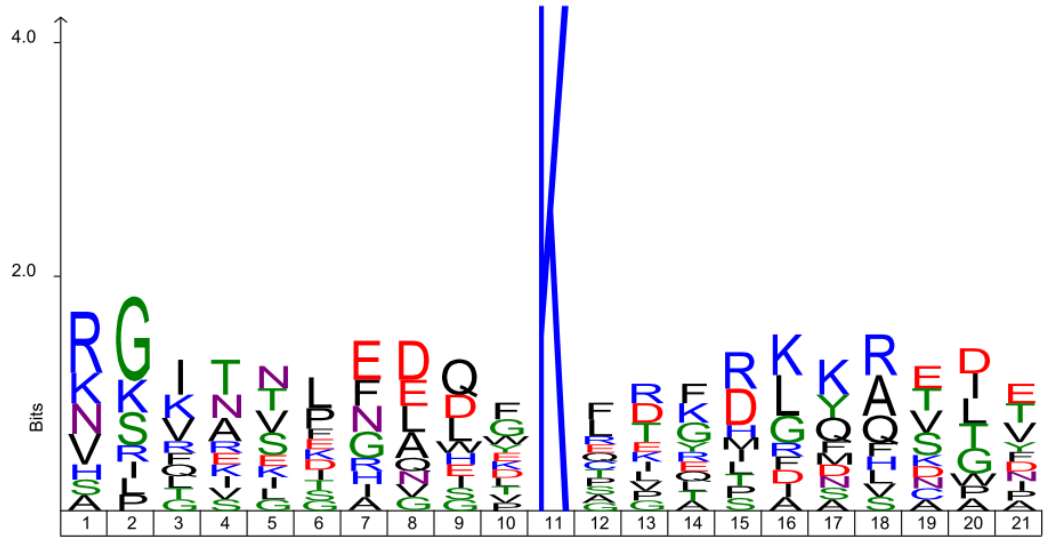



***Figure 2.18: Lysine methylation sequence motifs.*** *A sequence logo of ten amino acids upstream to ten downstream of the methylated lysines from all experiments in this study is shown.*

As the numbers of lysine methylation sites for each individual experiment in this study are low, a sequence motif was generated from a combination of all sites found via HILIC separation, protein IP and 3xMBT-based enrichment (Fig. 2.18). The enrichments of amino acid frequencies around the methylated lysine are not as clear as for methylated arginines (Fig. 2.14) and this is consistent with the current model that several different methyltransferases catalyse the methylations of individual lysines, while most of the arginine methylation events in *S. cerevisiae* are attributed to Hmt1 (Clarke 2013; Gary et al. 1996). For example, of the previously known lysine methylation sites identified here, K80 on histone H3 is methylated by Dot1 (Kmt4) (van Leeuwen et al. 2002), K30 and K79 on EF1A by EFM1 and K316 on the same protein by See1 (Couttas et al. 2012). On Rpl42A/B K40 is methylated by Rkm3 while the methyl-group is transferred to K55 by Rkm4 (Set7) (Webb et al. 2008). Nonetheless, a higher frequency of basic and acidic amino acids around the methylated lysine

than in the background yeast proteome is apparent. Additionally, a slight preference for asparagines and glycines N-terminal to the methylated lysine is also visible.

Two of the sites on EF1A (K30 and K79) are located in the guanine nucleotide binding domain and the site on TAF1 (K974) is within a coiled coil region. All other sites are not located in annotated domains, however, with the exception of K338 of Snu66, neither within extended disordered regions. A requirement for the absence of structure for lysine methylation therefore appears unlikely.

### 2.9.9. Proteome-wide Discovery Studies of Protein Methylation in *S. cerevisiae*

Before any proteome-wide studies aiming at the discovery of protein methylation sites in yeast had been attempted, sites had been predicted bioinformatically (Pang et al. 2010). This study was based on previously published peptide mass fingerprinting data on 2,607 purified proteins and methylation sites were assigned using the FindMod software (Wilkins et al. 1999) and subsequent filtering. This led to the prediction of 40 lysine-, 31 arginine- and 5 simultaneously lysine- and arginine- methylated proteins. Only a few of the reported methylation sites however have been verified experimentally to date.

Additional to the results originating from this work (Plank, 2015) several proteome-wide discovery studies of protein methylation in *S. cerevisiae* have been published during the time of this work:

A LC-MS/MS based study for the identification of lysine methylation sites detected methylation sites on three translation elongation factors and four ribosomal proteins (Couttas et al. 2012). This study did not use the MethylSILAC strategy and instead of enrichment of methylated proteins or peptides, a strategy to equalise the abundances of peptides in the sample was employed.

In a more recent study, 90 proteins in *S. cerevisiae* were found in a pull-down with an anti-Rme1 antibody and 15 methylation sites could be mapped to five of the proteins by tandem

mass spectrometry (Low et al. 2013). Additional 21 sites (mono- or di-methyl) on five proteins were subsequently verified by the same group by co-expressing the target proteins with the PRMT Hmt1 in *E. coli* (Erce et al. 2013).

Wang (2015) attempted the discovery of methylation sites by 2-dimensional LC-based MS analysis without prior enrichment. A modification of the MethylSILAC strategy in which heavy SAM instead of methionine was added to the medium was utilized and allowed the detection of 64 methylated proteins. Most of the methylation sites were mapped to lysine, while sites were also found on arginine, aspartate, asparagine, glutamine, histidine, glutamate, and cysteine residues.

Finally, Yagoub et al. (2015) increased the number of known arginine methylation sites on Npl3, Nop1, Gar1, Nsr1 and Rps2 using peptide-immunoaffinity purification with a different antibody than used in this work and showed by *in vitro* assays that the deposition of these methyl-groups is catalysed by Hmt1.

A picture of limited overlap originates from the comparison of all arginine (Fig. 2.19 A) and lysine (Fig. 2.19 B) methylated proteins identified in this and published studies. This low overlap may indicate that only a small portion of the methylome was identified in each case or that each study suffers from distinct biases towards certain proteins.

Particularly noteworthy is the fact that there is no overlap for both lysine and arginine methylation sites, between this study and the proteins reported by Pang et al. (2010), despite the study by Pang et al. being largest dataset. Additionally, the work of Pang et al. also exhibits hardly any overlap with other studies except with Low (2013) and it should be noted that the latter may be biased as only some of the proteins detected by IP were selected for verification. This may indicate a limited number of true positive predictions of methylated proteins from peptide mass fingerprinting (Pang et al. 2010). On the other hand, the studies by Yagoub (2015) and Couttas (2012) exhibit considerable overlap with the results obtained here

and it may be speculated that this is due to the detection of highly abundant and highly methylated proteins.

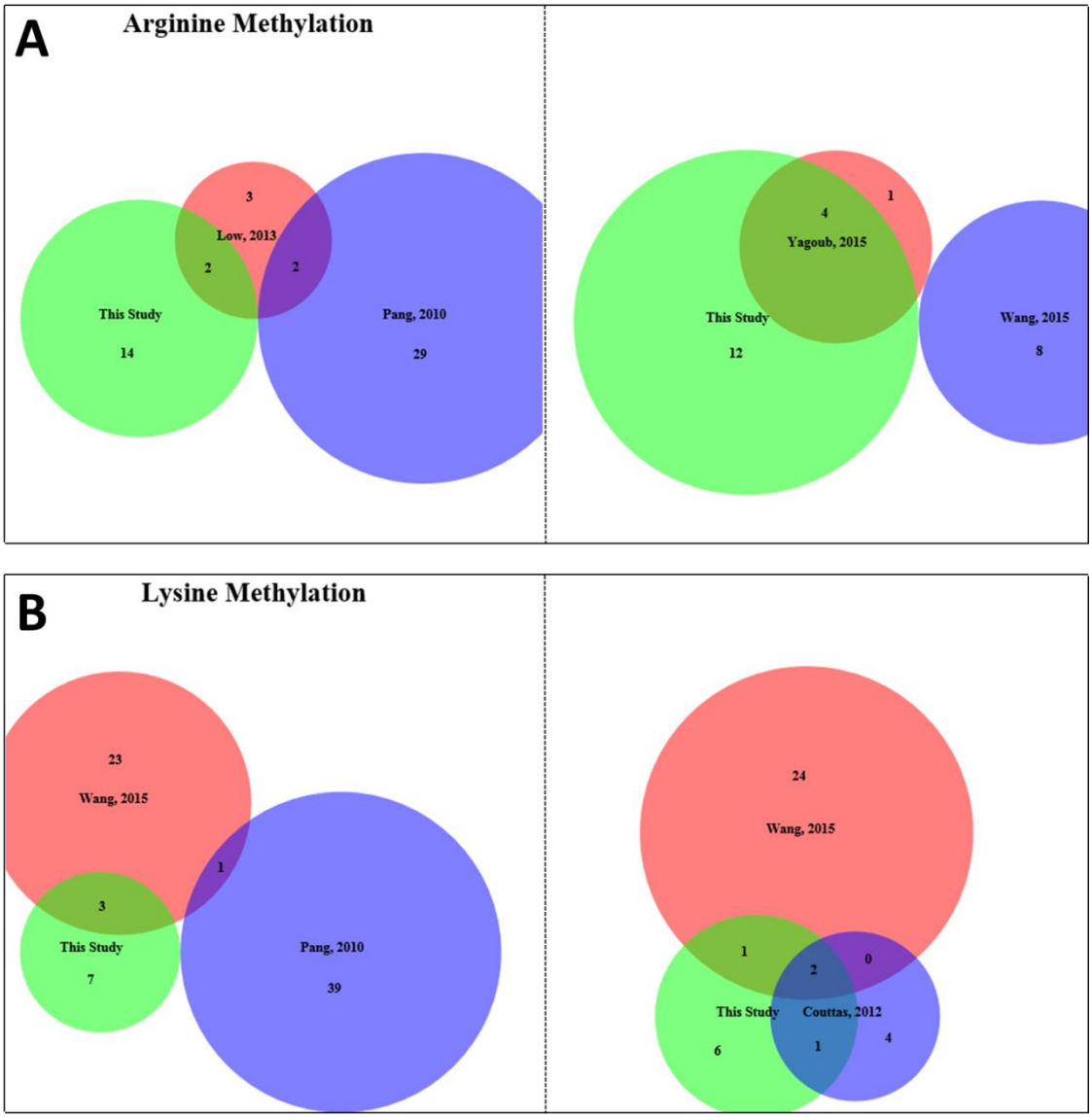



***Figure 2.19: Overlap of identifications of methylated proteins.*** *Venn diagrams of the overlap between arginine (A) and lysine (B) methylated proteins identified in this and published studies. Diagrams were created using BioVenn* (Hulsen et al. 2008).

To determine if the approaches for the identification of protein methylation sites presented here are limited by their sensitivity, a potential association between the detection of a protein as methylated and its cellular abundance was tested. For this previously published data on copy numbers were used (Ghaemmaghami et al. 2003). Fig. 2.20 shows the rank vs. the $\log_{10}$

of the copies per cell and highlights each of the proteins found to be methylated with different approaches. About 80% of the methylated proteins are part of the dataset on copy numbers. The mean of $\log_{10}$ (Molecules / Cell) for each experiment against the background of the total quantified yeast proteome is depicted in fig. 2.21. The mean for the copy numbers of methylated proteins detected after HILIC fractionation (A), peptide IP (B) and 3xMBT domain based enrichment (C) are all clearly above the mean for the total quantified proteome. Notably, in each case at least one protein of low cellular abundance (about 400 molecules / cell) was found to be methylated. While the identity of each of these proteins is different in each approach, the possibility of potential mis-quantifications of the copy numbers should be taken into account. As expected, the fractionation approach seems more limited by sensitivity that the enrichment approaches. The results for the enrichment by protein IP is omitted from the figures, as only one methylated protein was reliably identified. This protein, Elongation factor 1-alpha, is present at about 450 copies per cell. If this quantification is reliable, this result suggests a low limit of detection for this approach. It might also indicate a strong bias of the antibody towards this protein, allowing it to be enriched with high affinity. Notably, the mean copy numbers for yeast proteins found methylated in published studies are also all located above the mean of the quantified proteome, in a similar region as for the approaches presented here, i.e. about 5,000 to 30,000 copies per cell. Notably, the only study relying on sample fractionation alone without prior enrichment yields the lowest mean copy number and covers the widest range of protein abundances, down to 125 copies per cell (Wang et al. 2015). The study based on the analysis of purified proteins also covers a wide abundance range, however exhibits a clear bias towards more abundant proteins (Pang et al. 2010).

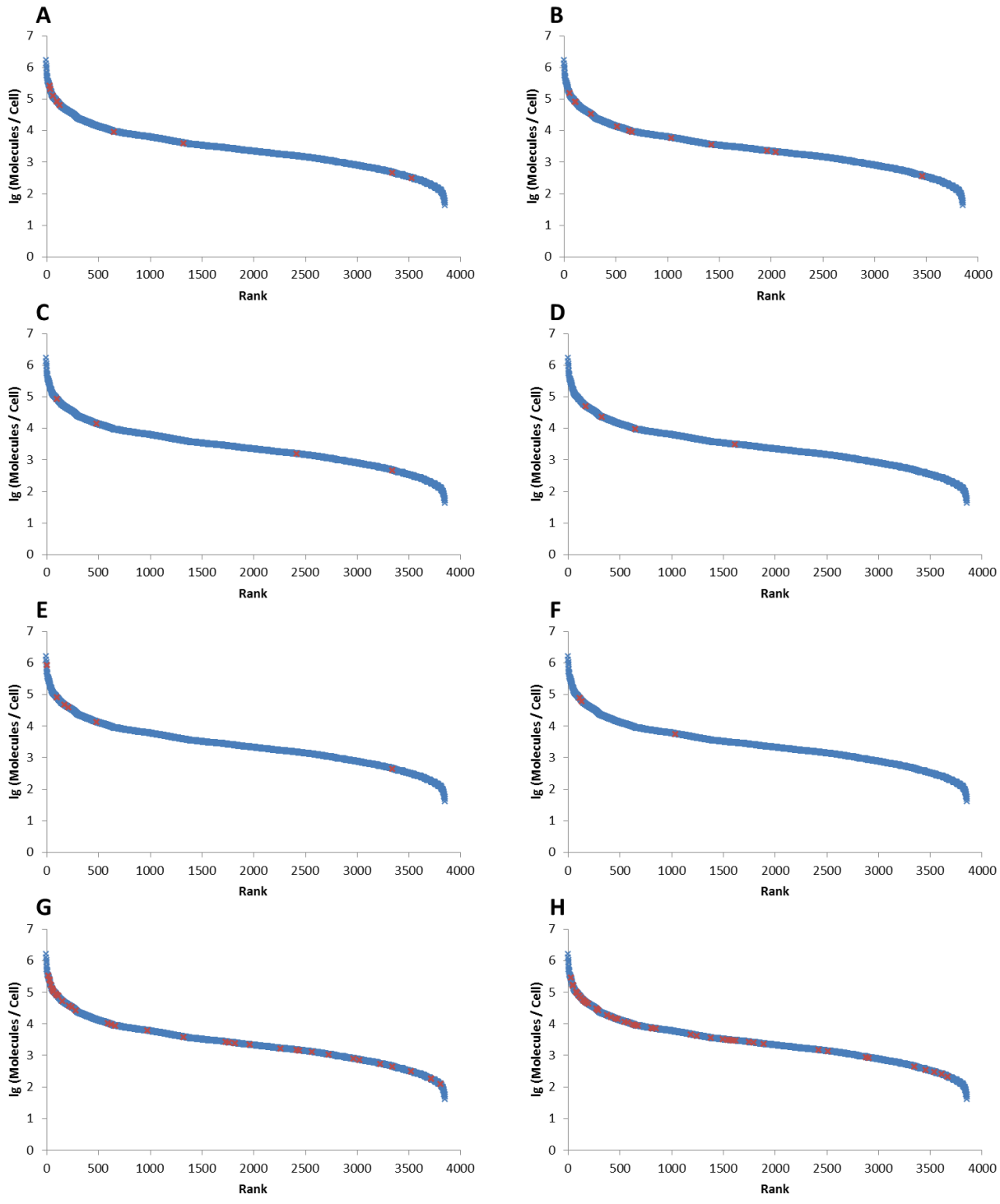



***Figure: 2.20: Cellular abundances of methylated proteins.*** *The cellular abundances for each protein as determined by Ghaemmaghami et al. (2003) are plotted as their rank vs. the $log_{10}$ of molecules per cell in blue. The proteins found to be methylated with different approaches in this study (A-C) and in published studies (D-H) are highlighted in red: A, Fractionation by HILIC, B, Peptide immunoaffinity enrichment, C, 3xMBT domain-based pull-down, D, Low et al., 2013, E, Couttas et al., 2012, F, Yagoub et al., 2015, G, Wang et al., 2015, H, Pang et al., 2010.*

Additionally to the cellular abundance of proteins, methylation stoichiometries should be taken into account to evaluate if methods are limited by their sensitivity. Unfortunately, to date insufficient data exist on protein methylation stoichiometries in yeast. Further, the current analysis does not consider the number of methylation sites on each protein. Finally, it cannot be ruled out that a biological bias towards methylation of more abundant proteins underlies these data.

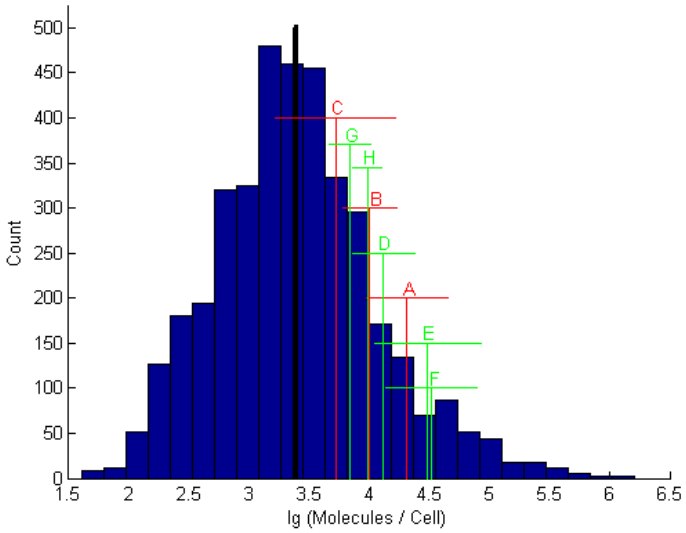



***Figure 2.21: Comparison of cellular abundances of methylated proteins to the yeast proteome.*** *The blue histogram depicts the $log_{10}$ of molecules per cell for all proteins determined by Ghaemmaghami et al (2003). Vertical bars represent means and horizontal bars the standard error of the mean. Black: all proteins, red: methylated proteins identified with different approaches in this study (A-C), green: methylated proteins identified in published studies (D-H). Labels correspond to labels of plots in fig. 2.20: A, Fractionation by HILIC, B, Peptide immunoaffinity enrichment, C, 3xMBT domain-based pull-down, D, Low et al., 2013, E, Couttas et al., 2012, F, Yagoub et al., 2015, G, Wang et al., 2015, H, Pang et al., 2010.*

In conclusion, all approaches towards the identification of methylated proteins in this study and by others are biased towards more abundant proteins. A bias for abundant analytes is not surprising for any experimental method that does not exhaustively cover all identifiable species. Nonetheless, most experiments also identify proteins present at lesss copies per cell. Therefore, the data are probably best explained by a combination of factors, including abundance, allowing a methylation of a protein of lower abundance to be detected if

favourable to the bias of the enrichment method and if the originating peptides are well-suited for analysis by LC-MS/MS.

### 2.9.10. Shortcomings of Current Approaches for the Identification of Protein Methylation Sites

In widely used standard LC-MS/MS-based proteomics approaches the identification of several peptides of a protein are generally required for the identification of the protein. In turn, the matching of fragment spectra to each individual peptide with a limited confidence is often sufficient. In contrast, the identification of a PTM usually relies on the observation of a single peptide, so that a much higher confidence in this identification would be required. The discovery of PTMs on peptides by global approaches therefore often only serves as a presentation of candidates that need to be verified by additional experimental techniques. This problem becomes even more severe when trying to map a modification to a specific residue if it can occur on several different amino acids. As it became apparent during the course of this study that methylation sites are much rarer in yeast than in mammals, it was aimed to provide a small set of high-confidence methylation sites rather than a more extended list of candidate sites. One way to achieve a higher confidence in the assignment of methylation sites is the use of MethylSILAC, which was employed throughout this study, to provide an indication for the presence of methylation orthogonal to the information of the MS/MS spectrum. A means to achieve confident site localisation is to require the presence of fragment ions directly adjacent to the methylation site, which however bears the danger of being overly restrictive and missing true methylation sites. In this respect it would be desirable to be able to increase the quality of fragment spectra to achieve sufficiently high confidence levels to base the assignment of a methylation site on a single spectrum.

While the MethylSILAC strategy was successfully used for the identification of methylation sites in this study, it suffers several shortcomings as discussed below.

First, it was observed that the labelling efficiencies of the heavy samples was generally lower than 90%, even though the expected labelling efficiency after eight generations is 96% ((1-$0.5^8$)*100%) and cells had been labelled for at least eight generations. While incomplete labelling is not expected to cause major problems for singly methylated peptides, the simulation shown in fig. 2.22 demonstrates the drastic decrease of the proportion of the fully labelled version of a peptide with several methyl-groups. For example, at a labelling efficiency of 75%, only about half of the peptides in the heavy sample are fully labelled if they have two methyl-groups (two mono- or one di-methylation).

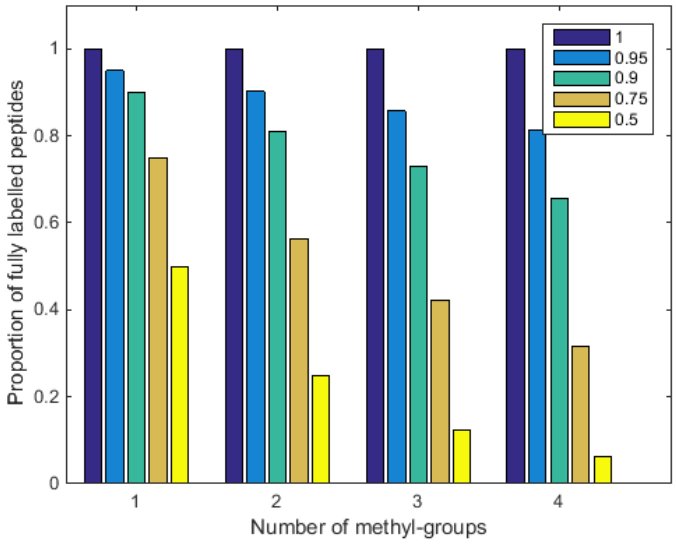



***Figure 2.22: Simulated proportion of fully labelled peptides at a given number of methyl-groups.*** *The number of methyl-groups on a peptide is plotted against the calculated proportion of fully labelled peptides at the labelling efficiencies indicated in the legend based on a binomial distribution.*

The incomplete labelling is likely attributed to metabolically generated light methionine. This is counter-intuitive at first thought, as the yeast strain used for this study was methionine auxotrophic (*met15Δ*). However, while this strain cannot synthesize methionine *de novo*, it can via the SAM-cycle regenerate light methionine from homocysteine originating from the methionine taken up from the medium (Fig. 2.23) (Scott & Weir 1981; Thomas & Surdin-

Kerjan 1997). Therefore, incomplete labelling even after many generations is possible, as methionine auxotrophy is not equivalent to SAM auxotrophy.

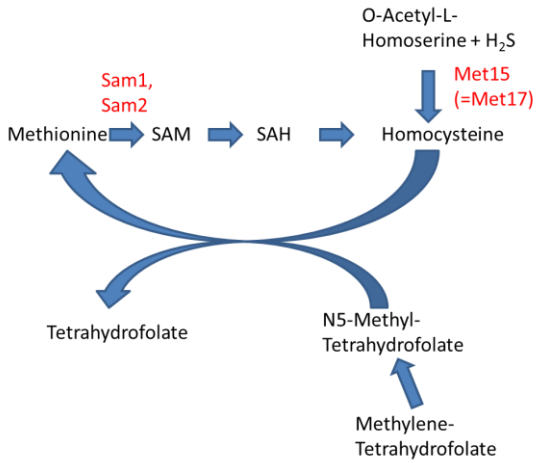



***Figure 2.23: Simplified scheme of the SAM-cycle.*** *Only selected reactants (black) and enzymes (red) are shown. Methionine is converted to S-Adenosyl methionine (SAM) by Sam1 or Sam2. SAM is used for methylation reactions, yielding S-Adenosyl homocysteine (SAH), which is converted to homocysteine and recycled to methionine using N5-Methyl-tetrahydrofolate. Homocysteine is also synthesized from O-Acetyl-L-homoserine and $H_2S$ in a Met15 (Met17) catalysed reaction.*

The principle of the MethylSILAC strategy is to make use of the presence of heavy-light pairs additionally to MS/MS information in identifying methylated peptides (see '2.1.3. MethylSILAC'). Two sources exist for false positive identifications of heavy-light pairs. Firstly, two peptide features may randomly occur at the distance resembling or nearly resembling a heavy-light pair. To date, no approach has been explored that makes use of the MS/MS information of both the light and the heavy partner in the MethylSILAC pair, rather than only the presence of pairs of peaks at MS1 level. However, considering the MS/MS spectra of both partners may improve the detection of heavy-light pairs, as with the increasing speed of mass spectrometers it becomes increasingly feasible to fragment both precursors. Secondly, heavy-light pairs will arise from methionine containing peptides. An elegant strategy to overcome this problem, iMethyl-SILAC, has been described in which heavy methionines of the same mass are used in both samples. However, in one of them the heavy isotopes are localized in the methyl-group, while in the other they are found in the $C_\beta$ or $C_\gamma$ position and their adjacent hydrogen atoms (Geoghegan et al. 2015). Therefore, while a heavy methyl-group is transferred upon methylation only in the first case, methionines incorporated into proteins are of equal mass. Consequently, methionine containing peptides will not appear

as heavy-light pairs. Unfortunately, the second variant of heavy methionine proved preventatively expensive for this study.

Beyond these problems, enrichment methods are prone to bias, e.g. to certain sequence contexts of methylation-sites, while upon fractionation without enrichment methylated peptides may not be present in samples at concentrations sufficient for detection by mass spectrometry.

In the following, a study aiming at the detection of methylation sites on purified candidate proteins is presented. This is expected to reduce the problems of biased enrichment and limited sensitivity.

# 3. Discovery of Protein Methylation by a Candidate Based Approach

## 3.1. Introduction

### 3.1.1. Rationale for Candidate-based Discovery

While global approaches are a fast way to obtain an overview of the identities and characteristics of methylation sites in an organism, they suffer from several shortcomings that may lead to missing identifications. As detailed in section 2.9, biases and insufficient sensitivity to detect low-abundance methylation sites may be limiting the results to a subset of detectable sites. Furthermore, when using LC-MS/MS, some methylation sites may reside on peptides that are not of an adequate length for detection after protease digestion. Also, when searching a complete proteome, the number of false positives will increase dramatically when an increasing number of PTMs is considered in database searching. Therefore, methylation-sites may be assigned to incorrect amino acids. On the other hand, if restricting the search to a too low number of considered modifications, methylation sites may be missed if they only occur on peptides that also harbour additional modifications. These problems are less pronounced when focussing on a single protein, as identifications of peptides mapping to other proteins can be disregarded and identifications to the purified protein are unlikely to occur by chance.

Therefore a set of candidate proteins from *S. cerevisiae* was purified individually and tested for methylation.

### 3.1.2. Candidate Selection

To compile a list of proteins suspected to be arginine methylated, candidates were obtained from the following sources:

In a previous study aimed at the identification of arginine methylated proteins in *S. cerevisiae* by a protein-IP with an anti-Rme1 antibody, 90 proteins were identified. However, only 15 methylation sites could be mapped to five of the proteins by tandem mass spectrometry (Low

et al. 2013). The identified proteins were selected as candidates for overexpression and purification: Cdc10, Mam33, Nhp6A, Ses1, Rim1, Has1, Mbf1, Dbp6, Pob3, Lat1, Rpl8B, Rps14A, Rpp2B, Nap1, Cbf5, Nop56, Rps28A, Rpl5, Hsc82, Cdc33, Rfa1, Rpl38, Imd3, Imd4, Dys1, Pwp1, Ssc1, Ent5, Cdc12, Crp1, Rps22A, Kgd2, Crp6, Lpd1, Ndi1, Eno2, Shm1, Arc1, Nop6, Ykl107W, Sse1, Rpl30, Tim9, Rpl11A, Rpl26A, Tif34, Pdb1, Pyc1, Pab1, Rpl13A, Rpl31A, Rpl6B, Pil1, Eno1, Pgk1, Cdc3, Idp1, Mdh3, Pda1, Dur1,2, Rps19A, Rps7A, Rrb1, Rrp9, Srp40, Utp15, Ytm1, Cpr6, Mrt4, Nhp6B, Rpl22A, Rpl22B, Rpp0, Rpp1B, Rps5, gag, Ded1, Lhp1, Nop1, Nsr1, Sbp1, Yra1, Prt1, Ssb1, Cdc11, Gus1, Rpl7A/B, Ssb2, Rrp5.

Secondly, a computational re-analysis of peptide-mass finger printing spectra of proteins purified from *S. cerevisiae* predicted 27 arginine methylated proteins (Pang et al. 2010). Eight of these overlap with the list extracted from (Low et al. 2013). The remaining ones were added to the selection of candidates: Vps52, Yku70, Cdc39, Lea1, Ecm29, Psd2, Rsc1, Cdc19, Rpl18A, Ede1, Rpl27A, Tub2, Rpl2A, Rps11B, Fks1, Nog2, Rpb2, Rsc30 and Tdh3.

Further, a search was conducted for RGG-rich proteins that have not previously been reported to be arginine methylated. UniProt lists eight proteins in *S. cerevisiae* containing an RGG-box (region = “RGG-box”): Nop1, Gar1, Nsr1, Sbp1, Nab2, Dbp2, Pub1 and Ski2 (June 2014). Pub1 and Ski2 are the only ones that had not been described to be arginine methylated and were therefore included in the list of candidates.

Two novel arginine-methylated proteins, Lsm4 and Lge1, from the global discovery approach (“2. Identification of Protein Methylation by Proteomics Approaches,”) and Scd6, a marginal hit in this study with an RG-rich region were also added to test if they may contain further methylation sites. Finally, Npl3, a heavily arginine methylated protein was included as a positive control and Hmt1 to test whether auto-methylation of this PRMT occurs, as has been described for its human homologue PRMT1 (Lakowski et al. 2010). Therefore, the final

candidate list contained 106 proteins. After an initial attempt to detect methylation sites on these proteins, the modification state of two proteins involved in protein translation was more closely investigated.

### 3.1.3. Methylation of Proteins Involved in Translation

Several proteins involved in translation are methylated on various amino acid residues. Ribosomal proteins are methylated in eukaryotes, prokaryotes and archaea (Polevoda & Sherman 2007). In *S. cerevisiae* ribosomal protein methylation includes lysine methylation (Lhoest et al. 1984), arginine methylation on the ω- (Lipson et al. 2010) and δ-position (Zobel-Thropp et al. 1998), histidine (Webb et al. 2010) and N-terminal methylation on proline (Meng et al. 2004). In HeLa cells, arginine methylation was found to be the most prevalent ribosomal modification (Chang et al. 1976) and it was linked to ribosome biogenesis in human and *S. pombe* cells (Ren et al. 2010, p.10; Bachand & Silver 2004). Further, the nucleolar proteins Gar1, Nop1, Nsr1 are heavily arginine methylated, suggesting a role of this PTM in rRNA maturation (Xu et al. 2003).

Lysine methylation of *S. cerevisiae* translation elongation factors EF1A (Cavallius et al. 1993), EF2 and EF3A (Couttas et al. 2012) has been reported and several of the sites were confirmed in this study ("2. Identification of Protein Methylation by Proteomics Approaches"). Bacterial EF-Tu which corresponds to EF1A is also lysine methylated, with this modification leading to a reduction of its t-RNA dependent GTPase activity and potentially an increase in translational fidelity (L'Italien & Laursen 1979; Van Noort et al. 1986). Yeast EF1A is additionally methyl esterified on the C-terminus (Zobel-Thropp et al. 2000). Methylation is also a step in the conversion of histidine to diphthamine in EF2 (Chen & Bodley 1988). Glutamine methylation was observed on release factors RF1 and RF2 in *E. coli* and Mrf1 and Sup45 in *S. cerevisiae* (Heurgué-Hamard et al. 2002; Heurgué-Hamard et al. 2005).

Ded1 is a protein important for unwinding RNA to allow translation initiation and is methylated on several arginines (Chuang et al. 1997; Low et al. 2013; this study). Further knowledge about potential methylation of translation initiation factors is lacking to date. It has however been shown that a protein with an RG-rich region, Scd6, can function in the inhibition of translation initiation by association with the cap-binding complex and that this process is dependent on the RG-rich region (Rajyaguru et al. 2012, p.6) (see also "5. A Role of Scd6-Methylation Dependent Translational Repression upon Cell Stress").

As methylation of proteins linked to translation occurs on various residues and throughout phylogenetically widely separated organisms, detailed follow-up experiments after initial indications of methylation were conducted on the methylation of two proteins involved in translation initiation, Scd6 and Cdc33. Cdc33 (eIF4E) is a member of the mRNA cap-binding complex and a protein that had been pulled down with an anti-methyl arginine antibody (Low et al. 2013). The lower sample complexity in this experiment allowed additionally considering modifications other than lysine and arginine methylation.

### 3.1.4. Methylation on Residues other than Lysine and Arginine

As mentioned in the previous section, methylation of the N-terminus of proteins, glutamine and histidine has been observed, i.e. protein methylation is not limited to lysine and arginine. A potential reason for the predominant discovery of lysine and arginine methylation might even be attributed to an experimental artefact, as much of the initial discovery of protein methylation is based on amino acid analysis (Gary et al. 1996), i.e. requires harsh treatment in acidic conditions to which lysine and arginine methylations prove more resistant than methyl-esters. While the existence of glutamate methylation is well established in bacteria (Morgan et al. 1993), the identification of aspartate and glutamate methylation has also been reported repeatedly in eukaryotes in the last decade: For example, Sprung (2008) performed a bottom-up proteomics based screen for aspartate and glutamate methylation and detected five and

three proteins modified in this way in HeLa and *S. cerevisiae* cells respectively. Low et al. (2014) observed five methylated glutamates in purified *S. cerevisiae* poly-A binding protein and pointed out their preferential occurrence in an EE or DE context. Interestingly, enzymatic O-methylation also occurs on isoaspartate (D-aspartate). It is considered a mechanism for the recognition of this form of protein damage and exists in both pro- and eukaryotes (Johnson et al. 1987; Li & Clarke 1992). Furthermore, carboxy-terminal protein methylation has been described in eukaryotes, especially, but not only, of isoprenylated cysteines (Xie & Clarke 1993). Finally, asparagine (Swanson & Glazer 1990) and cysteine (Lapko et al. 2005) methylation have also been observed. Therefore, after an initial screen for arginine methylation, the possibility of methylation on other amino acids was considered for a subset of the candidate proteins.

### 3.2. Methods

#### Preparation of Yeast Strains

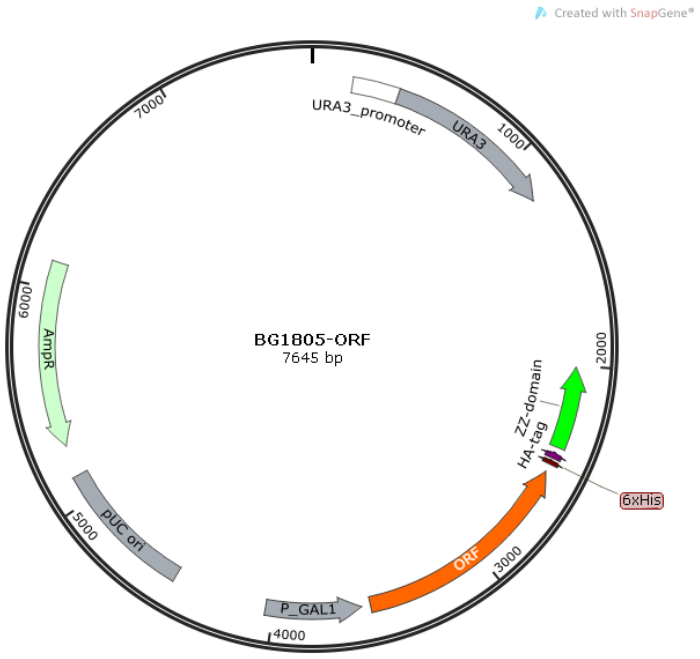



***Figure 3.1: Plasmid map of BG1805.*** *The plasmid is shown with annotation of features relevant to this study and an ORF insert of about 1000 bp. AmpR: ampicillin resistance gene; pUC ori: origin of replication; P_GAL1: GAL1 promoter*

The Movable Open Reading Frame (MORF) collection is a library of about 5,500 BG1805 plasmids with different open reading frames of the *S. cerevisiae* genome cloned as inserts

(Open Biosystems, YSC3868). The open reading frames are inserted as a Gateway-cassette (Hartley et al. 2000) under the GAL1-promoter and C-terminally tagged with at tandem-affinity purification tag of 6xHis, HA and the ZZ-domain of protein A (Fig. 3.1) (Gelperin et al. 2005).

Selected *E. coli* strains harbouring plasmids of the MORF collection were streaked onto LB agar plates containing 100 µg/ml ampicillin and grown at 37 ˚C overnight. Plasmids were purified after inoculation of 5 ml LB containing 100 µg/ml ampicillin and growth at 37 ˚C overnight using a QIAprep Spin Mini-prep kit (Qiagen, 27104) following the manufacturer`s instructions. For long term storage, bacterial strains were kept in terrific broth (Sigma, T5574) + 8% glycerol. Plasmids were transformed into *S. cerevisiae* BY4741 cells (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) using a yeast transformation kit (Sigma, YEAST-1) according to the manufacturer`s instructions. Transformed cells were plated on SC-Ura + 2% glucose (Glc) plates and single colonies, once visible, re-streaked onto SC-Ura + 2% Glc. For long term storage, transformed strains were kept in SC-Ura + 15% glycerol at -80 ˚C.

**Protein Expression and Purification**

Transformants were inoculated into 5 ml SD-ura with 2% Glc and 300 µg/ml G418. After growth at 30 ˚C and 180 rpm overnight, cultures were diluted into 25 ml SC-Ura with 2% raffinose and grown for eight hours at 30 ˚C. They were further diluted to into 300 ml light or heavy (containing methionine-[methyl-$^{13}CD_3$] instead of methionine) SC-Ura with 2% raffinose to an $OD_{600}$ of 0.02 and grown as above till an $OD_{600}$ of 0.8-1.2 was reached. At this point 150 ml SC-Ura (light or heavy as appropriate) with 6% galactose were added, followed by growth for 6 h. Cultures were harvested by centrifugation, washed with 3 ml ice-cold water and distributed to two 2 ml lysis tubes containing glass beads tubes. The tubes were spun down and pellets snap frozen in liquid nitrogen and stored at -80 ˚C.

Cells were resuspended in one pellet-volume cold TST buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% Tween 20 with 2 Complete Mini protease inhibitor tablets per 10 ml, 1 mM phenylmethylsulfonylflouride, 4 µg/ml leupeptin, 2.5 µg/ml pepstatin and 0.1% (v/v) benzonase). Cells were lysed by bead beating 10 times at 6,000 rpm for 30 s each, with 1 min on ice in between. The following steps were performed on ice or at 4 ˚C. Samples were centrifuged at 20,000 g and 4 ˚C for 10 min, before lysates corresponding to 225 ml culture were combined. If MethylSILAC was employed, samples were mixed in a 1:1 protein ration according to a BCA assay. The samples were adjusted to 1 mM EDTA, 4 mM $MgCl_2$, 10% glycerol and 0.75 M NaCl, incubated for 10 min, diluted with 5 volumes 10 mM Tris-Cl pH8 with 0.1% (v/v) NP40, 1 mM DTT and protease inhibitors as in the lysis buffer. The samples were centrifuged as above, while affinity-chromatography columns were prepared: 200 µl IgG Sepharose 6 Fast Flow (GE, 17-0969-01) was packed in chromatography columns, washed with 2.5 ml TST, 1.5 ml 0.5 M acetic acid pH 3.4, 1.5 ml TST, 1.5 ml 0.5M acetic acid pH 3.4 and at least 10 ml TST to reach a pH of 7-8. Columns were rotated with sample for 30 min, drained, washed with 5 ml TST and 1 ml 5 mM ammonium acetate pH 5 and proteins eluted with 1 ml 0.5 M acetic acid pH 3.4. The eluate was dried in a vacuum centrifuge and the resulting pellet resuspended in 15 µl 8 M urea or concentrated through VIVA-Spin 500, 5000 MWCO PES filters (Sigma, Z614009) before being subjected to separation on gel.

**Mass spectrometry and data analysis**

SDS-PAGE, in-gel digestion with trypsin and elastase and LC-MS/MS analysis were performed as described in section 2.2 with the following alterations: MS2 and MS3 spectra were acquired in the orbitrap at a resolution of 30,000. For the acquisition of MS2 spectra, fragmentation was performed using HCD at a normalised collision energy of 35% unless stated otherwise. For the generation of MS3 spectra, MS1-precursors were fragmented with CID and MS2 fragment ions with HCD.

MS data analysis using the Mascot search engine was performed as described in section 2.2, using carbamidomethylation of cysteine as a fixed and methionine oxidation and arginine mono- and di-methylation as variable modifications. For MethylSILAC samples, searches were additionally performed using carbamidomethylation of cysteine and a $^{13}CD_3$-label of methionine as fixed and heavy mono- and di-methylation of arginine as variable modifications. The MethylQuant software was used as in section 2.2. The Preview software (Protein Metrics) was used at default settings to estimate the mass accuracy of acquired spectra.

For data analysis with the Peaks software (Bioinformatics Solutions), MS/MS spectra were grouped according to fragmentation method (CID, HCD or ETD) and fragmentation level (MS2 or MS3) and stored in individual .mgf files using the Proteome Discoverer (Thermo) 'spectrum selector' node. In Peaks, precursor mass correction was enabled and a *de novo* search performed using a 10 ppm precursor and 0.05 Th fragment mass tolerance unless indicated otherwise. Carbamidomethylation on cysteine was selected as a fixed and oxidation of methionine as a variable modification. A database search (PeaksDB) (J. Zhang et al. 2011) was conducted against *S. cerevisiae* entries of the SwissProt database using the same tolerance and modification settings as for the *de novo* search. 'Trypsin' was specified as a protease for the analysis of tryptic digests and 'no enzyme' for elastase digests. When searching MS3 spectra, C-terminal dehydration was selected as an additional variable modification, as b-ions resemble a dehydrated peptide. Finally, the PeaksPTM (Han et al. 2011) search was performed against all built-in modifications, which comprise all natural modifications and mutations in UniMod (Creasy & Cottrell 2004) and include C, H, K ,N, Q, R, I, L, D, E, S, T, N-terminal and C-terminal mono-methylation and K, R, N, P and N-terminal di-methylation.

## 3.3. Results

### 3.3.1. Arginine Methylation Sites Detected by Tryptic Digestion of Candidate Proteins

To assess the methylation state of individual proteins, a list of 106 candidate proteins was selected. Of these, 33 candidate proteins were not part of the yeast movable open reading frame library (Gelperin et al. 2005), so that 73 plasmids were obtained. Proteins were expressed and purified according to a protocol adapted from the original description of the collection (Gelperin et al. 2005) to the equipment available in our laboratory and the purpose of the experiment. When concentrating eluates from column purification a striking advantage of using vacuum centrifugation over ultrafiltration was observed (Fig. A3.1 in the appendix).

The target protein could be purified in sufficient quantity to be visible as a Coomassie stained gel band from 38 of the yeast strains transformed with expression plasmids: The proteins purified were Hmt1, Mam33, Ses1, Rim1, Mbf1, Dbp6, Lat1, Rpp2B, Nap1, Nop56, Hsc82, Cdc33, Rfa1, Imd4, Pwp1, Ent5, Cdc12, Crp1, Kgd2, Lpd1, Ndi1, Eno2, Shm1, Arc1, Nop6, Ykl107w, Sse1, Rpl30, Tif34, Pil1, Eno1, Pgk1, Cdc11, Lea1, Ede1, Tub2, Lsm4, Scd6.

On 26 of these proteins arginine methylation sites with E-values < 0.05 in a database search were detected (Table 3.1). The majority (19) of these methylated proteins had been derived from the study by Low et al. (2013) while only four (one overlapping with the previous) had come from the study by Pang et al. (2010) (Table 3.1). Additionally, the mono-methylation of R119 on Lsm4, which had been found by a global approach could be confirmed and di-methylation on the same site was additionally detected. No further methylation sites were identified on the protein with a sequence coverage of 72%. Three di-methylation sites were detected on Scd6 (R18me2, R166me2, R264me2). Interestingly, six methylation sites were found on the PRMT Hmt1, with two of them both in mono- and di-methylated state (R42me2, R113me2, R247me1 and me2, R269me1 and me2, R317me2, R337me2).

**Table 3.1:** Methylated proteins found by the candidate approach: Arginine methylation sites are given with the corresponding E-value according to a Mascot search

| Gene symbol | Systematic Name | Reference | methylation sites (Mascot E< 0.05) | | | | |
|---|---|---|---|---|---|---|---|
| HMT1 | YBR034C | | R42me2; 0.011 | R113me2; 0.029 | R247me1; 0.0035 | R247me2; 0.00048 | R269me1; 0.014 |
| | | | R269me2; 0.0071 | R317me2; 0.015 | R337me2; 0.0011 | | |
| MAM33 | YIL070C | Low, 2013 | R52me2; 0.021 | R101me2; 0.00036 | | | |
| SES1 | YDR023W | Low, 2013 | R46me2; 0.0033 | R379me1; 0.0029 | | | |
| RIM1 | YCR028C-A | Low, 2013 | R42me2; 0.013 | R42me1; 0.041 | | | |
| LAT1 | YNL071W | Low, 2013 | R260me1; 0.011 | R321me2; 0.0014 | | | |
| RPP2B | YDR382W | Low, 2013 | R38me2; 8.6E-05 | | | | |
| NAP1 | YKR048C | Low, 2013 | R73me2; 0.00023 | R135me2; 0.0032 | R137me2; 0.028 | R234me1; 3.1E-6 | |
| HSC82 | YMR186W | Low, 2013 | R46me2; 0.0056 | R78me2; 0.0061 | R189me2; 0.0011 | R196me1; 0.0018 | R275me1; 0.026 |
| | | | R314me2; 0.00031 | R331me1; 0.0013 | R331me2; 0.0035 | R343me2; 0.03 | R376me1; 0.0028 |
| | | | R376me2; 0.00028 | R459me1; 5.4E-06 | R459me2; 6.2E-07 | R536me2; 2.9E-06 | R587me2; 1E-07 |
| | | | R587me1; 0.028 | R663me2; 0.00046 | | | |
| CDC33 | YOL139C | Low, 2013 | R101me2; 9.7E-05 | R101me1; 0.0051 | | | |
| RFA1 | YAR007C | Low, 2013 | R20me1; 0.037 | R35me1; 0.0002 | R35me2; 1.2E-06 | R91me2; 0.0037 | R109me2; 2E-07 |
| | | | R182me2; 0.0018 | R216me1; 0.0062 | R234me2; 4.2E-07 | R335me2; 8.8E-05 | R504me1; 0.00025 |
| | | | R504me2; 0.00055 | R514me1; 0.0073 | R578me1; 0.0016 | R578me2; 0.027 | R582me2; 0.00025 |
| | | | R591me2; 0.00024 | R593me1; 0.00024 | R593me2; 0.011 | R605me1; 5.8E-05 | R605me2; 0.00022 |
| PWP1 | YLR196W | Low, 2013 | R323me2; 0.00099 | R510me2; 0.0074 | | | |
| CDC12 | YHR107C | Low, 2013 | R195me2; 0.027 | | | | |
| KGD2 | YDR148C | Low, 2013 | R233me2; 4.7E-05 | | | | |
| SHM1 | YBR263W | Low, 2013 | R106me1; 3.1E-05 | R215me1; 0.00014 | R215me2; 0.0045 | | |
| ARC1 | YGL105W | Low, 2013 | R242me2; 9.2E-5 | | | | |
| YKL107W | YKL107W | Low, 2013 | R260me2; 2.7E-05 | R81me1; 0.0013 | | | |
| SSE1 | YPL106C | Low, 2013 | R639me2; 0.0044 | | | | |
| PIL1 | YGR086C | Low, 2013 | R11me2; 0.0029 | R44me2; 0.019 | R56me2; 0.00021 | R133me2; 0.00086 | R175me2; 0.00057 |
| | | | R192me2; 8.1E-05 | | | | |
| ENO1 | YGR254W | Low, 2013 | R9me2; 0.035 | | | | |
| PGK1 | YCR012W | Low, 2013 | R319me2; 0.023 | | | | |
| PRT1 | YOR361C | Pang, 2010 and Low, 2013 | R28me1; 0.00011 | R28me2; 0.012 | R165me2; 0.015 | R176me1; 0.0082 | R176me2; 0.00059 |
| | | | R199me1; 0.00046 | R228me2; 2.3E-06 | R237me1; 0.00049 | R266me1; 0.023 | R350me1; 4.8E-06 |
| | | | R350me2; 0.0012 | R394me2; 0.0043 | R472me1; 3.2E-06 | R537me1; 0.00019 | R559me1; 0.00079 |
| | | | R559me2; 2.2E-06 | R573me2; 0.0035 | R620me2; 0.0038 | R663me1; 0.026 | R681me2; 0.00053 |
| | | | R707me1; 0.028 | R717me1; 0.0035 | | | |
| IMD3 | YLR432W | Pang, 2010 | R5me1; 0.00025 | R19me2; 0.031 | R35me2; 0.01 | R143me1; 0.00016 | R157me1; 0.0023 |
| | | | R168me2; 0.044 | R227me2; 7.3E-07 | R263me2; 0.0091 | R311me2; 0.0019 | R326me1; 0.017 |
| | | | R437me2; 0.0029 | R520me1; 9.1E-05 | R520me2; 0.0064 | | |
| EDE1 | YBL047C | Pang, 2010 | R937me2; 0.00019 | R947me2; 0.0061 | | | |
| TUB2 | YFL037W | Pang, 2010 | R77me2; 0.00013 | | | | |
| LSM4 | YER112W | Plank, 2015 | R119me1; 5.2E-5 | R119me2; 0.049 | | | |
| SCD6 | YPR129W | | R18me2; 0.011 | R166me2; 0.0012 | R264me2; 0.031 | | |

For several of the sites additional observations increased the confidence in their identification: These included the presence of MS/MS spectra matching to the same peptide in different charge states or with oxidized and unmodified methionine. Also, identification of the mono- and di-methylated peptide and the detection of different peptide sequences harbouring the same methylation site provided additional confidence. Additionally, in several cases, the spectrum of the unmodified peptide was detected with a fragmentation pattern that resembled the one of the modified peptide. It should however be noted that while these observations corroborate the presence of +14.02 or +28.03 Da modifications, they generally did not provide additional information about the PTM localisation (see fig. 3.2 for an example).

Manual inspection of fragment spectra indicated that methylation sites were rarely unambiguously localized to a residue.

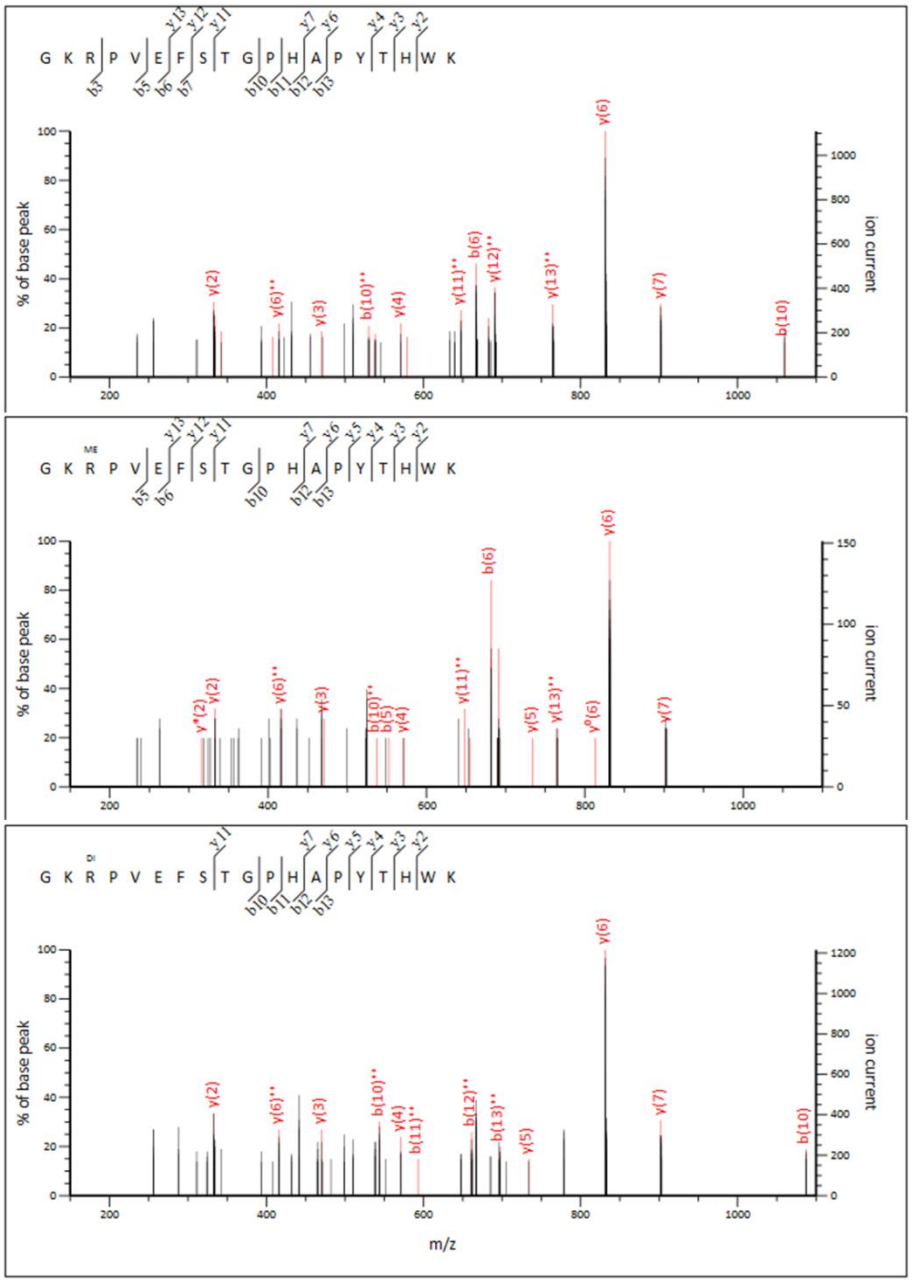



***Figure 3.2: Corresponding spectra of un-, mono- and di-methylated form of a peptide.*** *The spectrum assigned to the un- (top), mono- (middle) and di- (bottom) methylated version of peptide* $_{267}$*GKRPVEFSTGPHAPYTHWK*$_{285}$ *of Hmt1 are shown. Singly and doubly charged b- and y-ions matching to the sequence are annotated in red.*

### 3.3.2. Use of MethylSILAC to Distinguish True Methylations against Artefacts

Using a candidate approach, methylation sites at a Mascot E-value < 0.05 were identified on 26 proteins. However, the localisation of the methylation site was ambiguous in several cases.

Importantly, this may allow aspartate and glutamate methylation and ethylation, which can arise as artefacts from the sample preparation, to be mistaken with arginine mono- and di-methylation, respectively. While the use of methanol was avoided in the sample preparation, so that artificial methylation is not expected, ethanol was present in the gel staining solution. A way to avoid the erroneous identification of ethylation as di-methylation without the need to improve the quality of fragment spectra is to utilize MethylSILAC (see "2.1.3. MethylSILAC"). This is because heavy mono- and di-methylation (+18.04 and +36.08 Da respectively) are not expected to arise as artefacts in sample preparation.

Six proteins from the candidate list were randomly selected for the identification of methylation sites using the MethylSILAC strategy: Lge1, Lat1, Rim1, Lsm4, Cdc12 and Pub1. Three methylation events were identified on Lge1 (at $E < 0.05$): R30me1, R30me2 and R39me1 yielded identifications in the heavy sample and, while only R30me1 was also identified in the light sample, heavy-light pairs were found for all of them with heavy to light ratios of between 1 and 2. Additionally, R277me1 was only identified in the light sample and not detected as a heavy-light pair. It is therefore unclear if the latter may have arisen as an artefact, despite the absence of methanol in the sample preparation, while the former appear to be genuine methylation sites. Similarly, for R260me1 on Lat1, R119me1 and R119me2 on Lsm4 and R260me1 on Pub1 the heavy versions of the respective peptides were identified, indicating that these sites are not artefacts from sample preparation.

### 3.3.2. Detection of Additional Methylation Sites on Scd6 by Digestion with Elastase

The second limitation of the candidate approach was addressed next. The digestion of proteins by trypsin for LC-MS/MS often results in incomplete sequence coverages as only certain peptides produced by this highly specific protease will be suitable for mass spectrometric analysis, while others may for example be too long or too short (e.g. less than six or more than 30 amino acids). For low complexity samples, an alternative protease like elastase can be

employed, which due to its lower sequence specificity will result in a larger number of different peptides (Schlosser et al. 2002).

Scd6 was chosen as a target to test if digestion by elastase would result in broader sequence coverage and increased number of methylation sites identified. As the cleavage sites of this protease are not known *a priori*, a data analysis strategy based on *de novo* sequencing via the Peaks software was employed for data analysis. This allowed querying methylations on residues other than arginine at the same time. Sequence coverage of 83% was achieved by this method, which included methylation sites on five arginines (R288, R292, R294, R298 and R301) of the RG-rich region, as well as arginines 18 and 39 (Fig. 3.3). Additionally, methylations on other amino acids were identified (at -10 log(P) > 15) which require further validation, similar as shown for Cdc33 below.

### 3.3.3. Increased Sequence Coverage and Additional Potential Methylation Sites on Cdc33 by Digestion with Elastase

An initial experiment following a standard approach for the identification of methylation sites from a purified protein with tryptic digestion and a database search using Mascot, allowing oxidation of methionine and mono- and di-methylation of arginine as variable modifications identified protein Cdc33 with 49% sequence coverage. Mono- and di-methylation of R101 were identified at E < 0.05 (see "3.3.1. Arginine Methylation Sites detected by Tryptic Digestion of Candidate Proteins").

It was decided to next digest Cdc33 with elastase, a protease of lower sequence specificity than trypsin (Wang et al. 2008) to obtain higher sequence coverage and analyse the data via the Peaks software, allowing all 386 built-in modifications to be considered, which included mono-methylation on C, H, K, N, Q, R, I, L, D, E, S, T, the N- and C-terminus (Han et al. 2011).

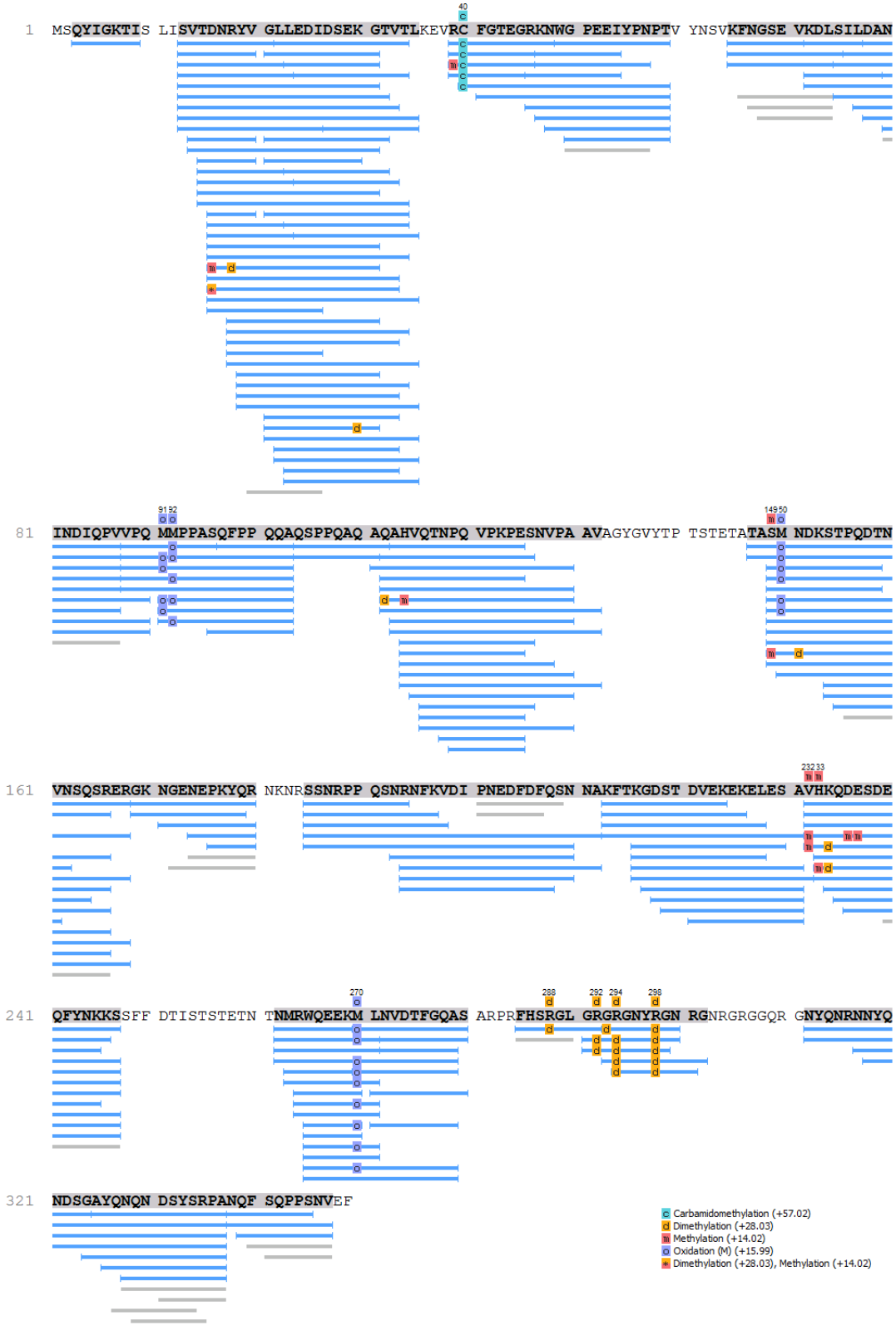



***Figure 3.3: Sequence coverage of Scd6 by elastase digest.*** *Scd6 was overexpressed in the BY4741 strain, purified, subjected to digestion with elastase and analysed by LC-MS/MS. Data were analysed using Peaks. Identified peptides are shown as blue bars and the corresponding covered sequence is bold and highlighted in grey. Modifications shown in the legend are displayed on the bars and, if the intensity of MS/MS peaks confirming these identifications was above 5% of the basepeak, also above the sequence. Grey bars indicate sequences only found by the de novo search.*

This search resulted in 99% sequence coverage of Cdc33 and ten mono-methylation sites at a cut-off of -10logP > 15 (Table 3.2). These included two cases of peptide C-terminal modifications of amino acids that are not expected to be methylated metabolically (V52 and A108). They were nonetheless included in the further analysis, as they may represent mislocalisations. Also, five sites with +28.03 Da modifications on glutamates or aspartates were detected. However, as ethanol was not avoided in the sample preparation, these may constitute ethylation rather than di-methylation events and were therefore not pursued further. The modifications on E107 and D92 (in one of two spectra) appeared as +71.04 Da mass shifts which were interpreted as simultaneous methylation and carbamidomethylation by the analysis software. Most identified mono-methylations were located on glutamate, but some also on serine, threonine, asparagine and aspartate. In most cases the modified amino acid was found supported by directly neighbouring fragment ions and most spectra showed a series of at least four consecutive b- or y-ions at a matching tolerance of 0.05 Da. However, often low abundance ions were used for scoring and some major peaks remained un-annotated, so that further confirmation was sought to support these unusual modifications (See fig. 3.4 for two examples of fragment spectra.). Note that the E107 methylation in this study and R101 methylation identified in "3.3.1. Arginine Methylation Sites detected by Tryptic Digestion of Candidate Proteins" localize to the same peptide in both cases so that therefore one may be a misassignment of the other.

An alternative Cdc33 peptide match to the spectrum yielding the sequence WYTKPAV(+14.02) exists: LWYTKPA. Interestingly, while there is a large sequence overlap between the peptides, one would not expect them to share fragment ions, as they do not have any terminus in common. However, the $b_1$-ion derived from W is very similar in mass to the $y_2$-ion from PA (0.02 Da difference) and similarly the $b_4$ (W+YTK) and $y_5$-ion (PA+YTK) are near isobaric, which explains why the same spectrum could yield these

different interpretations. This posed the question whether a more restrictive MS2 tolerance could be used to avoid false positives.

**Table 3.2.** Cdc33 peptides identified as mono-methylated at -10logP > 15 via the Peaks software

| Position | Modified Peptide | -10logP | ID at 0.01 Th MS2 tol. | Plausible alternative* |
|---|---|---|---|---|
| E11, E12 | SKKFE(+14.02)E(+14.02)NV | 24.39 | n | |
| T41 | HFDVKH(+57.02)PLNT(+14.02) | 32.49 | n | |
| V52 | WYTKPAV(+14.02) | 33.54 | n | |
| E56 | KPAVDKSE(+14.02)SWSDLLR | 23.98 | n | |
| S91 | S(+14.02)DYHVFR | 38.65 | y | |
| D92 | D(+57.02)(+14.02)YHVFR | 30.81 | y | |
| D92 | SD(+14.02)YHVFR | 37.21 | y | |
| E107 | DYHVFRNDVRPEWEDE(+14.02)(+57.02) | 22.84 | y | y |
| S201 | LKLTDDGHLEFFPH(+57.02)SS(+14.02) | 24.78 | y | y |
| S201 | KLTDDGHLEFFPH(+57.02)SS(+14.02) | 30.65 | y | y |
| N203 | DDGHLEFFPHSSAN(+14.02)(+0.98) | 15.80 | y | |

*Please refer to "3.3.4. Increased Confidence through alternative Fragmentation Regimes"

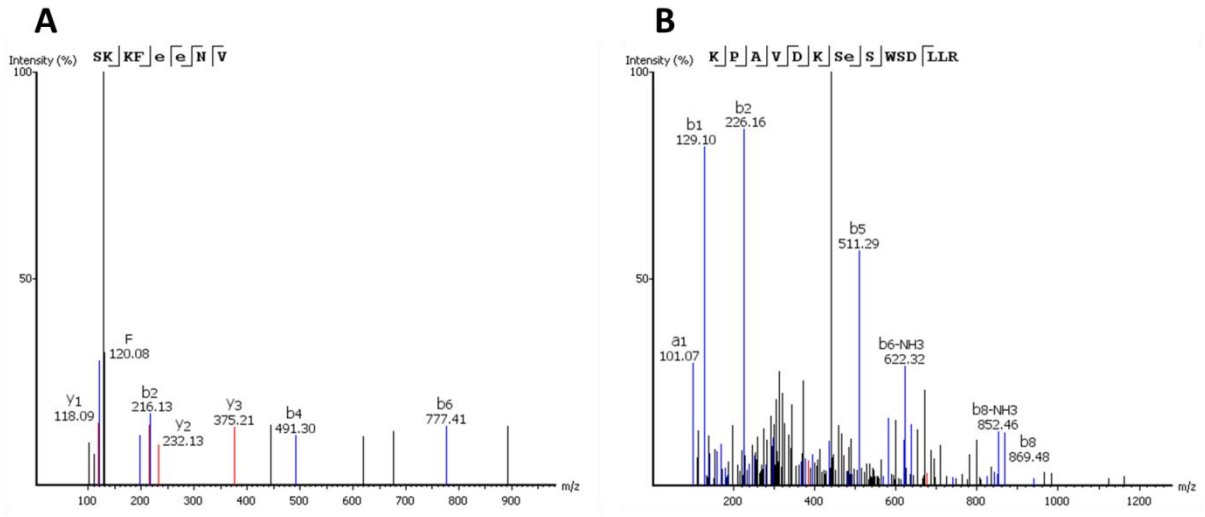



***Figure 3.4: Example spectra of peptides from a Cdc33 digest with elastase and analysed by Peaks.*** *Cdc33 was overexpressed in S. cerevisiae, purified via its ZZ-domain tag and in-gel digested. Peptides were subjected to mass spectrometry and data analysed with the Peaks software, returning identifications* $_{7}$*SKKFE(+14.04)E(+14.04)NV*$_{14}$ *(A) and* $_{49}$*KPAVDKSE(+14.04)SWSDLLR*$_{63}$ *(B). Fragment peaks are shown in red if assigned to y-ions and blue if assigned to a- or b-ions.*

As it is not straightforward to extract delta masses of MS2 assignments from the outputs of spectral matching software, the Preview software (Protein Metrics) was used for this purpose, which stated a median fragment accuracy of 0.0006 Th. As a value for the spread of this estimate was not given, a conservative threshold of 0.01 Th was used for a new analysis with the Peaks software. Previously observed peptides containing methylation sites on E11, E12, T41, V52 and E56 were not identified (-10logP > 15) at these settings (Table 3.2).

### 3.3.4. Increased Confidence through Alternative Fragmentation Regimes

Identifications for methylation sites on the following residues remained after lowering the MS2 matching tolerance: S91, D92, E107, S201 and N203 (Table 3.2). To further test if these constituted correct identifications, alternative fragmentation methods (CID and ETD), energies (normalised collision energies 28% and 42%) and additional acquisition of MS3 spectra were applied. As the application of alternative fragmentation settings on the same precursor will inevitably lead to increased cycle times, an inclusion list was generated from the peptides found methylated in the previous run.

Glutamate 107 had been found methylated in the peptide DYHVFRNDVRPEWE DE(+14.02,+57.02) by HCD fragmentation at a normalised collision energy (NCE) of 35. An alternative interpretation, DYHVFRNDVRPEWEDEA, was provided by HCD fragmentation with lower collision energy, as well as CID and ETD fragmentation (Fig. 3.5). As the modification masses in the former add up to 71.04, the mass of alanine, and the peptide is followed by this residue, the latter provides a reasonable alternative. Abundant $b_{16}^{++}$ and $c_{16}^{++}$ ions that distinguish between the two interpretations in favour of the unmodified version were observed (red arrows in fig. 3.5 B). MS3 fragmentation of the $b_{14}^{++}$ and $b_{15}^{++}$ ions of the CID spectrum contributed no further information to distinguish the two possibilities. Interestingly, the highly abundant, but unannotated peak at 696.65 Th (green arrow in fig. 3.5 C) was identified as $b_{16}^{+++}$ ion by MS3. The additional fragmentation step therefore provided a

reasonable explanation for the unassigned peak and indicated that searching for triply charged fragment ions, which is not performed by the software, can be useful to annotate abundant peaks in an MS2 spectrum.

A similar situation was encountered for peptides containing methylation of S201, where alternative hits for the same spectra without modifications, but instead an additional alanine at the C-terminus were observed when employing alternative fragmentation methods. In the case of D92, the methylation had been identified with a simultaneous carbamidomethylation on the same residue (D(+57.02)(+14.02)YHVFR). However, neither the residue C- nor N-terminal to this sequence was alanine. Additionally, the methylation on this residue was also found in the match of another spectrum to the non-carbamidomethylated peptide (D(+14.02)YHVFR), indicating that this assignment is likely to be true.

No further indications against or in favour of S91 and N203 methylation were obtained.

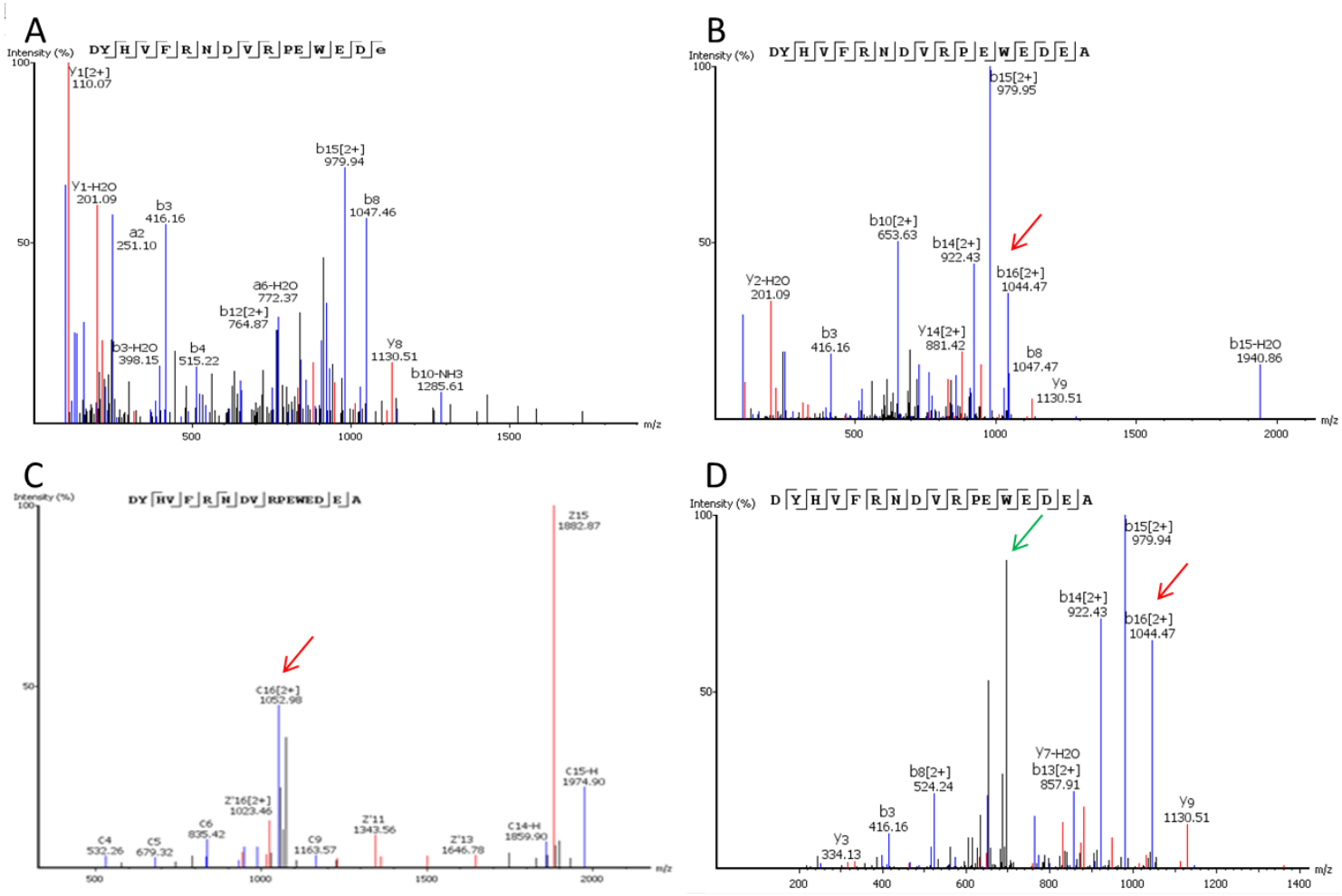



***Figure 3.5, Fragment spectra for the same precursor obtained with different fragmentation methods.*** *Cdc33 was overexpressed in S. cerevisiae, purified and in-gel digested. In the subsequent LC-MS/MS analysis, the same precursor was subjected to fragmentation by HCD with a normalized collision energy of 35% (A) or 28% (B), ETD (C) and CID (D). Matches to the sequence DYHVFRNDHDVRPEWEDE(+57,+14) (A) or DYHVFRNDHDVRPEWEDEA (B, C, D) were returned by the Peaks software. The red arrows indicate* $b_{16}^{++}$ *or* $c_{16}^{++}$ *ions and the green arrow a* $b_{16}^{+++}$ *ion (identified via MS3) distinguishing between the assignments. Red: y- (A, B, D) and z-ions (C); blue: b- (A, B, D) and c-ions (C).*

## 3.4. Discussion

### 3.4.1. Identification of Arginine Methylation Sites by Tryptic Digestion of Purified Proteins

As the results obtained from a global analysis of methylation sites are potentially limited by enrichment bias and insufficient sensitivity, 106 candidate proteins were selected to assess for methylation. Of these, 38 were successfully overexpressed and purified and methylation sites were mapped to 26 proteins by a database search considering arginine mono- and di-methylation. For several of the assignments, not only a low E-value, but also additional indications of true identifications, like observation of the peptide in different charge states or with and without oxidized methionine could be obtained. Note that at a low overall false discovery rate a random false positive match to the single protein of interest in each sample is unlikely, while the likelihood of an incorrect localisation of the modification site is not reduced by low sample complexity. Therefore, two main issues remained to be addressed in the identification of methylation sites via a candidate approach: a, Protein sequence coverages achieved were generally far from complete and this was likely due to the nature of some of the tryptic peptides, which were suboptimal for analysis by LC-MS/MS. b, The localisation of the methylation site on a peptide may be ambiguous, especially when considering that methylations may also occur on residues other than arginine. Importantly, a potential danger from insufficiently localized modification sites is the confusion of aspartate and glutamate methylations and ethylations, which may occur as artefacts of sample preparation (Jung et al. 2008), with arginine mono- and di-methylations. To test whether such artefacts may be a main reason for the sites identified, six randomly selected proteins were analysed with the use of MethylSILAC. Only for one of seven identified methylation sites a heavy-light pair was not detected, which may therefore constitute an artefact. While this indicates that the majority of sites identified are likely not to be artefacts, this needs to be confirmed for each di-methylation site individually.

Assuming, the majority of methylated proteins pass this additional validation, the efficiency of identification, with 26 out of 38 successfully purified proteins observed to be methylated, is remarkable. On the one hand this may point towards a high prevalence of arginine methylation in the *S. cerevisiae* proteome and therefore severe limitations of the proteome wide approach for the discovery of methylation sites ("2. Identification of Protein Methylation by Proteomics Approaches"). On the other hand, it may indicate that the experiments considered for generating the list of candidates, especially the study by Low (2013), towards which the set of purified proteins is skewed, indeed mainly reported methylated proteins.

It is noteworthy that R119 on Lsm4 which had been detected in mono-methylated state by the global approach was additionally found di-methylated on the purified protein. Overall, di-methylation was observed more frequently in the dataset than mono-methylation. This suggests that the more successful identification of arginine mono-methylation over di-methylation sites by proteome-wide approaches was due to reagent-, rather than sample-bias (see "2. Identification of Protein Methylation by Proteomics Approaches").

Another interesting discovery was the detection of six methylation sites on the PRMT Hmt1. This is important in the light that its human homologue, PRMT1, has been reported to undergo automethylation (Lakowski et al. 2010). Considering that Hmt1 is the predominant MTase in *S. cerevisiae*, automethylation is more likely than methylation by another MTase, however this warrants experimental testing. Note however, that in an *in vitro* assay no evidence for Hmt1 automethylation was detected (see "4.3.5. Hmt1 Methylates Scd6 *in vitro*", fig. 4.6, column 9). The fact that six methylation sites were detected on the protein, two of which were found both mono- and di-methylated, renders it unlikely that all these identifications are false positives.

### 3.4.2. Arginine and Potential Non-Arginine Methylation in Translation Initiation

Ribosomal proteins and translation factors are methylated in both prokaryotes and eukaryotes on various amino acids. This hints at an important role of protein methylation in translation. Two proteins involved in translation initiation were further examined for the presence of methylation sites here: Cdc33 (eIF4E) is a translation initiation factor that had been immunoprecipitated with an anti-methyl arginine antibody in a previous study (Low et al. 2013). Scd6 is a repressor of translation initiation, a process that is dependent on its RG-rich region from about amino acids 288 to 311 (Rajyaguru et al. 2012), and had been found to be methylated with the proteome-wide approach in this study.

Sequence coverage of 55% had been achieved for Scd6 with tryptic digestion, which however only included two (R289 and R292) of the eight arginines of the RG-rich region, probably due to the small size of the peptides originating from tryptic cleavage in this sequence. No arginine methylated residue was detected in the RG-rich region. One residue on Ccd33, R101, had been found both mono- and di-methylated with this approach while 49% of the protein sequence was covered.

To detect methylation sites that had been missed due to the limited sequence coverages achieved with trypsin, an additional digest of proteins Scd6 and Cdc33 with the lower specificity protease elastase was performed. As no method currently exists to predict its cleavage sites, a large number of *in silico* peptides would have to be considered in a database search against data generated with this enzyme. The data were therefore acquired with high mass accuracy and the data analysis based on a strategy supported by *de novo* sequencing. In this analysis, sequence tags were first assigned to spectra based on the m/z differences between abundant peaks, before the spectra were matched against *in silico* spectra of peptides containing the corresponding tag-sequence. Due to the lower number of spectra evaluated in the database search, methylation on amino acids other than arginine could be considered

without an extensive increase in the search space that would render classical database searching impossible.

This experiment led to the identification of peptides corresponding to 83% of the sequence of Scd6 and covering five arginines of the RG-rich region (Fig. 3.3). Four of them were detected in di-methylated state. Sequence coverage for the remaining three arginines followed by glycine in this region was not obtained, i.e. the actual number of methylation sites may be higher than four. Peptides with methylation sites on aspartate, asparagine, glutamine, serine and histidine were also observed, however these need additional validation.

For Cdc33 the analysis returned ten mono-methylation sites, none of which was located on arginine. While data analysis by database search (e.g. using Mascot) with a limited number of allowed modifications may force methylation sites onto incorrect positions, the strategy employed here with an extended number of allowed modifications is likely to result in a higher number of false positive identifications due to an increased search space. This was not corrected for by target-decoy FDR estimation as the number of hits is too low when analysing an individual protein.

It was noted that it is helpful to perform data acquisition with high mass accuracy and analysis with narrow tolerances, as limiting the fragment tolerance from 0.05 to 0.01 Th reduced the number of methylation sites on Cdc33 to five. The use of additional fragmentation techniques, energies and the acquisition of MS3 spectra allowed identifying two further methylation sites as likely false positives. These sites had been assigned to peptides with additional carbamidomethylation. Further analysis indicated that the true peptide is unmodified, but contains an additional alanine on one terminus which is exactly isobaric to the sum of a methyl- and carbamidomethyl-group. While it is reasonable in future analyses to disregard methyl-groups that are only identified with additional carbamidomethylation on peptides that are adjacent to alanine in the protein sequence, additional ways of fragmentation can

nonetheless convey important information. For example MS3 fragmentation of a major peak in a MS2 spectrum revealed this peak to be a fragment ion which had not been assigned due to its triply charged nature.

It is likely that, with the advent of high resolution, high mass accuracy mass spectrometry, an analysis that involves the digestion of a purified protein with a low specificity protease, a data analysis that considers a large set of possible modifications and subsequent confirmation of identified PTMs will become more widely used. While standardised procedures for these endeavours are not currently in place, one way in which this challenge may be approached was demonstrated here.

Despite these efforts, it was not possible to rule out or confirm methylation of S91, D92 and N203. While the use of MethylSILAC does not improve the localisation of a methylation site on a peptide it can indicate if a methyl-group is indeed present and may therefore be a logical next step to take. Modifications of +28.03 Da, which may constitute di-methylations, had not been considered in this analysis as they are likely to be ethylations since ethanol was present during sample preparation. Besides avoiding ethanol, MethylSILAC could also be employed to overcome this problem, as no ethylation artefact corresponding to the heavy di-methylation occurs. It may therefore be advisable to routinely include MethylSILAC in the candidate approach despite the additional cost and effort.

In conclusion, 99% sequence coverage of Cdc33 was achieved that lead to the apparent identification of ten methylation sites on this protein. The majority of these sites can be regarded as false positives as they were not detected at stricter analysis criteria or as plausible alternative matches to the corresponding fragment spectra were identified. Three of the sites detected after elastase digest and one after tryptic digest require further analysis to be validated or disproven. In contrast, it could be shown that the RG-rich region of Scd6 is

heavily methylated. The assignment of the PRMT responsible for these methylations is addressed in the next chapter.

# 4. Assigning Arginine Methylation Sites to Methyltransferases

## 4.1. Introduction

To gain a better understanding of the biological role of a protein methylation it is important to determine which MTase is responsible for depositing the modification. This knowledge can be informative about which methylation sites are co-regulated by the same enzyme and/or which stimuli lead to changes in methylation levels as they may coincide with stimuli that regulate the activity of the corresponding MTase.

In *S. cerevisiae* four arginine methyltransferases have been identified to date (see also "1.4. Protein Arginine Methyltransfrases"): Hmt1 is a type I PRMT and homologous to human PRMT1. It is considered to be the predominant PRMT in budding yeast (Gary et al. 1996). Hsl7 is a type II PRMT and a homologue of PRMT5 (Lee et al. 2000, p.7). Rmt2 is the only type IV PRMT known in all species, methylating the δ-position of the arginine guanidino-group (Zobel-Thropp et al. 1998). Finally, Sfm1 belongs to the SPOUT-family of MTases which normally catalyse the methylation of RNA and is classified as a type III PRMT based on the methylation of its only known target Rps3 (Young et al. 2012). Both Rmt2 and Sfm1 do not have a known homologue in higher eukaryotes (Krause et al. 2007). Based on the observation of limited change of global methylation levels in strains with disruption of one or several genes encoding these PRMTs, it has been argued that additional PRMTs are likely to exist in yeast (Low et al. 2013).

The PRMT activities responsible for the deposition of the R119 methyl-group of Lsm4 and the methyl-groups in the RG-rich region of Scd6 which were identified in this study ("2.3.6. Identification of Arginine Methylation Sites via Peptide Immunoaffinity Purification", "3. Discovery of Protein Methylation by a Candidate Approach") were investigated here. Lsm4 was chosen as its human homologue (LSM4) is also arginine methylated, albeit in a region with no apparent sequence homology to budding yeast Lsm4 (Fig. 4.1). Methylation of

human LSM4 at R88, R90, R102, R109, R115, R117, R125 exists in the symmetric di-methyl form and is carried out by PRMT5 (Brahms et al. 2001, p.4). Similarly, the *Arabidopsis thaliana* orthologue of PRMT5 (AtPRMT5) methylates the *Arabidopsis* homologue of Lsm4 (AtLSm4), a process which is important for correct mRNA splicing (Deng et al. 2010). While it has been shown to symmetrically di-methylate calf histones *in vitro*, no *in vivo* substrate is currently known for Hsl7 (Miranda et al. 2006). Symmetrically di-methylated arginine residues do, however, exist in the yeast proteome (Lakowski et al. 2015). The type of di-methylation on yeast Lsm4 was not determined, but it is reasonable to assume that it is symmetric due to the homology to the human protein. It can be hypothesized that Lsm4 may therefore prove to be the first *in vivo* target of Hsl7.

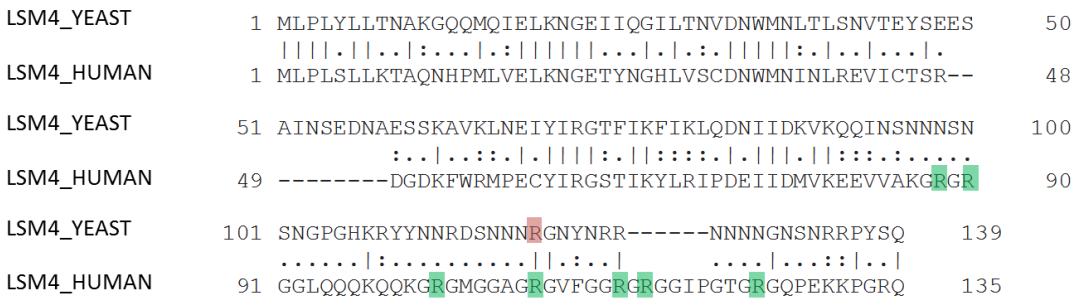



***Figure 4.1: Sequence alignment of S. cerevisiae and human Lsm4.*** *Sequences were aligned using Clustal Omega (Sievers & Higgins 2014). The Rme1/Rme2 methylation site in the yeast protein (this study) is highlighted in red; Rme2sym sites according to Brahms (2001) are highlighted in green.*

As a second target Scd6 was selected, because a function of the RG-rich region of Scd6 in translational repression has been observed. Arginine methylation events within this domain may be implicated in translational repression (see "5. A Role of Scd6-Methylation Dependent Translational Repression upon Cell Stress").

As the methyltransferases Rmt2 and Sfm1 are atypical with respect to their type of MTase activity and structure, respectively, and as they only have one known target each, the main focus was placed on Hmt1 and Hsl7. Additionally, according to current knowledge, all di-methylation events should originate from Hmt1 and Hsl7.

## 4.2. Methods

### Strains and Plasmids

Yeast strains carrying plasmids for the overexpression of GST-tagged proteins by galactose induction were obtained from the yeast GST-tagged ORFs collection (GE, YSC4515) and single gene disruption strains from the *Saccharomyces* Genome Deletion Project (Thermo, YSC1053). *E. coli* strains harbouring plasmids with different *S. cerevisiae* open reading frames inserted into the BG1805 backbone were acquired from the MORF collection (Open Biosystems, YSC3868). The pOPIN-RSF and pOPIN-M plasmids were kindly provided by Prof. R. Owens (Oxford Protein Production Facility) and pDLB2211 and pDLB2212 were kind gifts of Dr. Lew (Duke University) (Theesfeld et al. 2003).

### Preparation of *S. cerevisiae* strains overexpressing GST-tagged proteins

*S. cerevisiae* cells from the GST-tagged ORFs collection were streaked onto SC-Ura agar with 2% glucose and a single colony inoculated into 4 ml of the corresponding medium and grown overnight at 180 rpm and 30 ˚C, before they were harvested and washed in distilled water. Plasmids were extracted according to the manufacturer`s instructions for the QIAprep Spin Mini-prep kit (Qiagen, 27104), except that the cells were vortexed with 250 µl glass beads for 2 min after the addition of lysis buffer. Plasmids were transformed into TOP10 chemically competent *E. coli* cells (Invitrogen, C4040-10) according to the manufacturer`s instructions. Plasmids were purified by inoculating 5 ml LB containing 100 µg/ml ampicillin, growth at 37˚C overnight and extraction using the QIAprep Spin Mini-prep kit (Qiagen, 27104). For long-term storage, bacterial strains were kept in terrific broth + 8% glycerol. Plasmids were transformed into *hmt1Δ::KanMX, hsl7Δ::KanMX, rmt2Δ::KanMX* and *sfm1Δ::KanMX* strains of the yeast gene deletion collection (Thermo, YSC1053) (in a BY4741 background) using a yeast transformation kit (Sigma, YEAST-1). Transformed cells were plated on SC-Ura + 2% Glc plates and single colonies were re-streaked onto SC-Ura +

2% Glc. For long-term storage, transformed strains were kept in SC-Ura + 15% glycerol at -80 ˚C.

**Protein purification and mass spectrometry analysis**

Yeast cultures were maintained as described in section 2.2. ZZ-domain tagged (in the following: ZZ-tagged) proteins were overexpressed from plasmids of the MORF collection and purified as detailed in section 3.2. GST-tagged proteins were expressed from plasmids of the yeast GST-tagged ORFs collection (GE, YSC4515) in an identical way, except that proteins were bound to glutathione-coupled sepharose beads (GE, 17-0756-01) and eluted in 50 mM Tris-HCl, pH 8.0 with 10 mM reduced glutathione (Sigma, G4251). In-gel digestion and mass spectrometry analysis using a QExactive mass spectrometer were performed as described in detail in section 2.2.

**Western blotting**

Fluorescence-based western blotting was performed as in section 2.2, with the additional use of primary antibodies ICP0810 (Immunechem, ICP0810) and anti-GST (Sigma, G1160) at 1 : 1,000 (v/v) dilutions. For enhanced chemiluminescent (ECL) western blots, membranes were stripped with stripping solution (Millipore, 2504) according to the manufacturer`s instructions and re-probed as above, except for the secondary antibody incubation step, in which horseradish peroxidase coupled anti-mouse (Dako, P0447) and anti-rabbit (Dako, P0448) antibodies were applied in a 1 : 10,000 (v/v) dilution ratio and incubation was performed for 10 min at room temperature. The blots were imaged by addition of 1 ml ECL Prime (GE, RPN2232) and exposure to photographic film.

**Mass spectrometry data analysis**

For the identification of methylation sites mass spectrometry raw data were converted to .mgf files using MSConvert (Proteowizard), retaining the 200 most intense peaks. Data were searched with the Mascot search engine (v2.5) against a concatenation of SwissProt entries

corresponding to *S. cerevisiae* strain S288c and the sequence of the affinity tag encoded on plasmids of the MORF library. Chymotryptic cleavage with up to three missed cleavages, carbamidomethylation on cysteines as a fixed and arginine mono- and di-methylation as well as asparagine deamidation as variable modifications were specified. A mass tolerance of 10 ppm on precursor and 0.05 Da on fragment ions were used.

For label-free, quantitative analysis data were processed using the Progenesis QI software (Waters). The peak areas for each peptide were normalized to the average area of all peptides derived from Scd6 and its affinity tag from the same sample. Differences between samples were evaluated by a two sided Student`s t-test assuming homoscedasticity.

**Construction of expression plasmids**

Site specific mutagenesis of plasmid BG1805 containing a *HMT1* insert (Open Biosystems, YSC3868) was performed to yield a substitution of glycine 68 to arginine using the QuickChange Lightning mutagenesis kit (Agilent, 210518) according to the manufacturer`s instructions with primers 5'-GGATAAAATACCGGTACGGCAACCGACGTCTAAAA-3' and 5'-TTTTAGACGTCGGTTGCCGTACCGGTATTTTATCC-3'. This mutation has been found to render Hmt1 catalytically inactive (McBride et al. 2000).

Construction of bacterial expression plasmids was performed with the kind help of the Oxford Protein Production Facility (OPPF). The open reading frames for *HMT1*, *HMT1-G68R* and *SCD6* lacking their start codon were amplified from the corresponding BG1805 plasmids by polymerase chain reaction (PCR) using primers 5'-AAGTTCTGTTTCAGGGCCCGAGC AAGACAGCCGTGAAAGATTCTGC-3' and 5'-ATGGTCTAGAAAGCTTTAATGCATT AAATAAGAACCTTCGTTTTTTCTGC-3' for the wt and mutant version of Hmt1 and 5'-AAGTTCTGTTTCAGGGCCCGTCGCAGTACATCGGTAAAACTATTTCTTTAATCTCT -3' and 5'-GTGATGGTCTAGAAAGCTTTAAAATTCAACGTTGGAAGGAGGTTGCG-3'

for Scd6. PCRs were performed in 50 µl volumes with Phusion Flash polymerase (Thermo, F-548L) and the following reaction conditions according to manufacturer`s instructions:

1, Initial denaturing: 98 °C, 10 seconds
2, Denaturing: 98 °C, 1 second
3, Annealing: 60 °C, 5 seconds
4, Extension: 72 °C, 30 seconds
5, Repeat steps 2 to 4 29 times
6, Final extension: 72 °C, 2 minutes

To digest the template, 5 µl DpnI (NEB, R0176S) were added and samples incubated for 30 min at 37 °C. For purification of the PCR product, 90 µl AMPure magnetic beads (Beckman, A63881) were added, samples mixed, incubated at room temperature for 5 min, placed on a magnet and the supernatant removed. The beads were washed twice with 200 µl of 70% ethanol with 5 min incubation at room temperature each time. After air drying for 30 min, DNA was eluted in 30 µl 10 mM Tris pH 8.0.

PCR products were cloned into plasmid pOPIN-RSF or pOPIN-M using the In-Fusion enzyme (Clontech, 638912) and KpnI and HindIII linearized plasmids. Reaction mixtures were transformed into OmniMaxII competent *E. coli* cells (Invitogen, C8540-03) according to manufacturer`s instructions. Positive clones were selected for by plating on LB agar plates containing 50 µg/µl ampicillin. GST-tagged Hsl7 and Hsl7-G386A, a catalytically inactive mutant of Hsl7, were expressed from plasmids pDLB2211 and pDLB2212, respectively.

**Protein expression and purification from *E. coli***

Plasmids were transformed into BL21 DE3 *E. coli* cells (Invitrogen, C6000-03) by heat-shock according to the manufacturer`s instructions. Single colonies were inoculated into 5 ml LB with 50 µg/ml ampicillin and grown with shaking at 225 rpm and 37 ˚C overnight. Cultures were diluted 1 : 100 and incubated for 2 h at the same conditions before addition of isopropyl-

thio-galactoside (IPTG) to a final concentration of 1 mM and incubation for 3 h. Cultures were harvested by centrifugation and washed once with PBS. Bacteria expressing 6xHis-tagged proteins were resuspended in 20 ml lysis buffer (PBS pH 8.0, 10 mM imidazole, 1 mg/ml lysozyme, 1 mM PMSF, 1 protease inhibitor tablet (Roche) per 10 ml, 3 units/ml of benzonase). Cells were incubated on ice for 30 min and lysed by 12 cycles or sonication (10 s on, 30 s off) at 40% amplitude. Lysates were cleared by centrifugation for 20 min at 20,000 g and 4 ˚C and the resultant supernatant incubated with 500 µl pre-washed Ni-NTA beads (Merck, 69670) for 1 h at 4 ˚C. The beads were poured into a chromatography column, washed with 12.5 ml PBS pH 8.0 with 50 mM imidazole and bound proteins eluted in 500 µl fractions with PBS pH 8.0 with 250 mM imidazole. For GST-tagged plasmids PBS pH 7.3, 1 mg/ml lysozyme, 1mM PMSF, 1 protease inhibitor tablet per 10 ml with 3 units/ml of benzonase was as a lysis buffer and glutathione sepharose beads (GE, 17-0756-01) were used instead. Beads were washed with PBS pH 7.3 and proteins eluted with PBS pH 8.0 with 10 mM L-glutathione in this case. Samples were dialyzed against 1 l PBS + 15% (v/v) glycerol for 3 h with one buffer exchange and diluted to 1 µg/µl in the same buffer according to a reducing-agent compatible BCA assay (Pierce, 23250).

### *In vitro* methylation assay

*In vitro* methylation reactions were carried out with 1 µg of enzyme and substrate each, 2 µl adenosyl-L-methionine, S-[methyl-$^{3}$H] (Perkin Elmer, NET155H250UC; 0.55 mCi/mL) in a total volume of 20 µl 50 mM MOPS pH 7.2, 0.3 M NaCl, 2 mM EDTA for 1 h at 30 ˚C. Histones from calf thymus (Sigma, H7755) were used as substrate in a positive control reaction. For scintillation counting, the solution was spotted onto P81 filters (Whatman, 3698-321) and air dried. The filters were then washed 3 times for 10 min in 200 mM sodium bicarbonate buffer pH 9.2, rinsed in acetone and air dried again. Scintillation counts of each filter were measured in 4 ml scintillation fluid.

## 4.3. Results

### 4.3.1. Global Differences in Methylation Levels in Methyltransferase Deletion Strains

To obtain a global overview of changes in arginine methylation upon deletion of PRMTs, western blotting was performed against lysates from wild type (BY4741) and deletion strains of all known PRMTs from the *Saccharomyces* Genome Deletion Project with antibodies recognizing arginine mono-, asymmetric or symmetric di-methylation. No clear differences could be observed between the global methylation levels in BY4741 and *rmt2Δ*, *Δsfm1* and *hsl7Δ* strains with any of the antibodies tested. However, a striking reduction of methylation becomes apparent when using two different antibodies targeting arginine mono-methylation (Me-R4-100 and D5A12) upon deletion of *hmt1*. No clear difference was observed with antibodies for symmetric and asymmetric di-methyl arginine (SYM10 and ASYM24 respectively) (Fig. 4.2).

To test whether low levels of arginine mono-methylation are present in *hmt1Δ* strains that were not observable with fluorescence based western blotting the membrane was re-probed with chemiluminescence based readout. Despite an increase in signal for the other strains, bands were still completely absent for the *hmt1Δ* sample (Fig. 4.3).

### 4.3.2. Differences in Lsm4 and Scd6 Methylation Levels in Wild-Type vs. *hmt1Δ* and *hsl7Δ* Strains Assessed by Western Blotting

It was next aimed to identify differences in arginine methylation between wild type and MTase deletion strains using western blots with anti-methyl arginine antibodies on purified Lsm4 and Scd6 proteins. This experiment was performed on GST-tagged proteins, as it was found that the ZZ-tag of proteins of the MORF collection is still capable of binding antibodies after denaturing gel electrophoresis and transfer to a membrane (data not shown). First, several commercially available antibodies were tested to determine if they would be able to recognize methylation on the target proteins. As shown in fig. A4.1 in the appendix, antibodies sdme-RG (CST) and SYM10 (Millipore) targeting symmetrically di-methylated

arginine and adme-R (CST) and ASYM24 (Millipore) targeting asymmetrically di-methylated arginine did not yield bands for GST-Lsm4 or GST-Scd6 above the level of signal for the GST-tag alone.

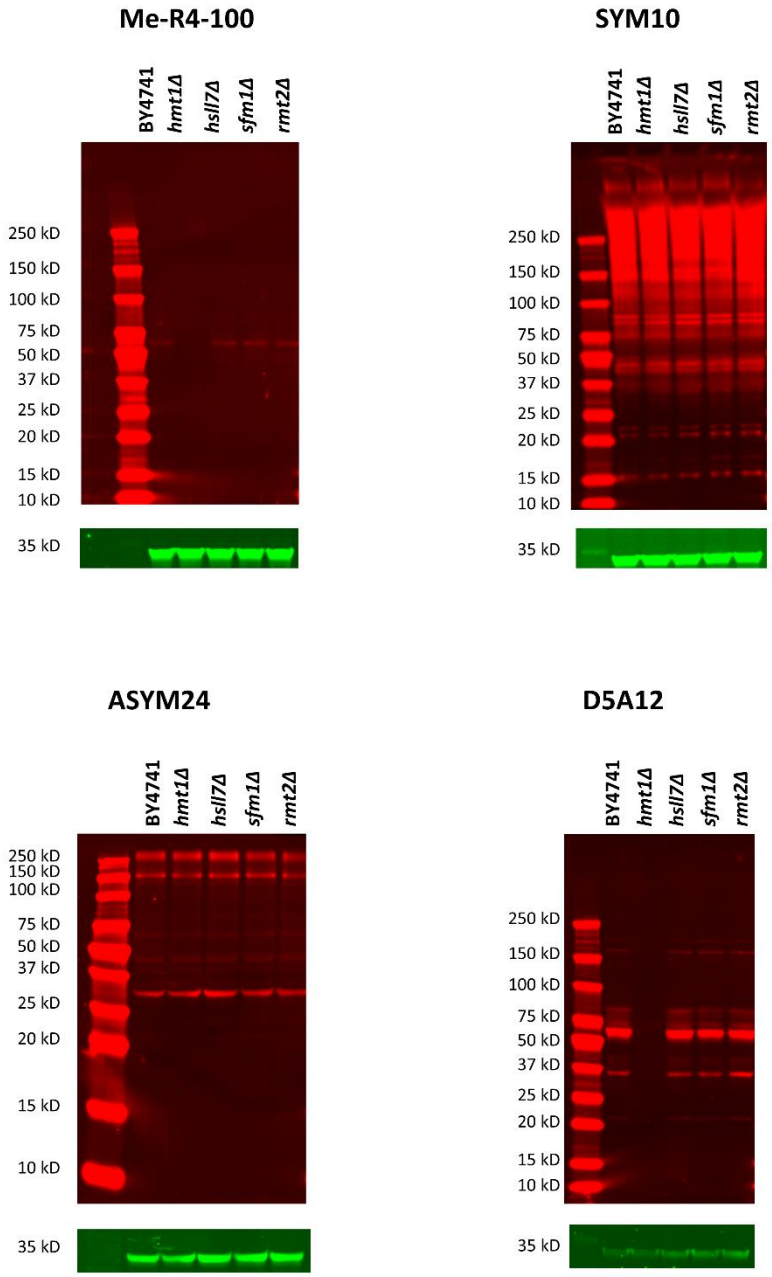



***Figure 4.2: Global arginine methylation levels in wild type and PRMT deletion strains.*** *Crude lysates were obtained from BY4741, hmt1Δ, hsl7Δ, sfm1Δ and rmt2Δ cells and analysed by western blotting with anti-methyl-arginine antibodies Me-R4-100 (CST), SYM10, ASYM24 (both Millipore) and D5A12 (CST) (red) and anti-GAPDH (green).*

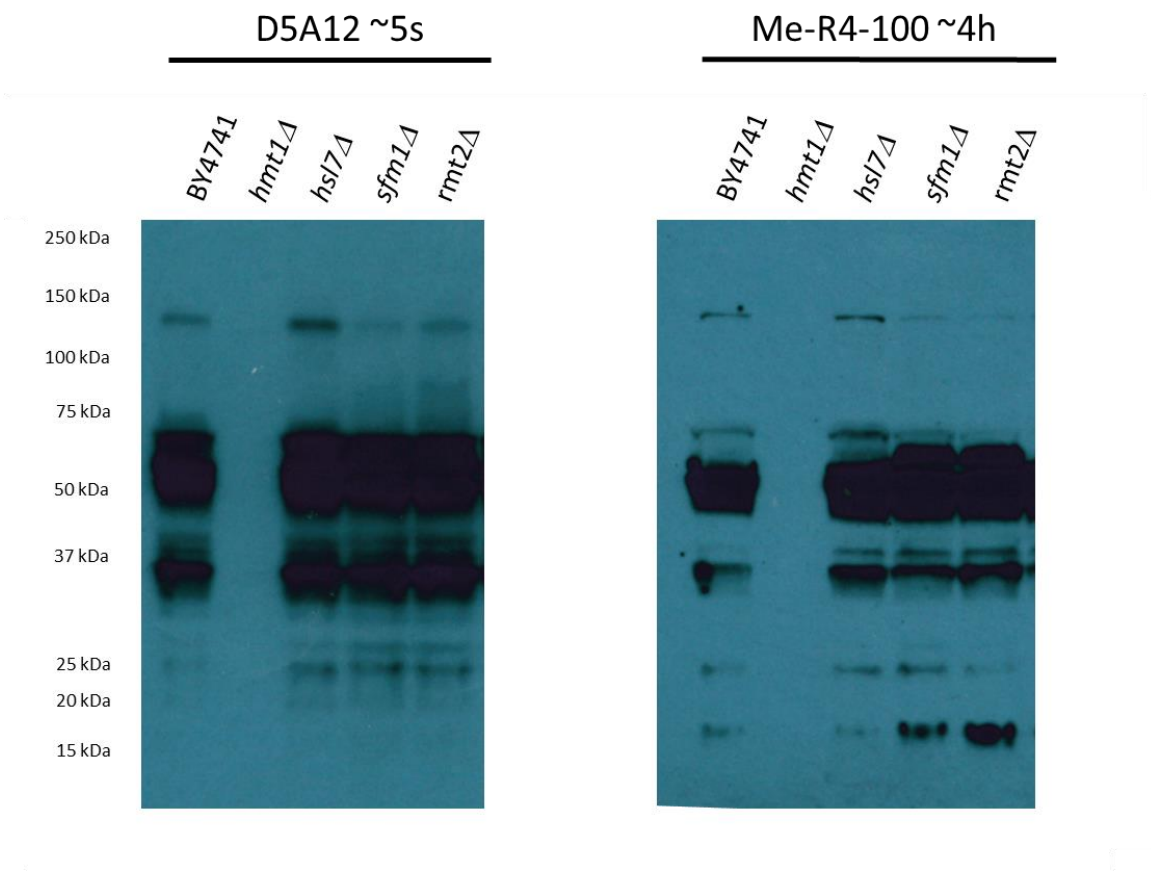



***Figure 4.3: Anti-methyl arginine western blots on BY4741 and PRMT candidate disruption strains.*** *Lysates of BY4741, hmt1Δ, rmt2Δ, sfm1Δ and hsl7Δ cells were resolved by SDS-PAGE and transferred to a PVDF membrane. Membranes from fig. 4.2 were striped and re-probed with anti-mono-methyl arginine specific antibodies D5A12 and Me-R4-100. The primary antibodies were detected with horseradish-peroxidase coupled secondary antibodies and blots developed by enhanced chemiluminescence for about 5 s and 4 h for plots with D5A12 and Me-R4-100 respectively. For anti-GAPDH loading control visualized by fluorescence-based western blotting see fig. 4.2.*

Similarly, western blots with antibodies Me-R4-100 (anti-Rme1) and ICP0810 (anti-Rme2as) did not recognize methylation of GST-Lsm4 purified from BY4741, *hmt1Δ* or *hsl7Δ* strains. In contrast, clear bands at the expected molecular weight were obtained with these antibodies for GST-Scd6 purified from BY4741 and *hsl7Δ* strains, while no signal was observed for GST-Scd6 from an *hmt1Δ* strain. The background signal observed from the GST-tag alone was much weaker than the signal from GST-Scd6 in each case. An alternative anti-Rme2as antibody, ASYM24, failed to detect Rme2as on GST-Scd6 (Fig. 4.4). Overall, it was concluded that GST-Scd6 methylation involves the MTase Hmt1.

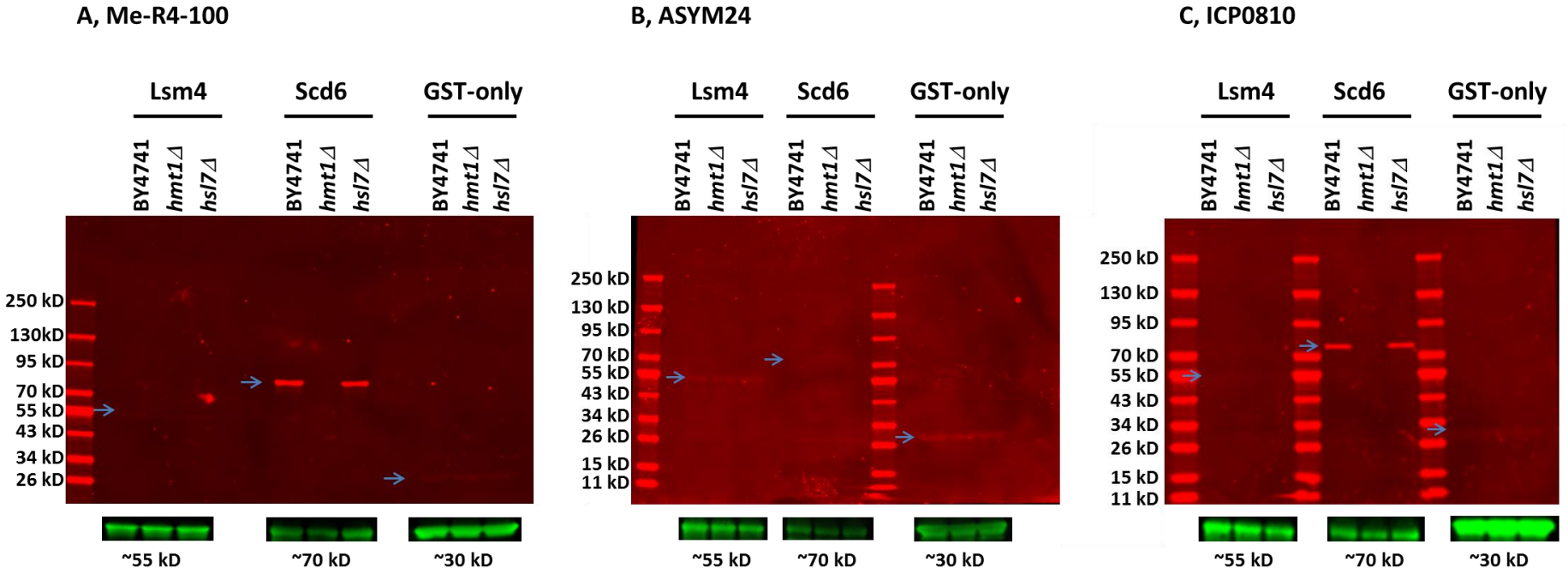



***Figure 4.4: Arginine methylation levels of purified proteins.*** *GST-tagged Lsm4 and Scd6 and the GST-tag alone were purified from BY4741, hmt1Δ and hsl7Δ cells as indicated and subjected to western blotting with anti-Rme1 antibody Me-R4-100 (A) and anti-Rme2as antibodies ASYM24 (B) and ICP0810 (C) respectively (red). All blots were additionally probed with anti-GST antibody (green). Blue arrows indicate the expected molecular weight of the purified proteins.*

### 4.3.3. Identification of Lsm4 Methylation in Wild-Type, *hmt1Δ* and *hsl7Δ* Strains by Mass Spectrometry

As Lsm4 methylation could not be detected by western blotting, it was next attempted to quantify Lsm4 R119 methylation levels by mass spectrometry. However, as the strains used in this study are not arginine auxotrophic, attempts at SILAC labelling with heavy arginine resulted in too low labelling efficiency to allow a reliable quantification (data not shown). Efforts at label-free quantification gave rise to high variability between samples due to pronounced levels of deamidation of the asparagine rich peptides $_{111}$NNRDSNNNRGNY$_{122}$ or $_{110}$YNNRDSNNNRGNY$_{122}$ and poor retention on reverse phase columns due to their hydrophilic nature (data not shown). The choice of chymotrypsin as a protease instead of trypsin was due to the (potential) inhibition of trypsin at modified arginines (Baldwin & Carnegie 1971).

The samples were therefore compared qualitatively by asking in which strains mono- and di-methylated R119 could be observed. The results are summarized in table 4.1: Unmodified, mono- and di-methylated forms of R119 were identified in the BY4741, *hmt1Δ* and *hsl7Δ* samples (E < 0.05), but not in the blanks between the samples (at any E-value < 1). Each of these identifications originates from a peptide with missed cleavage, YNNRDSNNNRGNY in charge state 3+. In some cases, also the peptide NNRDSNNNRGNY and the doubly charged form were identified. Neutral losses or diagnostic ions distinguishing symmetric from asymmetric di-methyl arginine were not observed.

| | R119unmod | R119me1 | R119me2 |
|---|---|---|---|
| BY4741 | y | y | y |
| *hmt1Δ* | y | y | y |
| *hsl7Δ* | y | y | y |
| blanks | n | n | n |

**Table 4.1:** Identifications of R119 modification states in wild type and MTase deletion strains (y = identified; n = not identified)

### 4.3.4. Differences in Scd6 Methylation Levels in Wild-Type vs. *hmt1Δ* and *hsl7Δ* Strains Assessed by Mass Spectrometry

To quantify methylation levels of individual arginines of Scd6, the ZZ-tagged protein (Scd6-ZZ) from three replicates each of strains BY4741, *hmt1Δ* and *hsl7Δ* was purified, in-gel digested and subjected to mass spectrometric analysis. Label-free quantification was used, as only limited labelling efficiencies can be achieved in a standard SILAC approach with the given strains. Chymotrypsin was chosen as a protease to avoid potential differences in cleavage efficiencies on methylated and unmethylated proteins. Scd6 was identified with a sequence coverage of 68% in this analysis; the peptide $_{291}$GRGRGNY$_{297}$ was present in an unmodified, di-methylated and doubly di-methylated states and with both one mono- and one di-methylation, however not with one or two mono-methyl-groups only. Different isoforms of the di-methylated and simultaneously mono- and di-methylated peptide mapped to the same features respectively and can therefore not be quantified independently. No further methylated peptides were detected.

All three arginine methylated forms of the peptide $_{291}$GRGRGNY$_{297}$ were detected at much higher intensities in the wild-type than in the *hmt1Δ* strain, while the unmethylated peptide was present in a much lower amount in the wild type than in the *hmt1Δ* strain (Fig. 4.5 A). In each case, the methylated peptides were only present at background intensities in the *hmt1Δ* sample, consistent with a complete absence of methylation without Hmt1. Apart from the singly di-methylated peptide (E-value = 0.03), the changes in methylation level between wt and *hmt1Δ* strains did however not reach statistical significance at an E-value of 0.05 due to the high variance in the wild-type sample, probably attributed to the label-free nature of the experiment. In contrast, no comparable changes were detected between wild-type and *hsl7Δ*, indicating that Hsl7 does not measurably contribute to the methylation of R292 and R294 of Scd6-ZZ (Fig. 4.5 B).

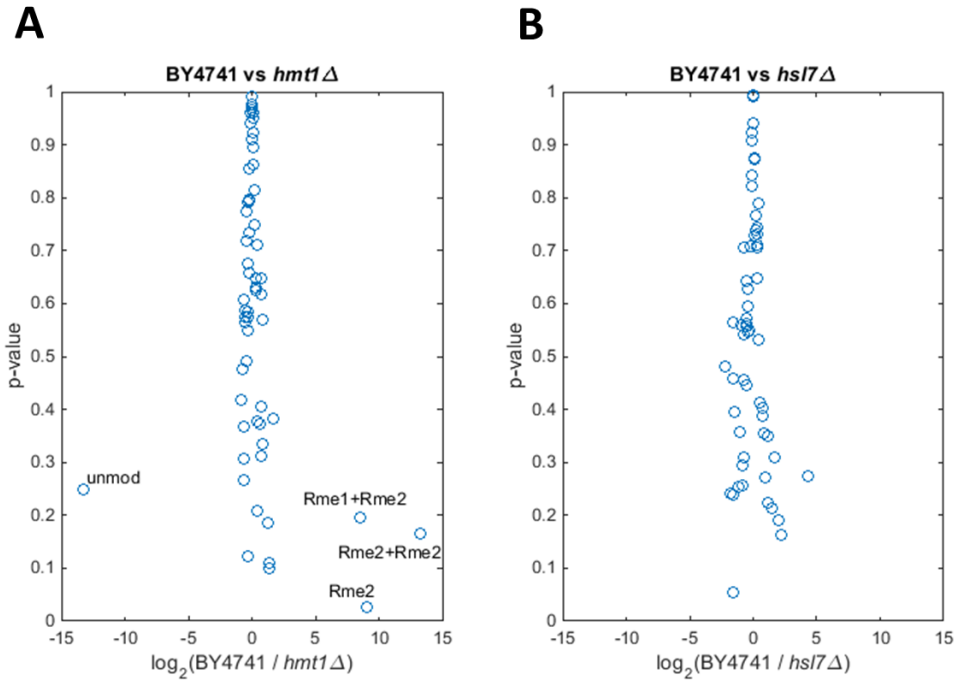



***Figure 4.5: Quantification of Scd6-ZZ arginine methylation in BY4741 vs. hmt1Δ and hsl7Δ.*** *Protein Scd6 was purified via a ZZ-domain-tag from three replicates each of BY4741, hmt1Δ and hsl7Δ, in-gel digested and peptides quantified by mass spectrometry. Volcano plots of the $log_2$ of the ratio of the mean signal for BY4741 to hmt1Δ (A) or hsl7Δ (B) against p-values from a two-sided t-test are shown. Only peptides originating from Scd6 or its affinity tag are shown. The outliers in A, are forms of the $_{291}$GRGRGNY$_{297}$ peptide and the unmodified (unmod), di-methylated (Rme2), doubly di-methylated (Rme2+Rme2) and mono- and di-methylated (Rme1+Rme2) states are labelled. As the unmodified form was not detected in the wild-type sample, it was plotted at the maximum absolute fold change.*

### 4.3.5. Hmt1 Methylates Scd6 *in vitro*

As quantitative mass spectrometry indicated a strong dependence of R292 and R294 methylation of Scd6-ZZ on the presence of MTase Hmt1, it was sought to confirm that the methylation is carried out by Hmt1. For this, an *in vitro* methylation assay with enzymes and substrate purified from *E. coli* was used.

It was initially tried to express Hmt1 and Hmt1-G68R from plasmid pOPIN-RSF with an N-terminal hexa-histidine tag. However, using this system, little soluble protein was produced under non-denaturing lysis conditions. The fact that a much higher amount was detected when denaturing lysis was used indicated accumulation in inclusion bodies (data

not shown). To improve solubility, the protein was therefore produced with a dual hexa-histidine - maltose binding protein tag (in the following: MBP) using the pOPIN-M vector. Recombinant MBP-Scd6, MTase and adenosyl-L-methionine, S-[methyl-$^{3}$H] were combined and incubated at 30℃ for 1 h. The radioactive decay was monitored by scintillation counting after spotting the reaction mixtures on filters with subsequent rinsing.

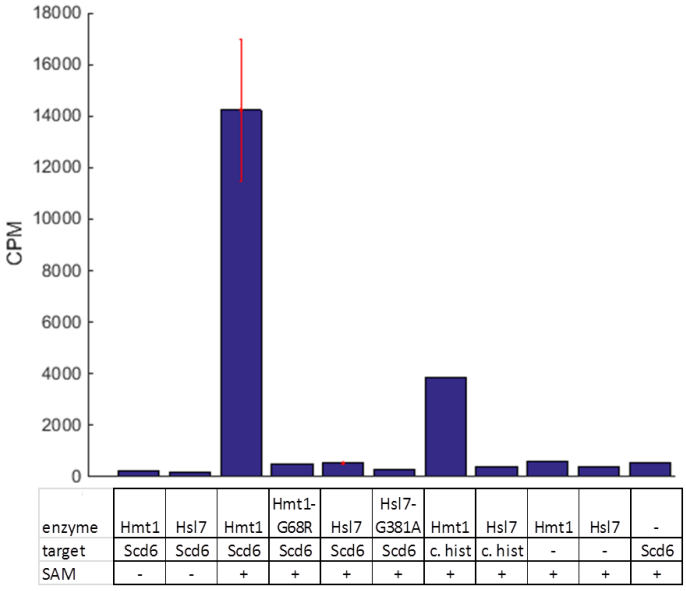



***Figure 4.6: Scintillation counts for in vitro methylation of Scd6.*** *MBP-Scd6 or calf histones (c. hist) were in vitro methylated with 6xHis-Hmt1, GST-Hsl7 or their inactive mutant versions 6xHis-Hmt1-G68R and GST-Hsl7-G386A purified from E. coli. Minus signs indicate omission of protein. Reactions were performed in the presence or absence of adenosyl-L-methionine, S-[methyl-$^{3}$H] (1.1 µCi) as indicated. Error bars for methylation of Scd6 by Hmt1 or Hsl7 represent standard deviations from three technical replicates. CPM: counts per minute.*

Reactions where enzyme, substrate or radioactive co-factor had been omitted exhibited a background level of incorporation of about 500 counts per minute (CPM) (Fig. 4.6). The methylation of MBP-Scd6 by GST-Hsl7 yielded similarly low counts to the catalytically inactive version of the enzyme, GST-Hsl7-G386A. As calf histones, which are the only known target of Hsl7 (Miranda et al. 2006, p.7) were not methylated, it is likely that GST-Hsl7 was not active under the conditions used.

In contrast, a much higher signal (mean of 14,229 CPM +/- 2,768 CPM standard error of the mean) was observed for the methylation of MBP-Scd6 by 6xHis-Hmt1. 6xHis-Hmt1 was also found to methylate calf histones (3,859 CPM), while the control reaction with MBP-Scd6 and 6xHis-Hmt1-G68R did not give signal above background level.

## 4.4. Discussion

### 4.4.1. Global Differences in Methylation Levels in Methyltransferase Deletion Strains

To gain a qualitative overview of the extent to which different methyltransferases contribute to arginine methylation in *S. cerevisiae* western blots with pan-methyl antibodies for the three different types of arginine methylation were performed. A striking reduction of total protein arginine mono-methylation in an *hmt1Δ* strain was apparent when compared to wild type *S. cerevisiae* cells. This observation is consistent with previous reports that Hmt1 is the major arginine MTase in *S. cerevisiae* (Gary et al. 1996) even though a complete absence of arginine methylation upon *hmt1* deletion was not observed in this study. The results also resemble those obtained by an *in vivo* methylation assay comparing a wt and *hmt1Δ* strain (Xu et al. 2003). On the other hand, they are in marked contrast to more recent observations by amino acid analysis on the whole proteome, which indicated that ω-MMA levels are reduced to only about half upon *hmt1* single or simultaneous *hmt1*/*hsl7* disruptions (Lakowski et al. 2015). They are also in contrast to a previous report in which only a subset of the bands in an anti-mono-methyl arginine western blot (with antibody ICP0801) were absent in *hmt1Δ* compared to wt cells (Low et al. 2013). It should be noted that the number of bands observed here was lower than by Low *et al.* (2013). It is possible that both anti-mono-methyl arginine antibodies used here (Me-R4-100, D5A12) are biased towards certain sequence contexts in which the arginine methylation occurs. In the case of such bias, the sequence motif in which methylation is recognized by the antibodies may be similar to the one methylated

by Hmt1. This motif may be the RGG-motif, as it is a preferential target of Hmt1 (Low & Wilkins 2012). The D5A12 antibody was raised against peptides with methylation sites in an RGG context, while Me-R4-100 was created from animals immunized with methylated peptides without sequence bias. It should be noted that the bands observed with the second antibody might be a subset of the ones detected with the former, indicating that the signal obtained from the Me-R4-100 antibody might originate from proteins methylated in an RGG context. As in many cases the methylation of arginines in an RGG motif occurs on several sites of a protein, it is also possible that methylation in this context is detected because of its stronger signal, while proteins methylated at a single site might not reach the detection limit of this method. Again, an antibody bias for RGG motifs coinciding with the target motif of Hmt1 may explain the results in a scenario where mono-methylation is not completely absent in *hmt1Δ* strains.

No changes in methylation levels were observed with antibodies for symmetric and asymmetric di-methyl arginine between wt and any of the deletion strains and therefore it is not evident if the detected signal is specific for methylation. The results for asymmetric di-methyl arginine resemble those of a previous report which did not observe differences between wt and a *hmt1Δ/hsl7Δ/rmt2Δ* strain in a western blot with antibody ICP0810 (Low et al. 2013). Notably, also no reduction of signal was observed in this study when adding free Rme2as during the western blot. No global assessment of Rme2sym levels by western blotting has been reported in *S. cerevisiae*. A striking reduction of Rme2as and a slight reduction of Rme2sym levels in an *hmt1Δ* strain were recently observed by amino acid analysis (Lakowski et al. 2015).

To determine which MTases are responsible for the methylation of specific proteins in *S. cerevisiae* the dependence of methylation states of two targets, Lsm4 and Scd6, on two PRMTs, Hmt1 and Hsl7, was tested.

Scd6 is a protein with several methylation sites in its RG-rich region (Fig. 3.3), while a single site (R119) was detected on Lsm4. Hmt1 and Hsl7 are the only known orthologues of human PRMTs in *S. cerevisiae* and according to current knowledge should explain all arginine di-methylation in this organism (Gary et al. 1996; Lee et al. 2000, p.7).

### 4.4.2. Lsm4 is neither Exclusively Methylated by Hmt1 nor Hsl7

When attempting to detect GST-Lsm4 methylation by western blotting, no signal above the level for the GST-tag alone was observed with the antibodies tested (sdme-RG and SYM10 for Rme2sym, Me-R4-100 for Rme1 and adme-R, ASYM24 and ICP0810 for Rme2as). This may be partially attributed to the nature of the only known methylation site in Lsm4, R119, which occurs in an RG, but not an RGG context and not in a sequence containing multiple methylation sites. The later factor renders the site a suboptimal target for some of the antibodies used (Table A2.1 in the appendix). However, the absence of signal may also be due to limited sensitivity as Lsm4 only has one known methylation site of unknown stoichiometry.

The mass spectrometric quantification of methyl-groups on both Lsm4 and Scd6 were challenged by the facts that sufficient SILAC labelling could not be achieved in the BY4741 background, as this strain is not arginine auxotrophic, and that an alternative protease to trypsin had to be used due to the potential (partial) inhibition of trypsin at methylated arginine. Because of these limitations coupled with the variable levels of deamidation and the hydrophilic nature of the peptides harbouring R119, reliable quantification of Lsm4 methylation levels was not achieved. However, a qualitative analysis identified Lsm4 R119 mono- and di- methylation in wt as well as *hmt1Δ* and *hsl7Δ* strains, while it was not identified in the blank runs between the samples. This allows one to conclude that this site is neither exclusively methylated by Hmt1, nor Hsl7. This finding sets Lsm4 apart from most other known arginine methylated proteins in *S.*

*cerevisiae*, which are thought to be mono- and asymmetrically di-methylated by Hmt1 only. For example, the signal obtained in radioactive *in vitro* methylation assays for Npl3, Nsr1, Nop1, Gar1 and Rps2 was completely lost in *hmt1Δ* strains (Henry & Silver 1996; Lipson et al. 2010; Xu et al. 2003). The detection of Nop1 and Nsr1 in an anti-methyl arginine immunoprecipitate from an *hmt1/hsl7/rmt2* triple deletion strain seemingly contradicts the idea of exclusive methylation by Hmt1, however as methylation sites were not identified in this assay, the proteins may have been unspecifically co-immunoprecipitated (Low et al. 2013).

Further work to confirm this finding could include establishing a SILAC method to allow quantification despite high variations introduced during sample preparation (e.g. by deamidation) and of a method to distinguish symmetric from asymmetric di-methylation. It will also be important to test simultaneous disruptions of multiple PRMTs for the absence of Lsm4 methylation.

#### 4.4.3. Scd6 is Methylated by Hmt1

In western blotting assays, mono- and asymmetric di-methylation was recognized on purified GST-tagged Scd6 by antibodies Me-R4-100 and ICP0810, however, not by the other antibodies for Rme2as, ASYM24 and adme-R. The complete absence of this signal in the *hmt1Δ*, but not the *hsl7Δ* strain for both Rme1 and Rme2as suggests that Hmt1 is the only enzyme catalysing methylation of Scd6. However, it is possible that methylation remains in the *hmt1Δ* strain on sites which are not recognized by the antibodies. Additionally, this assay did not rule out that Scd6 may additionally be symmetrically di-methylated.

It was therefore also aimed to quantify methylation levels by mass spectrometry. Using chymotryptic digestion, only two of the eight arginines in the RG-rich region of Scd6 were covered by an identified peptide. For this peptide, signal for both mono- and di-

methylation in all identified combinations was reduced to background level in the *hmt1Δ* strain. Taken together, these results strongly suggest complete absence of Scd6 arginine methylation in an *hmt1Δ* strain. Using an *in vitro* methylation assay with recombinant proteins it was shown that Scd6 methylation can be catalysed directly by Hmt1. While a better coverage of methylation sites and more robust quantification is still needed, these results will facilitate the investigation of the biological role of Scd6 methylation. Based on the knowledge of the PRMT, these assays will not need to rely on site-specific mutagenesis of arginines, which may not faithfully represent either the methylated or unmethylated state, but can instead make use of Scd6 *in vitro* methylated by Hmt1 or assays conducted in wt vs. *hmt1Δ* strains. In the case of the latter, it however needs to be kept in mind that the methylation state of other proteins will also change. Initial attempts towards elucidating the role of Scd6 methylation are presented in the following chapter.

# 5. A Role of Scd6-Methylation Dependent Translational Repression upon Cell Stress

## 5.1. Introduction

To assess the biological importance of arginine methylation in *S. cerevisiae*, the literature was surveyed for cases in which the sequences found to be arginine methylated in proteins have been demonstrated to play a role in a biological process.
In a previous report, it has been shown that the Scd6 protein is involved in translational stalling upon glucose deprivation and sodium azide stress, which is reflected by the formation of stress granules. Sodium azide treatment leads to a block in the respiratory chain by binding of azide to cytochrome oxidase (Keilin 1936).

The stress granule formation is less prominent in strains lacking Scd6 than in the wt, but occurs spontaneously in cells overexpressing the protein. Importantly, the process is dependent on the RG-rich region of Scd6, as stress granule formation occurs with similarly low frequency to Scd6 deletion strains in strains in which the RG-rich region has been ablated from the protein (Rajyaguru & Parker 2012). As arginine methylation of the RG-rich region of Scd6 was demonstrated in this report ("3. Discovery of Protein Methylation by a Candidate Approach"), it was hypothesized that this modification may be involved in controlling translational repression upon cellular stress. This research was explored in collaboration with the Rajyaguru group (Indian Institute of Science). While our collaborators addressed the question whether altered arginine methylation levels would influence the efficiency of translational repression and stress granule formation, we asked whether stress would alter the levels of arginine methylation on Scd6.

## 5.2. Methods

### Determining Scd6 methylation levels upon sodium azide stress

The overexpression and purification of GST-tagged Scd6, Npl3 and the GST-tag alone from the wild type strain and of Scd6 from an *hmt1Δ* strain was performed as described in section 4.2. Two 300 ml cultures were combined and distributed to two flasks after 6 h of induction of protein expression, before adding sodium azide to a final concentration of 0.5% (v/v) to one and a corresponding amount of water to the control culture. The cultures were shaken at 180 rpm at 30 $^{0}$C for the times indicated (30 min to 4 h). After protein purification, western blotting was performed with antibodies ICP0810 (Immunechem), Me-R4-100, sdme-RG (both CST) and SYM10 (Millipore) as described in section 2.2. Protein expression and purification was performed in duplicate for Scd6.

## 5.3. Results

To evaluate the effect of azide stress on Scd6 methylation levels, western blotting was performed against GST-Scd6 purified from BY4741 cells exposed to sodium azide and control cultures. GST-Npl3 and the GST-tag alone were used as positive and negative controls for the detection of protein methylation respectively. Additionally, GST-Scd6 was purified from an *hmt1Δ* strain, for which it had been shown that GST-Scd6 mono- and asymmetric di-methylation was abolished ("4. Assigning Arginine Methylation Sites to Methyltransferases"). The absence of signal in the latter with antibodies targeting Rme1 (Me-R4-100) and Rme2as (ICP0810) indicates that the corresponding signal in the wild type strain is methylation-specific (Fig. 5.1 A and B). In both cases no signal was observed from GST-alone.

While GST-Npl3 mono-methylation was detected with Me-R4-100, asymmetric di-methylation was not detected with ICP0810, which may be due to the absence of Rme2as

on GST-Npl3 under these growth conditions or due the inability of this antibody to recognize this modification in the given context.

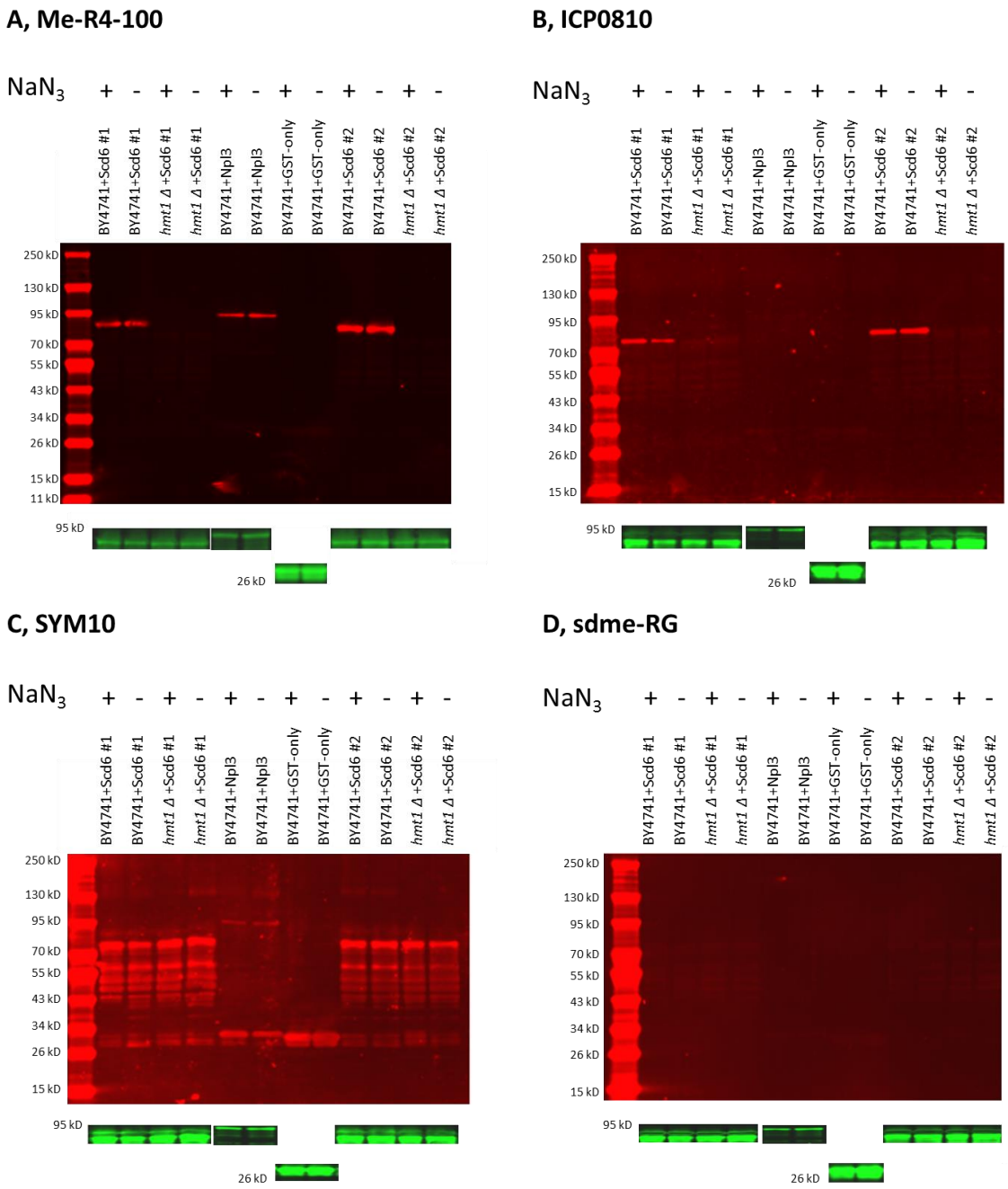



***Figure 5.1: Effect of sodium azide treatment on protein methylation levels.*** *Overexpressed proteins GST-Scd6, GST-Npl3 and the GST-tag alone (GST-only) were purified from sodium azide ($NaN_3$; 0.5% (v/v)) treated (+) or control (-) strains BY4741 or hmt1Δ as indicated. Western blots with anti-Rme1 antibody Me-R4-100 (A), anti-Rme2as antibody ICP0810 (B), and anti-Rme2sym antibodies SYM10 (C) and sdme-RG (D) are shown in red and anti-GST in green. The experiment was performed in two replicates (#1 and #2) for Scd6.*

No difference in signal between sodium azide treated and untreated strains was observed (Fig. 5.1). Symmetric arginine di-methylation was targeted with two antibodies (Fig 5.1 C and D). While no signal was obtained with sdme-RG, bands are visible not only at the expected, but also at lower molecular weights with SYM10. A reduction of signal in the *hmt1Δ* strain was not observed and would not be expected in this case if it was specific for symmetric di-methylation, as Hmt1 is a type I MTase, but the presence of a band for GST alone may indicate that this antibody binds to proteins in a non-specific manner. The levels of these modifications did not change upon sodium azide treatment.

After observing that arginine mono- and asymmetric di-methylation of Scd6 could be detected with antibodies Me-R4-100 and ICP0810 respectively, but that no differences in methylation levels were apparent after 30 min of sodium azide stress (Fig. 5.1 A, B) it was tested if application of the stress for a longer time period might lead to visible effects. The exposure to sodium azide stress was therefore extended to 2 and 4 h.

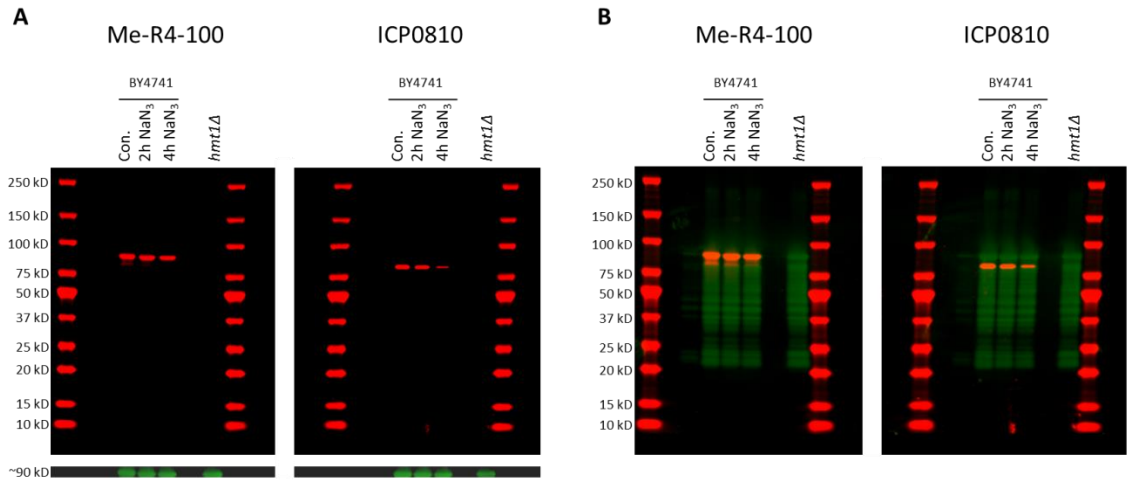



***Figure 5.2: Levels of arginine mono- and asymmetric di-methylation of Scd6 upon 2 and 4 h treatment with sodium azide.*** *Scd6 was overexpressed and cells were treated with sodium azide ($NaN_3$; 0.5 % (w/v)) for 2 or 4 h. Purified protein from treated BY4741 and untreated BY4741 (con.) and hmt1Δ strains was resolved on SDS PAGE gels and western blotting performed with antibodies Me-R4-100 and ICP0810 (red), recognizing mono- and asymmetric di-methyl arginine respectively, and anti-GST antibody (green). Blots are shown individually in A and overlaid in B.*

As above, signal was detected with both antibodies and was completely absent in an *hmt1Δ* strain, indicating that it indeed originates from methylation (Fig. 5.2). While only a slight decrease of arginine mono-methylation is perceptible with an increasing time of sodium azide treatment, the effect is more pronounced for asymmetric di-methylation. Notably, the bands for the latter do not correspond to the full length protein, but a form about 5 to 20 kDa smaller (Fig. 5.2 B).

### 5.4. Discussion

Low-glucose and sodium azide stress leads to an increase in stress granule formation, and the RG-rich region of Scd6 is necessary for this process (Rajyaguru et al. 2012). Scd6 is methylated in its RG-rich region ("3. Discovery of Protein Methylation by a Candidate Approach") and this methylation is catalysed by Hmt1 ("4. Assigning Arginine Methylation Sites to Methyltransferases"). These findings pave the way for studying translation inhibition and stress granule formation in dependence of methylation by employing *in vitro* methylated Scd6 for *in vitro* and *hmt1Δ* strains for *in vivo* experiments respectively.

On the other hand, if stress granule formation is regulated via arginine methylation, it is reasonable to assume that arginine methylation levels change after reduction of glucose levels or application of sodium azide stress. It was reasoned that methylation levels on Scd6 are most easily tested on tagged, overexpressed protein as it is challenging to assess the endogenous counterpart biochemically. However, as the system available for overexpression was based on galactose-induction, this is not readily compatible with a stress that involves the alteration of glucose levels. Scd6 arginine methylation levels were therefore tested after exposure to sodium azide, an inhibitor of respiration (Keilin 1936).

To test the model in which cellular stress leads to translational repression via changes in the methylation levels of Scd6, it was examined if its methylation levels would change

upon 30 min exposure to 0.5% (w/v) sodium azide. These conditions had been previously shown to induce the formation of stress granules in yeast cells (Rajyaguru et al. 2012). Methylation levels were assessed by western blotting with pan-methyl antibodies for different types of arginine methylation on overexpressed, purified proteins. While these experiments did not allow inferring if Scd6 is symmetrically di-methylated, a specific signal was obtained for both arginine mono- and asymmetric di-methylation. This signal was absent when the gene for the major type I PRMT, Hmt1, was deleted. No change in the levels of these modifications could be observed upon sodium azide treatment for 30 min, indicating that they are not the signals leading to stress granule formation.

When investigating the effect of longer exposure of cells to sodium azide, a marginal reduction of arginine mono- and clear decrease of asymmetric di-methylation of Scd6 was observed after 4 h. It can be speculated that sodium azide stress leads to a small reduction of methylation levels also on a short timescale, which could not be resolved by these experiments, but which nonetheless may play a role in the formation of stress granules. Alternatively, the decrease of methylation levels may be functional at a later point, e.g. in the maintenance of the granules. In this case, methylation would not explain the differences in stress granule formation between full length Scd6 and Scd6 lacking its RG-rich region reported after 30 min (Rajyaguru et al. 2012). Finally, the observed effect may be un-related to stress granule formation and be caused by global changes in the metabolism and physiology after long-term exposure to sodium azide. Similarly, the significance of observing asymmetric di-methylation on a shortened version of Scd6, but not the full length protein is unexplained to date. The strains used in this study lead to an expression of N-terminally GST-tagged proteins, while the RG-rich region of Scd6 is located close to the C-terminus (amino acids 288-311 of 349). Therefore, the shortest truncated version that would be purified via the GST-tag and

contain the RGG-region is about 5 kDa shorter than the full-length version. Other methylated, truncated versions of Scd6 that are not observed in this assay as they lack the GST-tag may potentially exist. It is therefore possible that asymmetric di-methylation is in this case linked to protein shortening, and testing this possibility warrants further investigation.

It is interesting to note that the growth of yeast strains under conditions of lower nutrient availability has been shown to cause a reduction of global arginine mono-methylation levels (Lakowski et al. 2013). In this case, methylation levels were measured after long-term exposure to the growth conditions and not as a short-term consequence of a sudden stress. These results may therefore be more relevant in reflecting the current findings of long-term changes in methylation levels.

# 6. Conclusions and Perspective

The field of arginine methylation has seen a surge in the number of known methylation sites in higher eukaryotes over recent years, however, our understanding of the functions of these modifications is lagging behind. Lower eukaryotes, e.g. *S. cerevisiae*, have successfully served as model organisms to enhance our understanding of cellular processes on multiple occasions (Duina et al. 2014). A better understanding of arginine methylation in yeast is therefore likely to inform the field also in higher eukaryotes. While there is a substantial body of literature regarding arginine methylation in *S. cerevisiae*, it is mainly centred around the same PRMT (Hmt1), the same context of methylation sites (RG-rich regions and RGG-motifs) and only a few methylated proteins (especially Npl3) (Low & Wilkins 2012). The application of state of the art proteomics approaches for the identification of further methylation sites in this organism has been largely missing before this study. Proteome-wide identification of methylation sites in *S. cerevisiae*, based on several enrichment and fractionation strategies and on the use of MethylSILAC to increase the confidence in identified sites was explored here.

The identified arginine methylation sites were mainly in RG-rich regions and were frequently found in RGG-motifs and often overlapped with previously known methylated proteins. However, 20 of the sites were novel and some did not conform to the previously observed trends. In particular, sites located in disordered regions that were likely not caused by a high RG-content, but a high frequency of asparagines, were detected.

As there is a high probability that this global approach is limited by biases inherent to affinity reagents and by its sensitivity, the analysis of methylation sites on purified candidate proteins was also employed. In the first instance, this strategy resulted in a surprisingly high number of methylation sites; however, follow-up experiments indicated

that considerable care needs to be taken to ensure that an observed site is not an artefact and is confidently localized.

Lsm4 is one protein for which mono- and di-methylation of one site (R119) was found by both approaches; the methylation site is located in an asparagine-, rather than RG-rich sequence context. While the role of arginine methylation on Sm proteins and LSM4 is well-established in higher eukaryotes (Brahms et al. 2001), the corresponding regions are not strongly conserved and little is known about the role of arginine methylation in splicing in yeast. Despite technical difficulties, the results indicate that the methylation of R119 on Lsm4 can neither be solely attributed to the activity of Hmt1 nor Hsl7. As arginine methylation plays a role in splicing mainly in the form of Rme2sym in higher eukaryotes, Lsm4 remains a candidate for a symmetrically di-methylated protein. While symmetric di-methylation of arginines exists in *S. cerevisiae*, no protein bearing this modification has been reported to date. Lsm4 may therefore prove useful in the quest for the elucidation of further PRMTs in this organism.

Furthermore, more is to be learned about the role of arginine methylation in RG-rich regions. Here, the methylation of Scd6 in its RG-rich region was described. It was shown that Hmt1 is the sole MTase catalysing arginine methylation of Scd6. Given the importance of the RG-rich region of this protein in translation inhibition upon stress (Rajyaguru et al. 2012), a change in Scd6 methylation levels upon prolonged (~2-4 h) sodium azide stress was demonstrated. This would be consistent with a hypothesis that translation inhibition may be altered upon stress via arginine methylation of Scd6.

One of the most immediate questions that need to be resolved in the detection of protein arginine methylation in *S. cerevisiae* is the apparent discrepancy between the low number of methylated proteins detected in global approaches compared to the high rate of detection of methylated proteins in the candidate-based approach. Three main

explanations may contribute to these observations: (a) Limited sensitivity and enrichment biases in global strategies prevent the detection of methylation sites that are detectable on purified proteins; (b) The analysis strategy used in the candidate approach may lead to an increased number of false positives and detection of methylation artefacts; (c) The methods used for determining candidate proteins may have been very selective for methylated proteins.

Therefore, firstly, the presence of false positives and artefacts in the candidate approach needs to be ruled out, e.g. by the use of MethylSILAC, as was demonstrated in this study for a subset of proteins. Secondly, it needs to be tested if the candidate selection indeed led to an enrichment of methylated proteins. This may be done by performing the candidate approach on a set of randomly selected proteins. Based on these results, a new strategy for the discovery of methylation sites may be devised: Methylated proteins may be enriched and identified in a first step, while the mapping of methylation sites is performed on purified proteins.

Subsequently, further methylated proteins that may be identified in this way and the methylated proteins presented in this study need to be tested for the MTase activities responsible for their methylation. This may especially enhance the understanding of methylation events catalysed by known or unknown PRMTs other than Hmt1. Attempts towards the assignment of PRMTs to methylation sites were made in this study and brought to light several technical limitations that need to be addressed.

Finally, the results presented here in combination with previous knowledge indicate that unstructured RG-rich regions are a good predictor of arginine methylation and proteins containing such sequences need to be tested for methylation and its functional consequences. If indeed any arginine or any arginine in a single RG-sequence context

within disordered regions, e.g. as in Lsm4, serves as a target for methylation, remains to be investigated.

# References


Anderson, J.T. et al., 1993. NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Molecular and Cellular Biology*, 13(5), pp.2730–2741.

Anon, 1994. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual* 1994 ed edition., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Pr.

Auclair, Y. & Richard, S., 2013. The role of arginine methylation in the DNA damage response. *DNA repair*, 12(7), pp.459–465.

Bachand, F., 2007. Protein arginine methyltransferases: from unicellular eukaryotes to humans. *Eukaryotic Cell*, 6(6), pp.889–898.

Bachand, F. & Silver, P.A., 2004. PRMT3 is a ribosomal protein methyltransferase that affects the cellular levels of ribosomal subunits. *The EMBO journal*, 23(13), pp.2641–2650.

Baldwin, G.S. & Carnegie, P.R., 1971. Isolation and partial characterization of methylated arginines from the encephalitogenic basic protein of myelin. *The Biochemical Journal*, 123(1), pp.69–74.

Bedford, M.T., 2007. Arginine methylation at a glance. *Journal of Cell Science*, 120(Pt 24), pp.4243–4246.

Bedford, M.T. et al., 2000. Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *The Journal of Biological Chemistry*, 275(21), pp.16030–16036.

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

Beggs, J.D., 2005. Lsm proteins and RNA processing. *Biochemical Society Transactions*, 33(Pt 3), pp.433–438.

Bertani, G., 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of Bacteriology*, 62(3), pp.293–300.

Beyer, A.L. et al., 1977. Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell*, 11(1), pp.127–138.

Biggar, K.K. & Li, S.S.-C., 2015. Non-histone protein methylation as a regulator of cellular signalling and function. *Nature Reviews. Molecular Cell Biology*, 16(1), pp.5–17.

Boeckel, J.-N. et al., 2011. Jumonji domain-containing protein 6 (Jmjd6) is required for angiogenic sprouting and regulates splicing of VEGF-receptor 1. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), pp.3276–3281.

Boisvert, F.-M. et al., 2003. A proteomic analysis of arginine-methylated protein complexes. *Molecular & cellular proteomics: MCP*, 2(12), pp.1319–1330.

Boisvert, F.-M. et al., 2002. Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. *The Journal of cell biology*, 159(6), pp.957–969.

Boisvert, F.-M., Chénard, C.A. & Richard, S., 2005. Protein interfaces in signaling regulated by arginine methylation. *Science's STKE: signal transduction knowledge environment*, 2005(271), p.re2.

Bond, A.T. et al., 2001. Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. *Molecular and Cellular Biology*, 21(21), pp.7366–7379.

Böttger, A. et al., 2015. The oxygenase Jmjd6--a case study in conflicting assignments. *The Biochemical Journal*, 468(2), pp.191–202.

Boulanger, M.-C. et al., 2005. Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. *Journal of Virology*, 79(1), pp.124–131.

Bouveret, E. et al., 2000. A Sm-like protein complex that participates in mRNA degradation. *The EMBO journal*, 19(7), pp.1661–1671.

Brahms, H. et al., 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA (New York, N.Y.)*, 7(11), pp.1531–1542.

Brahms, H. et al., 2000. The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *The Journal of Biological Chemistry*, 275(22), pp.17122–17129.

Brame, C.J., Moran, M.F. & McBroom-Cerajewski, L.D.B., 2004. A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. *Rapid communications in mass spectrometry: RCM*, 18(8), pp.877–881.

Bremang, M. et al., 2013. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Molecular bioSystems*, 9(9), pp.2231–2247.

Burd, C.G. & Dreyfuss, G., 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science (New York, N.Y.)*, 265(5172), pp.615–621.

Byvoet, P. et al., 1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of Biochemistry and Biophysics*, 148(2), pp.558–567.

Calnan, B.J. et al., 1991. Arginine-mediated RNA recognition: the arginine fork. *Science (New York, N.Y.)*, 252(5009), pp.1167–1171.

Cao, X.-J., Arnaudo, A.M. & Garcia, B.A., 2013. Large-scale global identification of protein lysine methylation in vivo. *Epigenetics*, 8(5), pp.477–485.

Castello, A. et al., 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149(6), pp.1393–1406.

Cavallius, J. et al., 1993. Characterization of yeast EF-1 alpha: non-conservation of post-translational modifications. *Biochimica et biophysica acta*, 1163(1), pp.75–80.

Chambers, M.C. et al., 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 30(10), pp.918–920.

Chang, B. et al., 2007. JMJD6 is a histone arginine demethylase. *Science (New York, N.Y.)*, 318(5849), pp.444–447.

Chang, F.N. et al., 1976. Methylation of ribosomal proteins in HeLa cells. *Archives of Biochemistry and Biophysics*, 172(2), pp.627–633.

Cheng, D. et al., 2007. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Molecular Cell*, 25(1), pp.71–83.

Chen, J.Y. & Bodley, J.W., 1988. Biosynthesis of diphthamide in Saccharomyces cerevisiae. Partial purification and characterization of a specific S-adenosylmethionine:elongation factor 2 methyltransferase. *The Journal of Biological Chemistry*, 263(24), pp.11692–11696.

Chen, Y.-C. et al., 2010. Protein arginine methylation facilitates cotranscriptional recruitment of pre-mRNA splicing factors. *Molecular and Cellular Biology*, 30(21), pp.5245–5256.

Chern, M.-K. et al., 2002. Yeast ribosomal protein L12 is a substrate of protein-arginine methyltransferase 2. *The Journal of Biological Chemistry*, 277(18), pp.15345–15353.

Cho, Jung-Nam, Jee-Youn Ryu, Young-Min Jeong, Jihye Park, Ji-Joon Song, Richard M. Amasino, Bosl Noh, and Yoo-Sun Noh. 2012. 'Control of Seed Germination by Light-Induced Histone Arginine Demethylation Activity'. *Developmental Cell* 22 (4): 736–48. doi:10.1016/j.devcel.2012.01.024.

Chuang, R.Y. et al., 1997. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science (New York, N.Y.)*, 275(5305), pp.1468–1471.

Cid, V.J. et al., 2001. Dynamic localization of the Swe1 regulator Hsl7 during the Saccharomyces cerevisiae cell cycle. *Molecular Biology of the Cell*, 12(6), pp.1645–1669.

Clarke, S.G., 2013. Protein methylation at the surface and buried deep: thinking outside the histone box. *Trends in Biochemical Sciences*, 38(5), pp.243–252.

Colaert, N. et al., 2009. Improved visualization of protein consensus sequences by iceLogo. *Nature Methods*, 6(11), pp.786–787.

Collins, R.E. et al., 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nature Structural & Molecular Biology*, 15(3), pp.245–250.

Corley, S.M. & Gready, J.E., 2008. Identification of the RGG box motif in Shadoo: RNA-binding and signaling roles? *Bioinformatics and Biology Insights*, 2, pp.383–400.

Côté, J. et al., 2003. Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Molecular biology of the cell*, 14(1), pp.274–287.

Côté, J. & Richard, S., 2005. Tudor domains bind symmetrical dimethylated arginines. *The Journal of Biological Chemistry*, 280(31), pp.28476–28483.

Couttas, T.A. et al., 2012. Methylation of translation-associated proteins in Saccharomyces cerevisiae: Identification of methylated lysines and their methyltransferases. *Proteomics*, 12(7), pp.960–972.

Creasy, D.M. & Cottrell, J.S., 2004. Unimod: Protein modifications for mass spectrometry. *Proteomics*, 4(6), pp.1534–1536.

Cuthbert, G.L. et al., 2004. Histone deimination antagonizes arginine methylation. *Cell*, 118(5), pp.545–553.

Dawson, M.A., Kouzarides, T. & Huntly, B.J.P., 2012. Targeting epigenetic readers in cancer. *The New England Journal of Medicine*, 367(7), pp.647–657.

Deng, X. et al., 2010. Arginine methylation mediated by the Arabidopsis homolog of PRMT5 is essential for proper pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44), pp.19114–19119.

Denman, R.B., 2002. Methylation of the arginine-glycine-rich region in the fragile X mental retardation protein FMRP differentially affects RNA binding. *Cellular & Molecular Biology Letters*, 7(3), pp.877–883.

Di Lorenzo, A. & Bedford, M.T., 2011. Histone arginine methylation. *FEBS letters*, 585(13), pp.2024–2031.

Duina, A.A., Miller, M.E. & Keeney, J.B., 2014. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. *Genetics*, 197(1), pp.33–48.

Erce, M.A. et al., 2013. Interactions affected by arginine methylation in the yeast protein-protein interaction network. *Molecular & cellular proteomics: MCP*, 12(11), pp.3184–3198.

Feng, Q. et al., 2009. Biochemical control of CARM1 enzymatic activity by phosphorylation. *The Journal of Biological Chemistry*, 284(52), pp.36167–36174.

Fernandez, C.F. et al., 2004. An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5. *Molecular Biology of the Cell*, 15(6), pp.2842–2852.

Fink, J.L. & Hamilton, N., 2007. DomainDraw: a macromolecular feature drawing program. *In Silico Biology*, 7(2), pp.145–150.

Formosa, T. & Nittis, T., 1998. Suppressors of the temperature sensitivity of DNA polymerase alpha mutations in Saccharomyces cerevisiae. *Molecular & general genetics: MGG*, 257(4), pp.461–468.

Frankel, A. & Clarke, S., 1999. RNase treatment of yeast and mammalian cell extracts affects in vitro substrate methylation by type I protein arginine N-methyltransferases. *Biochemical and Biophysical Research Communications*, 259(2), pp.391–400.

Friesen, W.J. et al., 2001. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Molecular Cell*, 7(5), pp.1111–1117.

Fu, Y. & Qian, X., 2014. Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry. *Molecular & cellular proteomics: MCP*, 13(5), pp.1359–1368.

Gao, W. et al., 2015. Arginine methylation of HSP70 regulates retinoid acid-mediated RARβ2 gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(26), pp.E3327–3336.

Gary, J.D. et al., 1996. The predominant protein-arginine methyltransferase from Saccharomyces cerevisiae. *The Journal of biological chemistry*, 271(21), pp.12585–12594.

Gehrig, P.M. et al., 2004. Fragmentation pathways of NG-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. *Journal of the American Society for Mass Spectrometry*, 15(2), pp.142–149.

Gelperin, D.M. et al., 2005. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development*, 19(23), pp.2816–2826.

Geoghegan, V. et al., 2015. Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. *Nature Communications*, 6. Available at: http://www.nature.com/ncomms/2015/150407/ncomms7758/full/ncomms7758.html [Accessed June 30, 2015].

Ghaemmaghami, S. et al., 2003. Global analysis of protein expression in yeast. *Nature*, 425(6959), pp.737–741.

Gilbert, W., Siebel, C.W. & Guthrie, C., 2001. Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA*, 7(2), pp.302–313.

Green, D.M. et al., 2002. Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. *The Journal of Biological Chemistry*, 277(10), pp.7752–7760.

Guo, A. et al., 2014. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Molecular & cellular proteomics: MCP*, 13(1), pp.372–387.

Hanakahi, L.A., Sun, H. & Maizels, N., 1999. High affinity interactions of nucleolin with G-G-paired rDNA. *The Journal of Biological Chemistry*, 274(22), pp.15908–15912.

Han, X. et al., 2011. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *Journal of Proteome Research*, 10(7), pp.2930–2936.

Hartley, J.L., Temple, G.F. & Brasch, M.A., 2000. DNA cloning using in vitro site-specific recombination. *Genome Research*, 10(11), pp.1788–1795.

Hart-Smith, G. et al., 2012. Enhanced methylarginine characterization by post-translational modification-specific targeted data acquisition and electron-transfer dissociation mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 23(8), pp.1376–1389.

Henry, M.F. & Silver, P.A., 1996. A novel methyltransferase (Hmt1p) modifies poly(A)+-RNA-binding proteins. *Molecular and Cellular Biology*, 16(7), pp.3668–3678.

Heurgué-Hamard, V. et al., 2005. The glutamine residue of the conserved GGQ motif in Saccharomyces cerevisiae release factor eRF1 is methylated by the product of the YDR140w gene. *The Journal of Biological Chemistry*, 280(4), pp.2439–2445.

Heurgué-Hamard, V. et al., 2002. The hemK gene in Escherichia coli encodes the N(5)-glutamine methyltransferase that modifies peptide release factors. *The EMBO journal*, 21(4), pp.769–778.

Higashimoto, K. et al., 2007. Phosphorylation-mediated inactivation of coactivator-associated arginine methyltransferase 1. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), pp.12318–12323.

Hilliker, A. et al., 2011. The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Molecular Cell*, 43(6), pp.962–972.

Hopkinson, R.J. et al., 2013. 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chemical Science*, 4(8), pp.3110–3117.

Hsieh, C.-H. et al., 2007. Expression of proteins with dimethylarginines in Escherichia coli for protein-protein interaction studies. *Protein Science: A Publication of the Protein Society*, 16(5), pp.919–928.

Huang, B., Johansson, M.J.O. & Byström, A.S., 2005. An early step in wobble uridine tRNA modification requires the Elongator complex. *RNA (New York, N.Y.)*, 11(4), pp.424–436.

Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), pp.1–13.

Hughes, R.M. & Waters, M.L., 2006. Arginine methylation in a beta-hairpin peptide: implications for Arg-pi interactions, DeltaCp(o), and the cold denatured state. *Journal of the American Chemical Society*, 128(39), pp.12735–12742.

Hulsen, T., de Vlieg, J. & Alkema, W., 2008. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics*, 9, p.488.

Hung, M.-L. et al., 2010. Arginine methylation of REF/ALY promotes efficient handover of mRNA to TAP/NXF1. *Nucleic Acids Research*, 38(10), pp.3351–3361.

Hwang, W.W. et al., 2003. A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Molecular Cell*, 11(1), pp.261–266.

Hyun, S., Jeong, S. & Yu, J., 2008. Effects of asymmetric arginine dimethylation on RNA-binding peptides. *Chembiochem: A European Journal of Chemical Biology*, 9(17), pp.2790–2792.

Iberg, A.N. et al., 2008. Arginine methylation of the histone H3 tail impedes effector binding. *The Journal of biological chemistry*, 283(6), pp.3006–3010.

Inoue, K. et al., 2000. Novel RING finger proteins, Air1p and Air2p, interact with Hmt1p and inhibit the arginine methylation of Npl3p. *The Journal of Biological Chemistry*, 275(42), pp.32793–32799.

Invernizzi, C.F. et al., 2007. Arginine methylation of the HIV-1 nucleocapsid protein results in its diminished function. *AIDS (London, England)*, 21(7), pp.795–805.

Invernizzi, C.F. et al., 2006. PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. *Retrovirology*, 3, p.93.

Iost, I., Dreyfus, M. & Linder, P., 1999. Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. *The Journal of Biological Chemistry*, 274(25), pp.17677–17683.

Johnson, B.A. et al., 1987. Protein carboxyl methyltransferase facilitates conversion of atypical L-isoaspartyl peptides to normal L-aspartyl peptides. *The Journal of Biological Chemistry*, 262(12), pp.5622–5629.

Jung, S.Y. et al., 2008. Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. *Analytical chemistry*, 80(5), pp.1721–1729.

Keilin, D., 1936. The Action of Sodium Azide on Cellular Respiration and on some Catalytic Oxidation Reactions. *Proceedings of the Royal Society of London B: Biological Sciences*, 121(822), pp.165–173.

Kiledjian, M. & Dreyfuss, G., 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *The EMBO journal*, 11(7), pp.2655–2664.

Kirmizis, A. et al., 2007. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature*, 449(7164), pp.928–932.

Kirmizis, A. et al., 2009. Distinct transcriptional outputs associated with mono- and dimethylated histone H3 arginine 2. *Nature Structural & Molecular Biology*, 16(4), pp.449–451.

Kondo, K. & Inouye, M., 1992. Yeast NSR1 protein that has structural similarity to mammalian nucleolin is involved in pre-rRNA processing. *The Journal of Biological Chemistry*, 267(23), pp.16252–16258.

Krause, C.D. et al., 2007. Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacology & Therapeutics*, 113(1), pp.50–87.

Krogan, N.J. & Greenblatt, J.F., 2001. Characterization of a six-subunit holo-elongator complex required for the regulated expression of a group of genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 21(23), pp.8203–8212.

Kufel, J. et al., 2003. Lsm Proteins are required for normal processing and stability of ribosomal RNAs. *The Journal of Biological Chemistry*, 278(4), pp.2147–2156.

Kufel, J. et al., 2002. Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Molecular and Cellular Biology*, 22(14), pp.5248–5256.

Lacoste, N. et al., 2002. Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *The Journal of biological chemistry*, 277(34), pp.30421–30424.

Lacroix, M. et al., 2008. The histone-binding protein COPR5 is required for nuclear functions of the protein arginine methyltransferase PRMT5. *EMBO reports*, 9(5), pp.452–458.

Lakowski, T.M. et al., 2015. Arginine methylation in yeast proteins during stationary-phase growth and heat shock. *Amino Acids*.

Lakowski, T.M. et al., 2013. MS3 fragmentation patterns of monomethylarginine species and the quantification of all methylarginine species in yeast using MRM3. *Journal of Proteomics*, 80, pp.43–54.

Lakowski, T.M. et al., 2010. Nη-substituted arginyl peptide inhibitors of protein arginine N-methyltransferases. *ACS chemical biology*, 5(11), pp.1053–1063.

Lapko, V.N. et al., 2005. Modifications of human betaA1/betaA3-crystallins include S-methylation, glutathiolation, and truncation. *Protein Science: A Publication of the Protein Society*, 14(1), pp.45–54.

Lee, J.H. et al., 2000. Hsl7p, the yeast homologue of human JBP1, is a protein methyltransferase. *Biochemical and Biophysical Research Communications*, 274(1), pp.105–111.

Lee, M.G. et al., 2006. Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chemistry & Biology*, 13(6), pp.563–567.

van Leeuwen, F., Gafken, P.R. & Gottschling, D.E., 2002. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell*, 109(6), pp.745–756.

Lee, W.C., Xue, Z.X. & Mélèse, T., 1991. The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. *The Journal of Cell Biology*, 113(1), pp.1–12.

Lee, W.C., Zabetakis, D. & Mélèse, T., 1992. NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. *Molecular and Cellular Biology*, 12(9), pp.3865–3871.

Lhoest, J. et al., 1984. Methylated proteins and amino acids in the ribosomes of Saccharomyces cerevisiae. *European journal of biochemistry / FEBS*, 141(3), pp.585–590.

Li, C. & Clarke, S., 1992. A protein methyltransferase specific for altered aspartyl residues is important in Escherichia coli stationary-phase survival and heat-shock resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 89(20), pp.9885–9889.

Li, H. et al., 2007. Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Molecular Cell*, 28(4), pp.677–691.

Lipson, R.S., Webb, K.J. & Clarke, S.G., 2010. Rmt1 catalyzes zinc-finger independent arginine methylation of ribosomal protein Rps2 in Saccharomyces cerevisiae. *Biochemical and biophysical research communications*, 391(4), pp.1658–1662.

L'Italien, J.J. & Laursen, R.A., 1979. Location of the site of methylation in elongation factor Tu. *FEBS letters*, 107(2), pp.359–362.

Liu, H. et al., 2013. A method for systematic mapping of protein lysine methylation identifies functions for HP1β in DNA damage response. *Molecular Cell*, 50(5), pp.723–735.

Liu, Q. & Dreyfuss, G., 1995. In vivo and in vitro arginine methylation of RNA-binding proteins. *Molecular and cellular biology*, 15(5), pp.2800–2808.

Low, J.K.K. et al., 2013. Analysis of the Proteome of Saccharomyces cerevisiae for Methylarginine. *Journal of Proteome Research*, 12(9), pp.3884–3899.

Low, J.K.K. et al., 2014. The Saccharomyces cerevisiae poly(A)-binding protein is subject to multiple post-translational modifications, including the methylation of glutamic acid. *Biochemical and Biophysical Research Communications*, 443(2), pp.543–548.

Low, J.K.K. & Wilkins, M.R., 2012. Protein arginine methylation in Saccharomyces cerevisiae. *The FEBS journal*, 279(24), pp.4423–4443.

Lukasiewicz, R. et al., 2007. The RGG domain of Npl3p recruits Sky1p through docking interactions. *Journal of Molecular Biology*, 367(1), pp.249–261.

Lu, R. & Wang, G.G., 2013. Tudor: a versatile family of histone methylation "readers." *Trends in Biochemical Sciences*, 38(11), pp.546–555.

Luscombe, N.M., Laskowski, R.A. & Thornton, J.M., 2001. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Research*, 29(13), pp.2860–2874.

Magrane, M. & Consortium, U., 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database: The Journal of Biological Databases and Curation*, 2011, p.bar009.

Malmström, L. et al., 2007. Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS biology*, 5(4), p.e76.

McBride, A.E. et al., 2000. Analysis of the yeast arginine methyltransferase Hmt1p/Rmt1p and its in vivo function. Cofactor binding and substrate interactions. *The Journal of Biological Chemistry*, 275(5), pp.3128–3136.

McBride, A.E. et al., 2005. Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions. *The Journal of Biological Chemistry*, 280(35), pp.30888–30898.

Mears, W.E. & Rice, S.A., 1996. The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *Journal of Virology*, 70(11), pp.7445–7453.

Meng, F. et al., 2004. Molecular-level description of proteins from saccharomyces cerevisiae using quadrupole FT hybrid mass spectrometry for top down proteomics. *Analytical Chemistry*, 76(10), pp.2852–2858.

Meyer, M. & Vilardell, J., 2009. The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. *Briefings in Functional Genomics & Proteomics*, 8(1), pp.60–67.

Migliori, V. et al., 2012. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nature Structural & Molecular Biology*, 19(2), pp.136–144.

Miranda, T.B. et al., 2006. Yeast Hsl7 (histone synthetic lethal 7) catalyses the in vitro formation of omega-N(G)-monomethylarginine in calf thymus histone H2A. *The Biochemical Journal*, 395(3), pp.563–570.

Moore, K.E. et al., 2013. A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Molecular cell*, 50(3), pp.444–456.

Morgan, D.G., Baumgartner, J.W. & Hazelbauer, G.L., 1993. Proteins antigenically related to methyl-accepting chemotaxis proteins of Escherichia coli detected in a wide range of bacterial species. *Journal of Bacteriology*, 175(1), pp.133–140.

Nady, N. et al., 2012. Histone recognition by human malignant brain tumor domains. *Journal of Molecular Biology*, 423(5), pp.702–718.

Najbauer, J. et al., 1993. 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. *The Journal of Biological Chemistry*, 268(14), pp.10501–10509.

Najbauer, J. & Aswad, D.W., 1990. Diversity of methyl acceptor proteins in rat pheochromocytoma (PC12) cells revealed after treatment with adenosine dialdehyde. *The Journal of Biological Chemistry*, 265(21), pp.12717–12721.

Ng, H.H. et al., 2002. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes & Development*, 16(12), pp.1518–1527.

Nichols, R.C. et al., 2000. The RGG domain in hnRNP A2 affects subcellular localization. *Experimental Cell Research*, 256(2), pp.522–532.

Niewmierzycka, A. & Clarke, S., 1999. S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase. *The Journal of biological chemistry*, 274(2), pp.814–824.

Olsson, I. et al., 2007. The arginine methyltransferase Rmt2 is enriched in the nucleus and co-purifies with the nuclear porins Nup49, Nup57 and Nup100. *Experimental Cell Research*, 313(9), pp.1778–1789.

Ong, S.-E., Mittler, G. & Mann, M., 2004. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature methods*, 1(2), pp.119–126.

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

Paik, W.K. & Kim, S., 1970. Omega-N-methylarginine in histones. *Biochemical and Biophysical Research Communications*, 40(1), pp.224–229.

Paik, W.K. & Kim, S., 1968. Protein methylase I. Purification and properties of the enzyme. *The Journal of Biological Chemistry*, 243(9), pp.2108–2114.

Pak, M.L. et al., 2011. A protein arginine N-methyltransferase 1 (PRMT1) and 2 heteromeric interaction increases PRMT1 enzymatic activity. *Biochemistry*, 50(38), pp.8226–8240.

Pang, C.N.I., Gasteiger, E. & Wilkins, M.R., 2010. Identification of arginine- and lysine-methylation in the proteome of Saccharomyces cerevisiae and its functional implications. *BMC genomics*, 11, p.92.

Passos, D.O., Quaresma, A.J.C. & Kobarg, J., 2006. The methylation of the C-terminal region of hnRNPQ (NSAP1) is important for its nuclear localization. *Biochemical and Biophysical Research Communications*, 346(2), pp.517–525.

Plank, M. et al., 2015. Expanding the yeast protein arginine methylome. *Proteomics*.

Polevoda, B. & Sherman, F., 2007. Methylation of proteins involved in translation. *Molecular microbiology*, 65(3), pp.590–606.

Poulard, C. et al., 2014. JMJD6 regulates ERα methylation on arginine. *PloS One*, 9(2), p.e87982.

Rajpurohit, R., Paik, W.K. & Kim, S., 1994. Effect of enzymic methylation of heterogeneous ribonucleoprotein particle A1 on its nucleic-acid binding and controlled proteolysis. *The Biochemical Journal*, 304 ( Pt 3), pp.903–909.

Rajyaguru, P. & Parker, R., 2012. RGG motif proteins: modulators of mRNA functional states. *Cell Cycle (Georgetown, Tex.)*, 11(14), pp.2594–2599.

Rajyaguru, P., She, M. & Parker, R., 2012. Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. *Molecular Cell*, 45(2), pp.244–254.

Ren, J. et al., 2010. Methylation of ribosomal protein S10 by protein-arginine methyltransferase 5 regulates ribosome biogenesis. *The Journal of Biological Chemistry*, 285(17), pp.12695–12705.

Rossmann, M.G., Moras, D. & Olsen, K.W., 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature*, 250(463), pp.194–199.

Ruault, M. & Pillus, L., 2006. Chromatin-modifiying enzymes are essential when the Saccharomyces cerevisiae morphogenesis checkpoint is constitutively activated. *Genetics*, 174(3), pp.1135–1149.

Salgado-Garrido, J. et al., 1999. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *The EMBO journal*, 18(12), pp.3451–3462.

Sanchez, S.E. et al., 2010. A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature*, 468(7320), pp.112–116.

Sayegh, J. & Clarke, S.G., 2008. Hsl7 is a substrate-specific type II protein arginine methyltransferase in yeast. *Biochemical and Biophysical Research Communications*, 372(4), pp.811–815.

Schimmang, T. et al., 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *The EMBO journal*, 8(13), pp.4015–4024.

Schlosser, A. et al., 2002. Identification of protein phosphorylation sites by combination of elastase digestion, immobilized metal affinity chromatography, and quadrupole-time of flight tandem mass spectrometry. *Proteomics*, 2(7), pp.911–918.

Scott, J.M. & Weir, D.G., 1981. The methyl folate trap. A physiological response in man to prevent methyl group deficiency in kwashiorkor (methionine deficiency) and an explanation for folic-acid induced exacerbation of subacute combined degeneration in pernicious anaemia. *Lancet (London, England)*, 2(8242), pp.337–340.

Shen, E.C. et al., 1998. Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes & Development*, 12(5), pp.679–691.

Shi, Y. et al., 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7), pp.941–953.

Siebel, C.W. et al., 1999. Conservation in budding yeast of a kinase specific for SR splicing factors. *Proceedings of the National Academy of Sciences of the United States of America*, 96(10), pp.5440–5445.

Sievers, F. & Higgins, D.G., 2014. Clustal omega. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, 48, pp.3.13.1–3.13.16.

Smith, P.K. et al., 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1), pp.76–85.

Smith, W.A. et al., 2004. Arginine methylation of RNA helicase a determines its subcellular localization. *The Journal of Biological Chemistry*, 279(22), pp.22795–22798.

Sprangers, R. et al., 2003. High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *Journal of Molecular Biology*, 327(2), pp.507–520.

Sprung, R. et al., 2008. Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *Journal of proteome research*, 7(3), pp.1001–1006.

Sun, L. et al., 2011. Structural insights into protein arginine symmetric dimethylation by PRMT5. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), pp.20538–20543.

Swanson, R.V. & Glazer, A.N., 1990. Phycobiliprotein methylation. Effect of the gamma-N-methylasparagine residue on energy transfer in phycocyanin and the phycobilisome. *Journal of Molecular Biology*, 214(3), pp.787–796.

Sylvestersen, K.B. et al., 2014. Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. *Molecular & cellular proteomics: MCP*, 13(8), pp.2072–2088.

Tang, J. et al., 1998. PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *The Journal of Biological Chemistry*, 273(27), pp.16935–16945.

Thandapani, P. et al., 2013. Defining the RGG/RG motif. *Molecular Cell*, 50(5), pp.613–623.

Theesfeld, C.L. et al., 2003. A monitor for bud emergence in the yeast morphogenesis checkpoint. *Molecular Biology of the Cell*, 14(8), pp.3280–3291.

Thomas, D. & Surdin-Kerjan, Y., 1997. Metabolism of sulfur amino acids in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews: MMBR*, 61(4), pp.503–532.

Tkaczuk, K.L. et al., 2007. Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC bioinformatics*, 8, p.73.

Tollervey, D. et al., 1991. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *The EMBO journal*, 10(3), pp.573–583.

Tripsianes, K. et al., 2011. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nature Structural & Molecular Biology*, 18(12), pp.1414–1420.

Tsukada, Y. et al., 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, 439(7078), pp.811–816.

Tutucci, E. & Stutz, F., 2011. Keeping mRNPs in check during assembly and nuclear export. *Nature Reviews. Molecular Cell Biology*, 12(6), pp.377–384.

Uhlmann, T. et al., 2012. A method for large-scale identification of protein arginine methylation. *Molecular & cellular proteomics: MCP*, 11(11), pp.1489–1499.

Valentini, S.R., Weiss, V.H. & Silver, P.A., 1999. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *RNA (New York, N.Y.)*, 5(2), pp.272–280.

Van Noort, J.M. et al., 1986. Methylation in vivo of elongation factor EF-Tu at lysine-56 decreases the rate of tRNA-dependent GTP hydrolysis. *European journal of biochemistry / FEBS*, 160(3), pp.557–561.

Waldherr, M. et al., 1993. A multitude of suppressors of group II intron-splicing defects in yeast. *Current Genetics*, 24(4), pp.301–306.

Wang, B. et al., 2008. Evaluation of the low-specificity protease elastase for large-scale phosphoproteome analysis. *Analytical Chemistry*, 80(24), pp.9526–9533.

Wang, H. et al., 2001. Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science (New York, N.Y.)*, 293(5531), pp.853–857.

Wang, K. et al., 2015. Proteomic analysis of protein methylation in the yeast Saccharomyces cerevisiae. *Journal of Proteomics*, 114, pp.226–233.

Webb, K.J. et al., 2010. A novel 3-methylhistidine modification of yeast ribosomal protein Rpl3 is dependent upon the YIL110W methyltransferase. *The Journal of Biological Chemistry*, 285(48), pp.37598–37606.

Webb, K.J. et al., 2008. Identification of two SET domain proteins required for methylation of lysine residues in yeast ribosomal protein Rpl42ab. *The Journal of biological chemistry*, 283(51), pp.35561–35568.

Webby, C.J. et al., 2009. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science (New York, N.Y.)*, 325(5936), pp.90–93.

Wei, H.-M. et al., 2014. Arginine methylation of the cellular nucleic acid binding protein does not affect its subcellular localization but impedes RNA binding. *FEBS Letters*, 588(9), pp.1542–1548.

Weiss, V.H. et al., 2000. The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nature Structural Biology*, 7(12), pp.1165–1171.

Wilkins, M.R. et al., 1999. High-throughput mass spectrometric discovery of protein post-translational modifications. *Journal of Molecular Biology*, 289(3), pp.645–657.

Winkler, G.S. et al., 2002. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), pp.3517–3522.

Wong, C.-M. et al., 2010. Yeast arginine methyltransferase Hmt1p regulates transcription elongation and termination by methylating Npl3p. *Nucleic Acids Research*, 38(7), pp.2217–2228.

Xie, H. & Clarke, S., 1993. Methyl esterification of C-terminal leucine residues in cytosolic 36-kDa polypeptides of bovine brain. A novel eucaryotic protein carboxyl methylation reaction. *The Journal of Biological Chemistry*, 268(18), pp.13364–13371.

Xu, C. et al., 2003. In vivo analysis of nucleolar proteins modified by the yeast arginine methyltransferase Hmt1/Rmt1p. *RNA*, 9(6), pp.746–759.

Xu, C. & Henry, M.F., 2004. Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p. *Molecular and Cellular Biology*, 24(24), pp.10742–10756.

Yagoub, D. et al., 2015. Yeast proteins Gar1p, Nop1p, Npl3p, Nsr1p and Rps2p are natively methylated and are substrates of the arginine methyltransferase Hmt1p. *Proteomics*.

Yan, C. & Mélèse, T., 1993. Multiple regions of NSR1 are sufficient for accumulation of a fusion protein within the nucleolus. *The Journal of Cell Biology*, 123(5), pp.1081–1091.

Yong, J., Wan, L. & Dreyfuss, G., 2004. Why do cells need an assembly machine for RNA-protein complexes? *Trends in Cell Biology*, 14(5), pp.226–232.

Young, B.D. et al., 2012. Identification of methylated proteins in the yeast small ribosomal subunit: a role for SPOUT methyltransferases in protein arginine methylation. *Biochemistry*, 51(25), pp.5091–5104.

Young, P.J. et al., 2003. The Ewing's sarcoma protein interacts with the Tudor domain of the survival motor neuron protein. *Brain Research. Molecular Brain Research*, 119(1), pp.37–49.

Yu, M.C. et al., 2004. Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes & development*, 18(16), pp.2024–2035.

Yu, M.C. et al., 2006. The role of protein arginine methylation in the formation of silent chromatin. *Genes & Development*, 20(23), pp.3249–3254.

Yun, C.Y. & Fu, X.D., 2000. Conserved SR protein kinase functions in nuclear import and its action is counteracted by arginine methylation in Saccharomyces cerevisiae. *The Journal of Cell Biology*, 150(4), pp.707–718.

Zanotti, K.J. et al., 2006. Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry*, 45(27), pp.8319–8330.

Zhang, J. et al., 2011. PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*, p.mcp.M111.010587.

Zhang, X. et al., 2005. Transcriptional regulation by Lge1p requires a function independent of its role in histone H2B ubiquitination. *The Journal of Biological Chemistry*, 280(4), pp.2759–2770.

Zhang, X. & Cheng, X., 2003. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure (London, England: 1993)*, 11(5), pp.509–520.

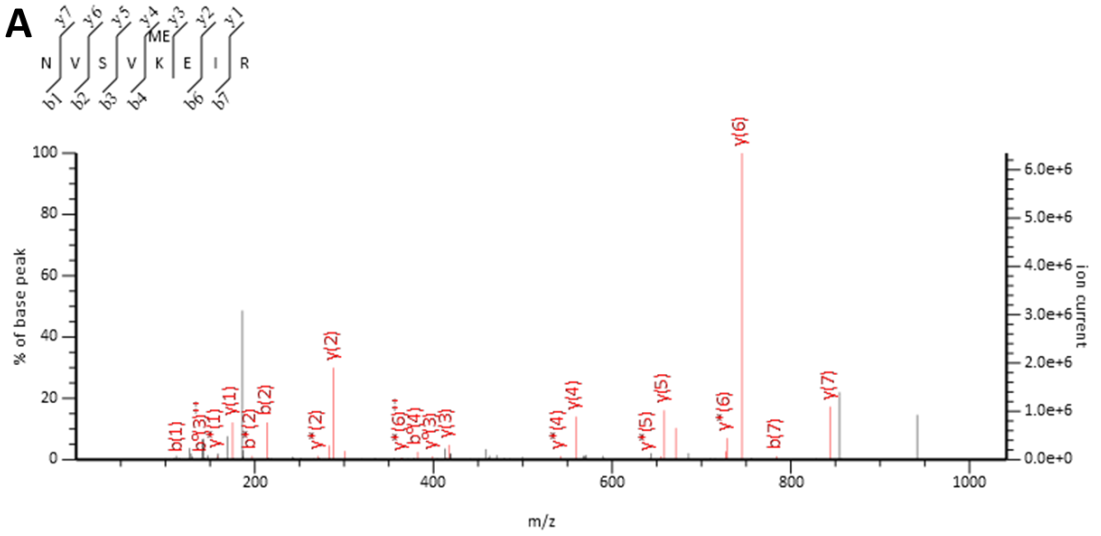
Zhang, X., Zhou, L. & Cheng, X., 2000. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *The EMBO journal*, 19(14), pp.3509–3519.

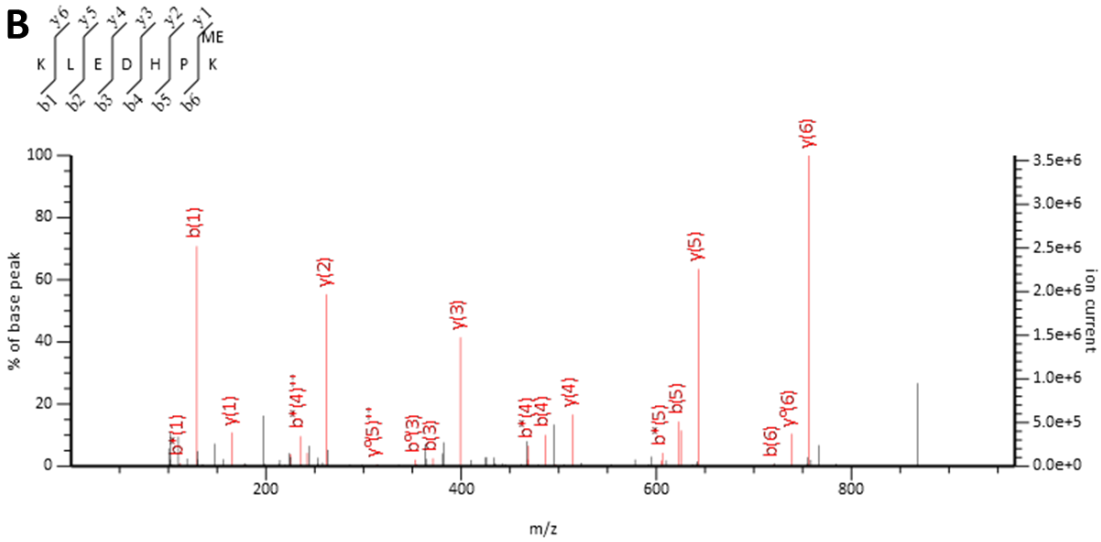
Zhang, Z. et al., 2011. Arabidopsis floral initiator SKB1 confers high salt tolerance by regulating transcription and pre-mRNA splicing through altering histone H4R3 and small nuclear ribonucleoprotein LSM4 methylation. *The Plant Cell*, 23(1), pp.396–411.

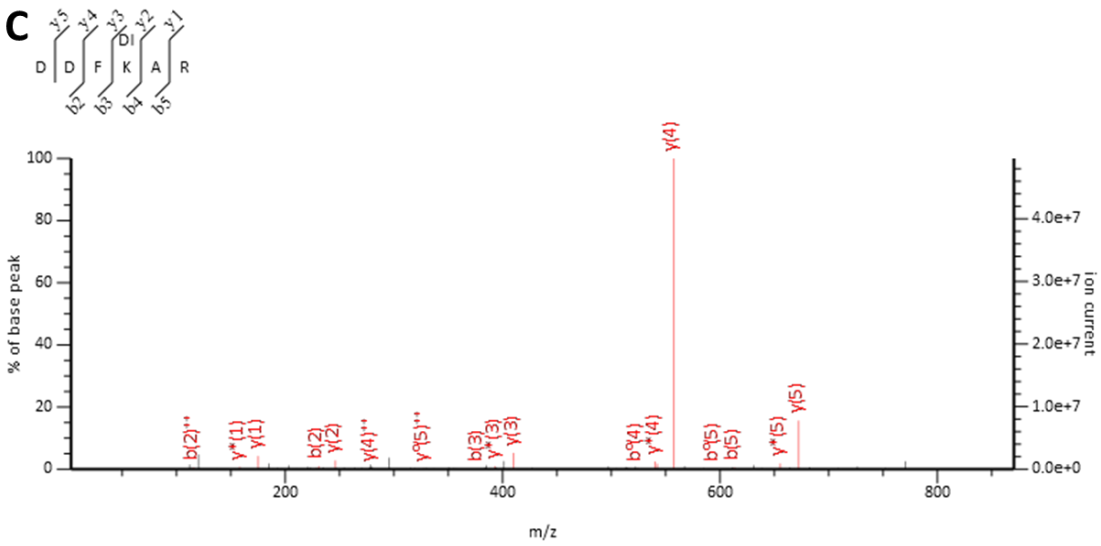
Zobel-Thropp, P. et al., 2000. A novel post-translational modification of yeast elongation factor 1A. Methylesterification at the C terminus. *The Journal of biological chemistry*, 275(47), pp.37150–37158.

Zobel-Thropp, P., Gary, J.D. & Clarke, S., 1998. delta-N-methylarginine is a novel posttranslational modification of arginine residues in yeast proteins. *The Journal of Biological Chemistry*, 273(45), pp.29283–29286.

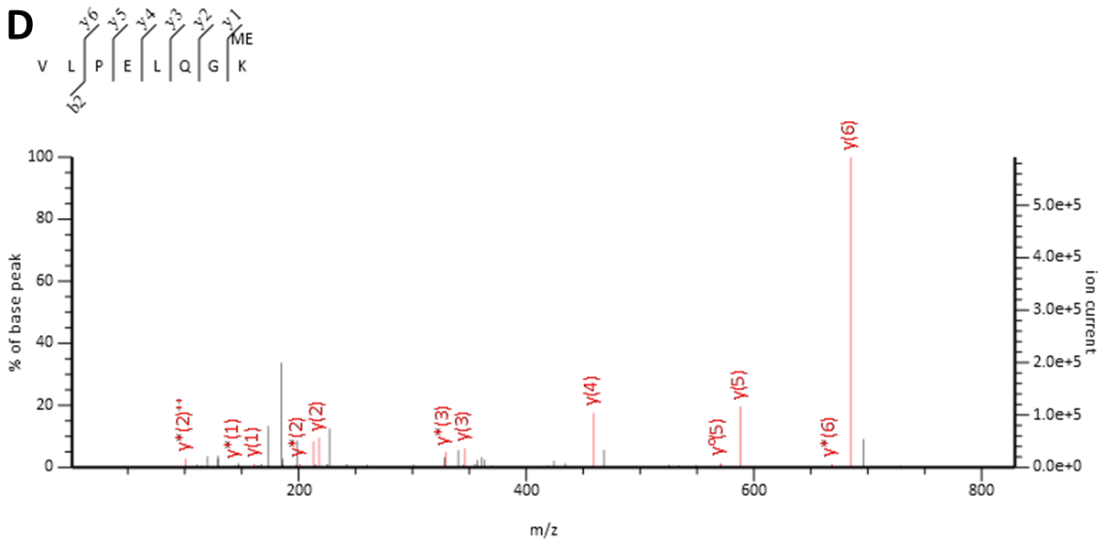
# Appendix Figures



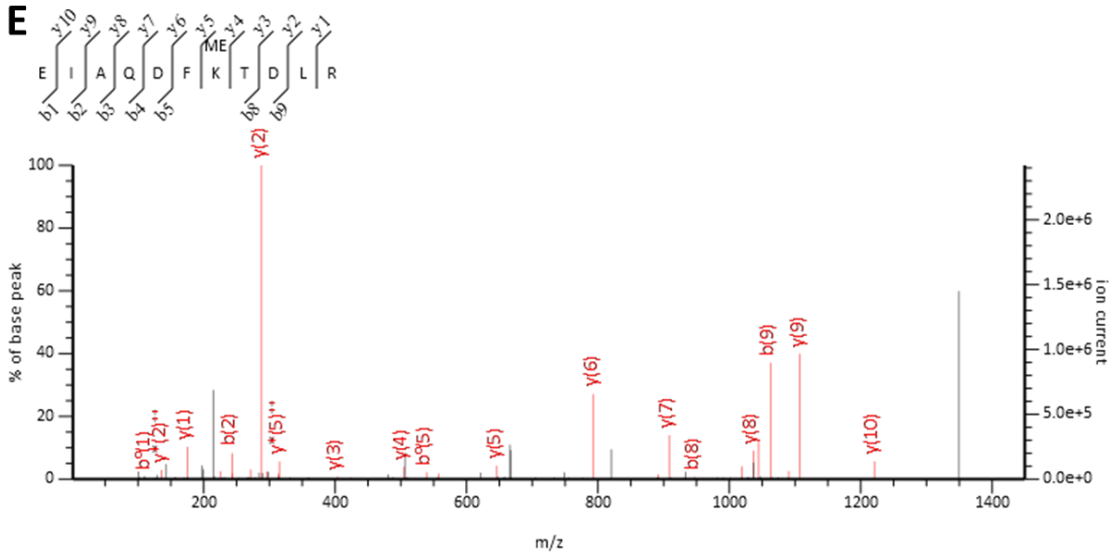



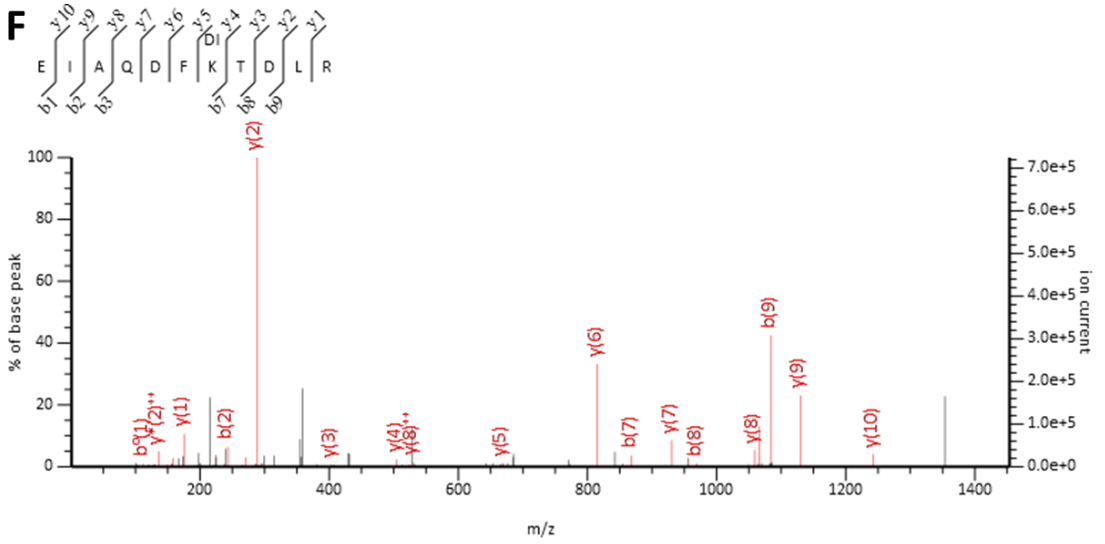
**D**


V L P E L Q G K ME
% of base peak
ion current
m/z

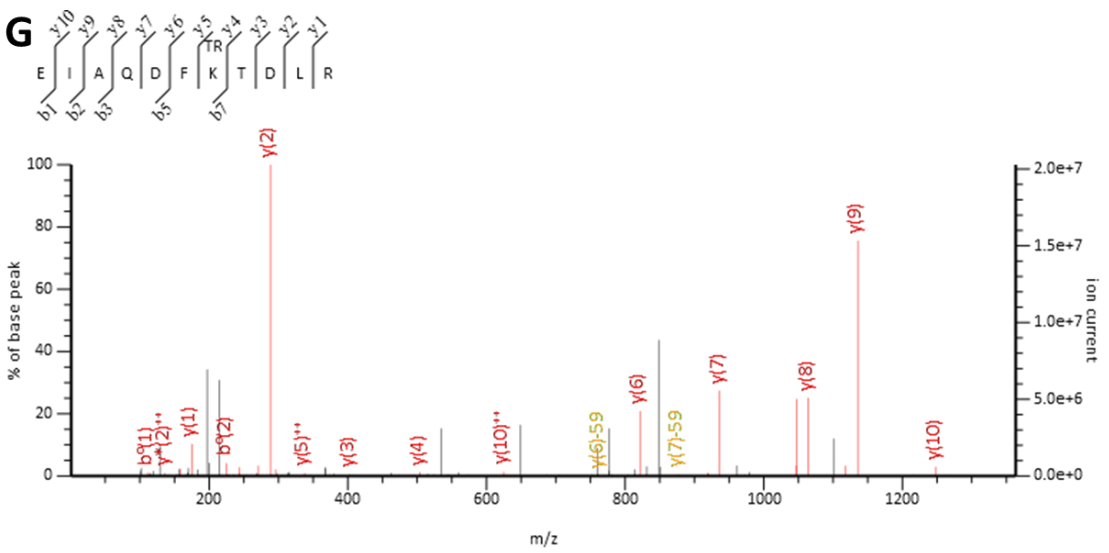
**E**


E I A Q D F K T D L R ME
% of base peak
ion current
m/z

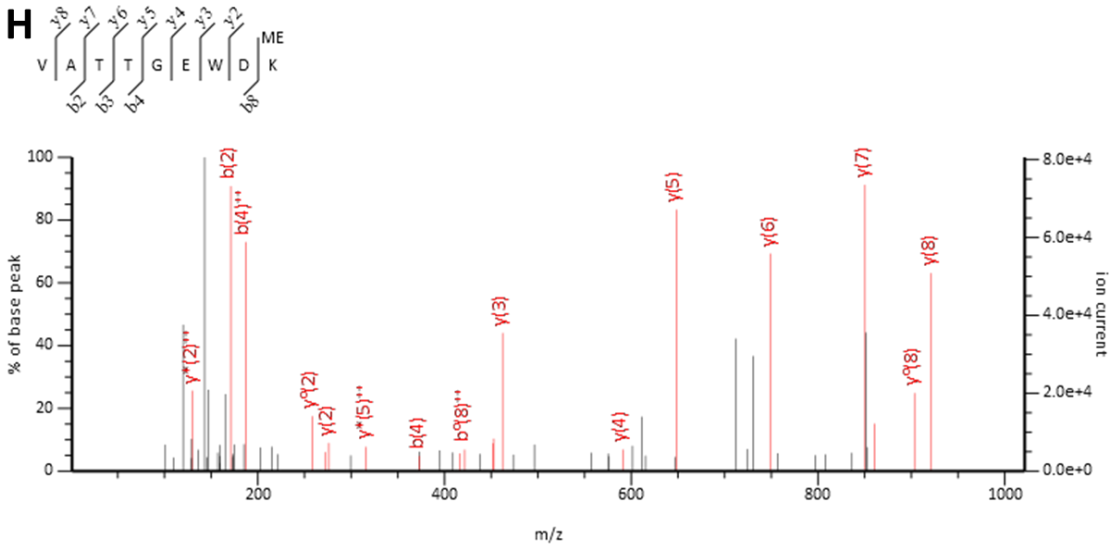
**F**


E I A Q D F K T D L R DI
% of base peak
ion current
m/z

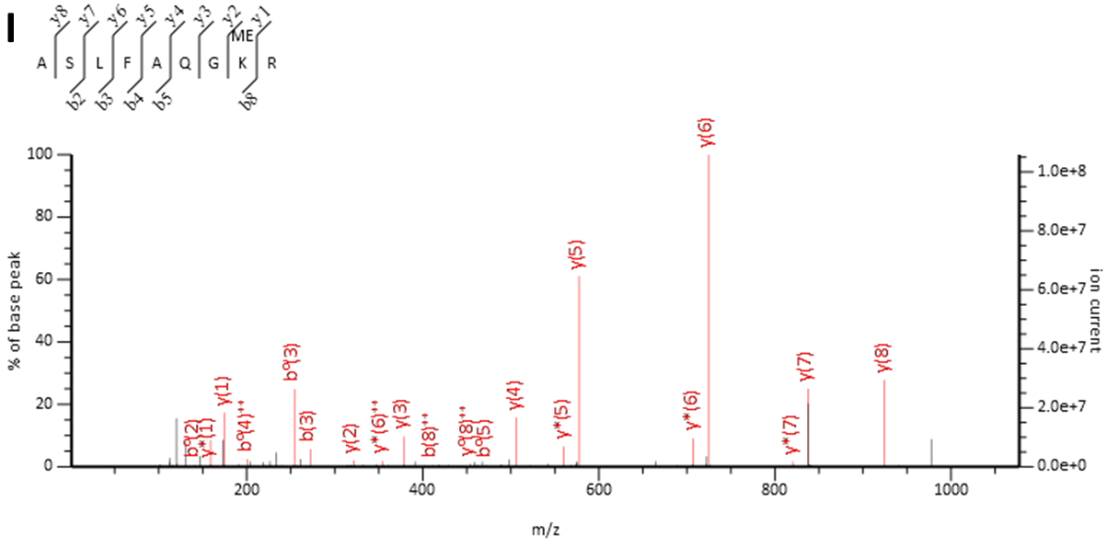
**G**


E I A Q D F TR K T D L R
% of base peak
m/z
ion current

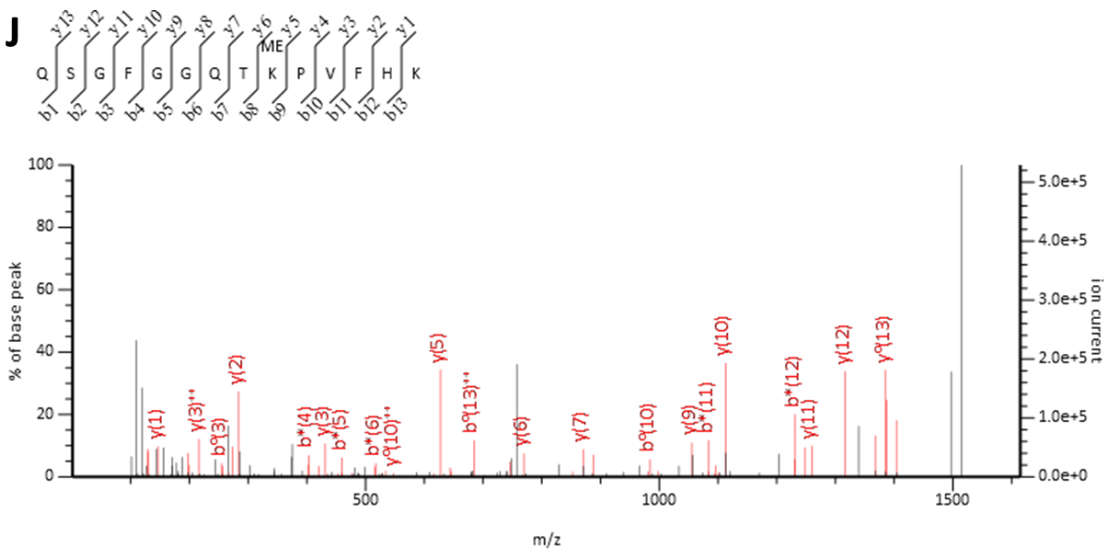
**H**


V A T T G E W D ME K
% of base peak
m/z
ion current

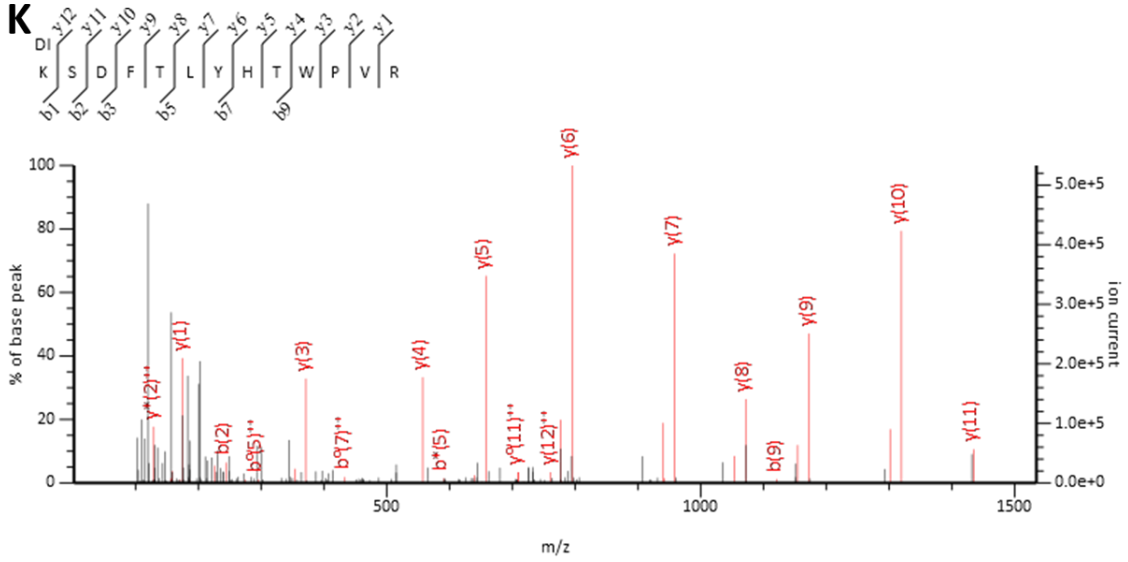
**I**


A S L F A Q G ME K R
% of base peak
m/z
ion current

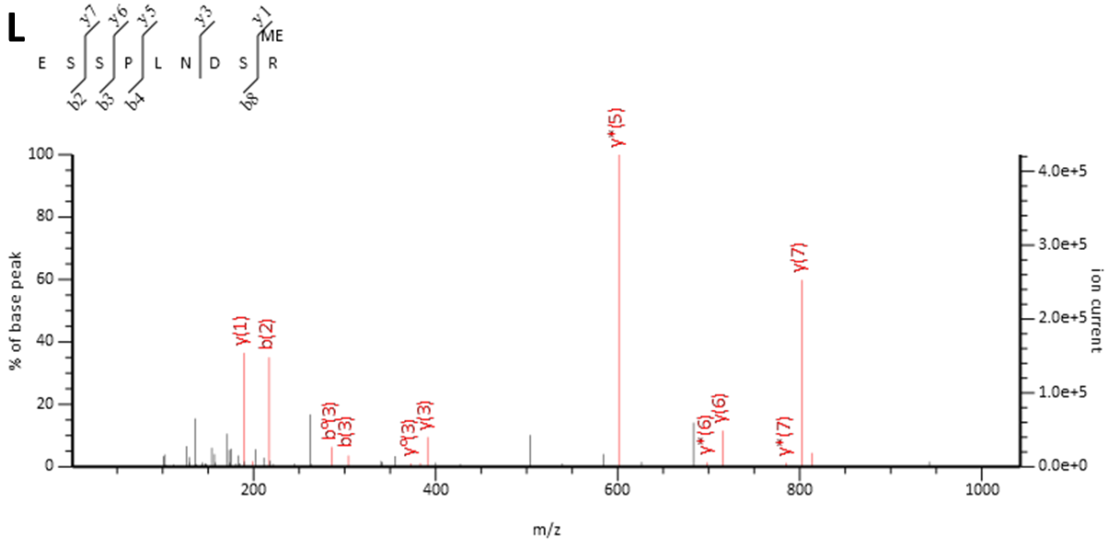
**J**


Q S G F G G Q T ME K P V F H K
% of base peak
ion current
m/z

**K**


DI K S D F T L Y H T W P V R
% of base peak
ion current
m/z

**L**


E S S P L N D S ME R
% of base peak
ion current
m/z

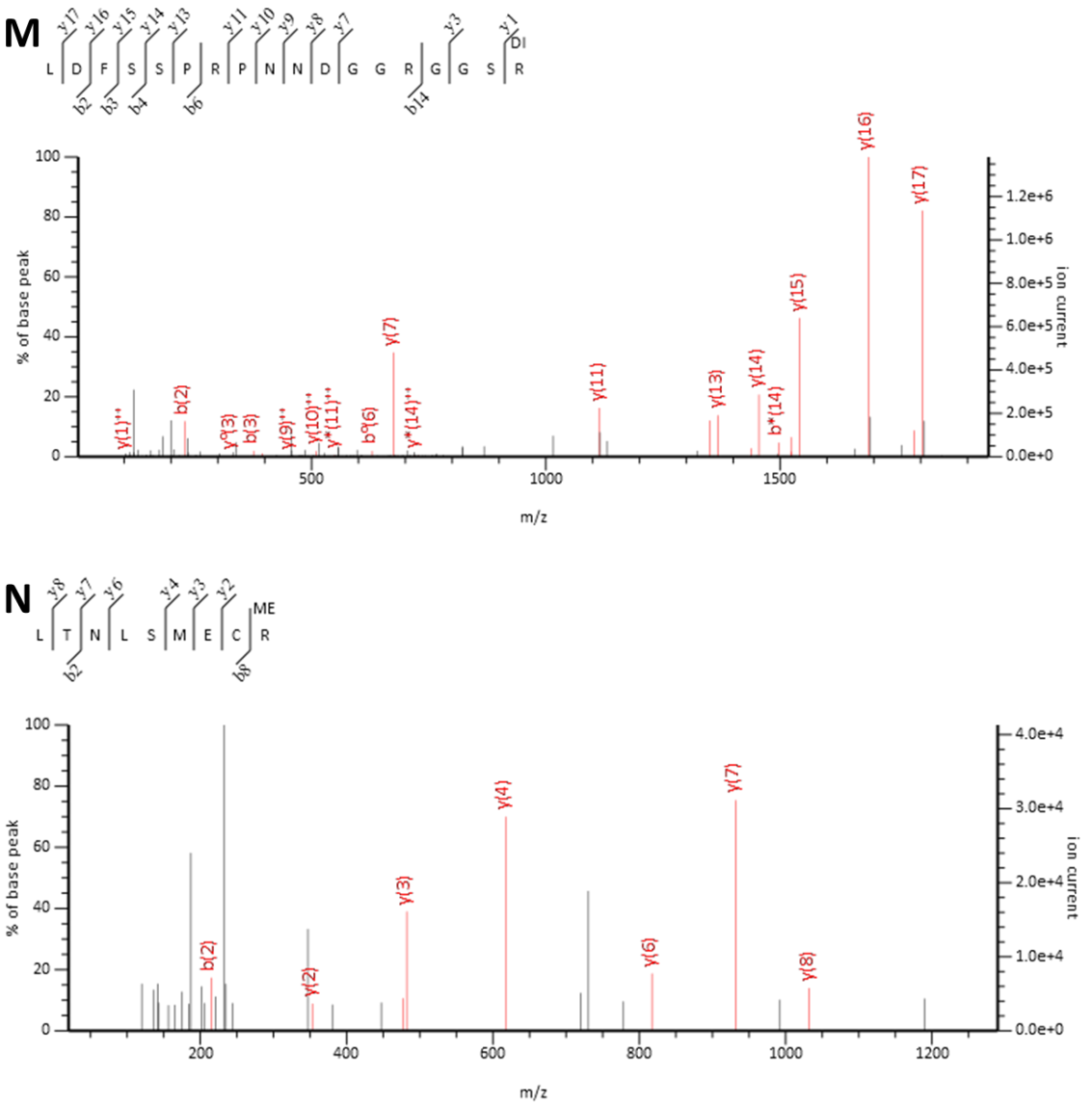
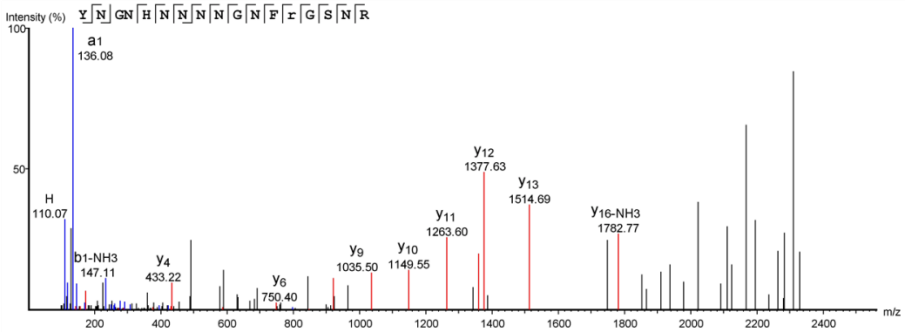
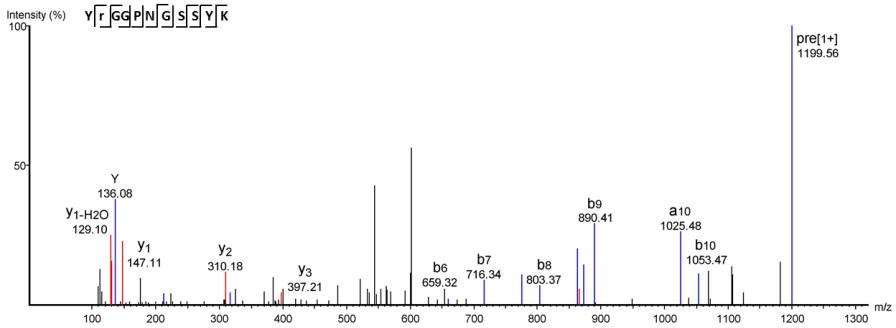



***Figure A2.1, Fragment spectra of methylated peptides identified via HILIC fractionation.*** *ME: mono-methyl, DI: di-methyl, TR: tri-methyl;* ***A****,* $_{312}$*NVSVK(me1)EIR*$_{319}$*, EF1A;* ***B****,* $_{384}$*KLEDHPK(me1)*$_{390}$*, EF1A;* ***C****,* $_{310}$*DDFK(me2)AR*$_{615}$*, EF1A;* ***D****,* $_{218}$*VLPELQGK(me1)*$_{225}$*, G3P1/G3P2/G3P3;* ***E****,* $_{74}$*EIAQDFK(me1)TDLR*$_{84}$*, H3;* ***F****,* $_{74}$*EIAQDFK(me2)TDLR*$_{84}$*, H3;* ***G****,* $_{74}$*EIAQDFK(me3)TDLR*$_{84}$*, H3;* ***H****,* $_{512}$*VATTGEWDK(me1)*$_{520}$*, Pdc1;* ***I****,* $_{33}$*ASLFAQGK(me1)R*$_{41}$*, Rpl42A/B;* ***J****,* $_{47}$*QSGFGGQTK(me1)PVFHK*$_{60}$*, Rpl42A/B;* ***K****,* $_{207}$*K(me2)SDFTLYHTWPVR*$_{219}$*, VatA;* ***L****,* $_{641}$*ESSPLNDSR(me1)*$_{649}$*, Lte1;* ***M****,* $_{340}$*LDFSSPRPNNDGGRGGSR(me2)*$_{357}$*, Nsr1;* ***N****,* $_{167}$*LTNLSMECR(me1)*$_{175}$*, Num1.*

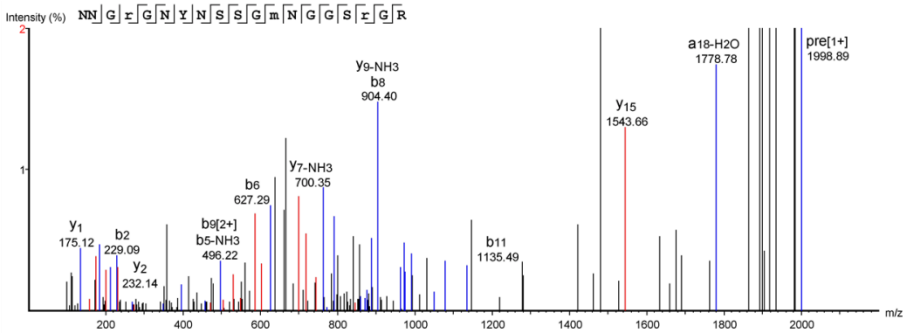
A


Intensity (%)
Y N GN H N N N N G N F r G S N R
a1 136.08
H 110.07
b1-NH3 147.11
y4 433.22
y6 750.40
y9 1035.50
y10 1149.55
y11 1263.60
y12 1377.63
y13 1514.69
y16-NH3 1782.77
m/z

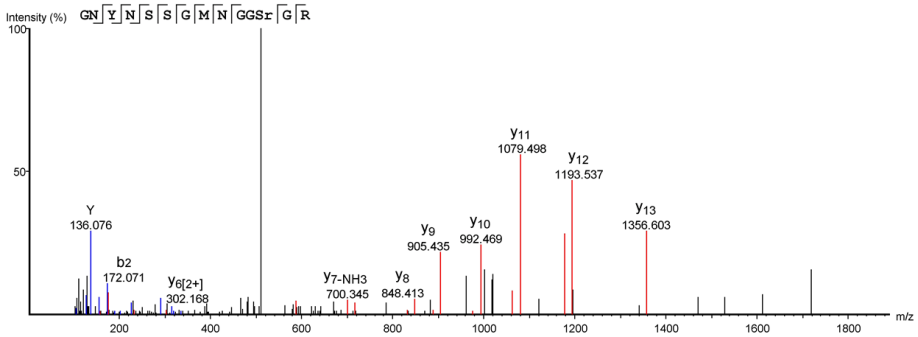
B


Intensity (%)
Y r GG P N G S S Y K
pre[1+] 1199.56
Y 136.08
y1-H2O 129.10
y1 147.11
y2 310.18
y3 397.21
b6 659.32
b7 716.34
b8 803.37
b9 890.41
a10 1025.48
b10 1053.47
m/z

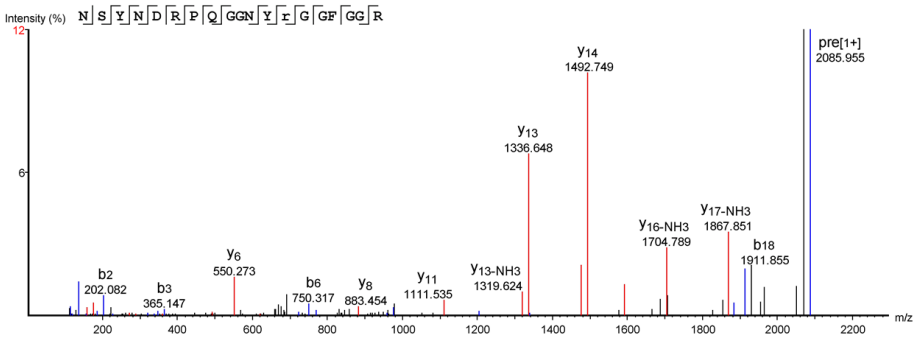
C


Intensity (%)
NN G r G N Y N S S G m N G G S r G R
a18-H2O 1778.78
pre[1+] 1998.89
y9-NH3 b8 904.40
y15 1543.66
y7-NH3 700.35
b6 627.29
y1 175.12
b2 229.09
y2 232.14
b9[2+] b5-NH3 496.22
b11 1135.49
m/z

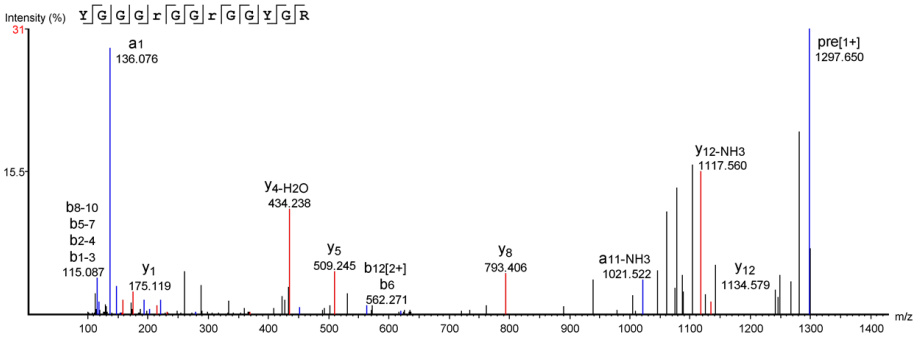
D


Intensity (%)
GN Y N S S G M N GGSr G R
y11 1079.498
y12 1193.537
Y 136.076
y13 1356.603
y9 905.435
y10 992.469
b2 172.071
y6[2+] 302.168
y7-NH3 700.345
y8 848.413
m/z

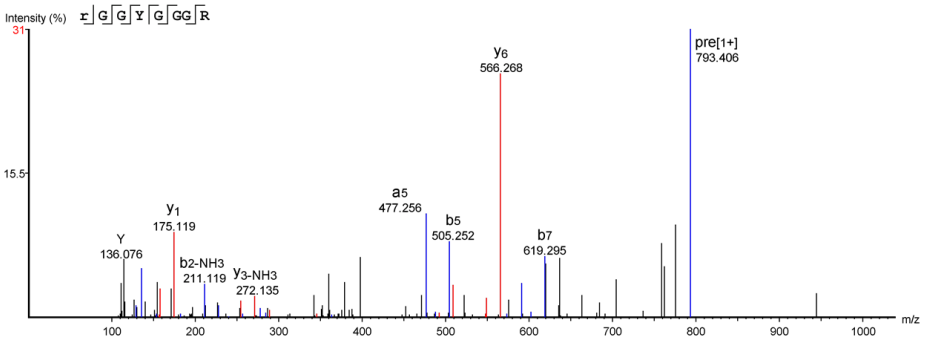
E


Intensity (%)
N S Y N D R P Q GGN Y r G GF GG R
12
6
y14 1492.749
pre[1+] 2085.955
y13 1336.648
y17-NH3 1867.851
y16-NH3 1704.789
b18 1911.855
y6 550.273
b2 202.082
b3 365.147
b6 750.317
y8 883.454
y11 1111.535
y13-NH3 1319.624
m/z

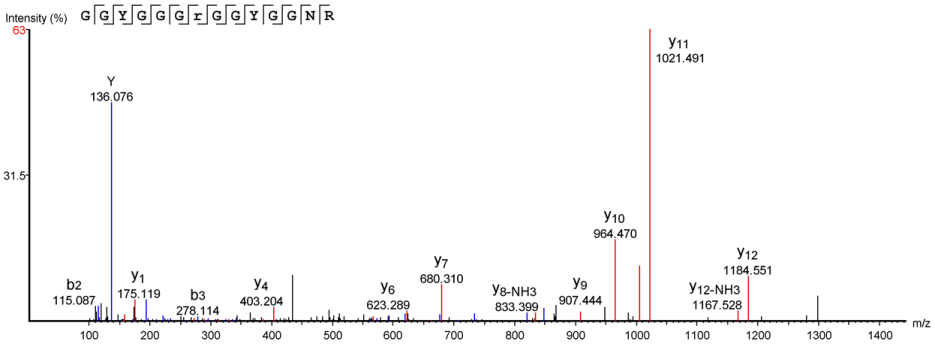
F


Intensity (%)
Y G G G r G G r G G Y G R
31
15.5
a1 136.076
pre[1+] 1297.650
y12-NH3 1117.560
y4-H2O 434.238
b8-10
b5-7
b2-4
b1-3 115.087
y1 175.119
y5 509.245
b12[2+]
b6 562.271
y8 793.406
a11-NH3 1021.522
y12 1134.579
m/z

G


Intensity (%)
r G G Y G GG R
31
15.5
y6 566.268
pre[1+] 793.406
a5 477.256
y1 175.119
b5 505.252
Y 136.076
b7 619.295
b2-NH3 211.119
y3-NH3 272.135
m/z

H


Intensity (%)
G G Y G G G r G G Y G G N R
63
31.5
y11 1021.491
Y 136.076
y10 964.470
y12 1184.551
y7 680.310
b2 115.087
y1 175.119
b3 278.114
y4 403.204
y6 623.289
y8-NH3 833.399
y9 907.444
y12-NH3 1167.528
m/z

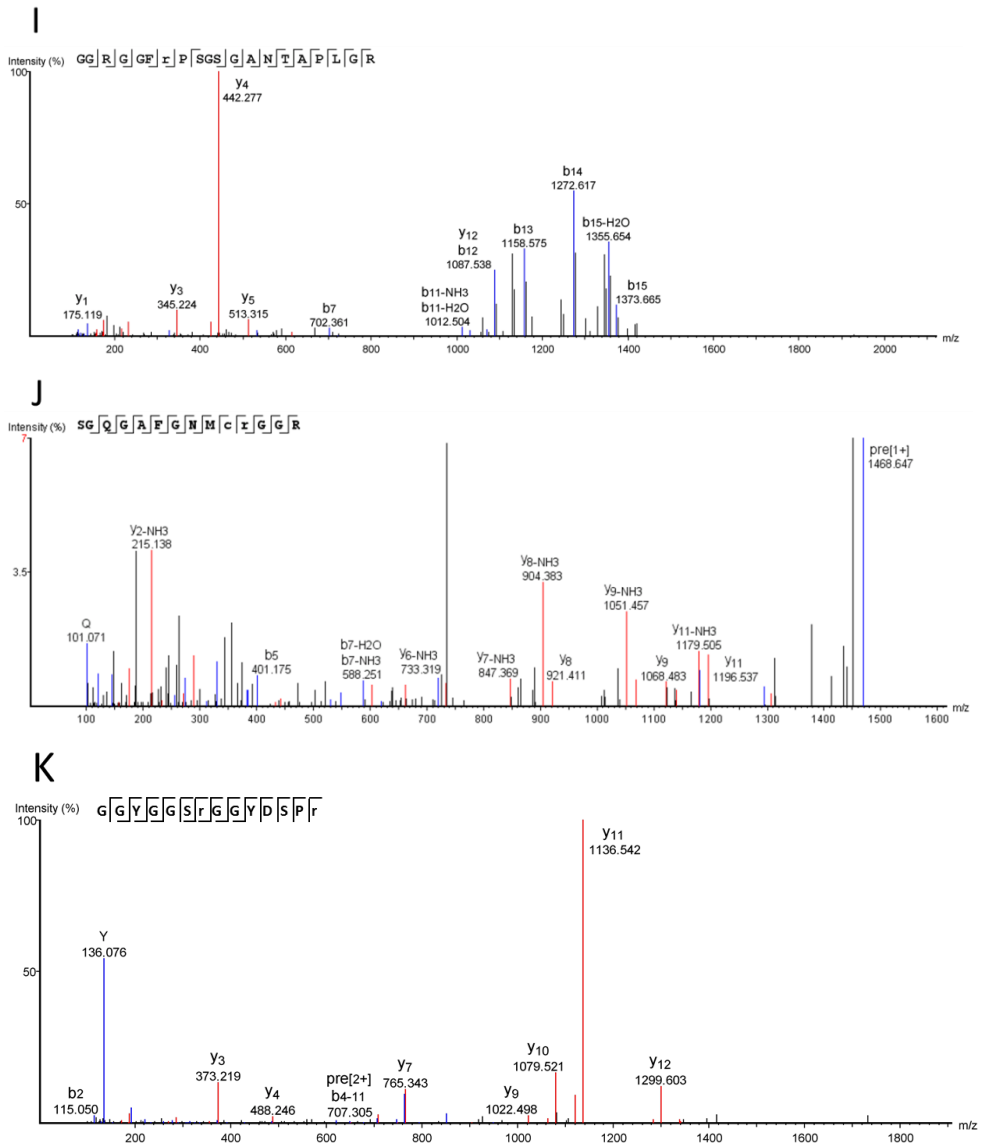



***Figure A2.2: Peptide MS/MS spectra demonstrating protein arginine methylation of methylation sites identified via peptide immunoaffinity purification and not previously reported***. *Modified residues are indicated by lowercase letters (c: carbamidomethylated cysteine, m: oxidized methionine, r: mono-methylated arginine). In D the position of the methylation site is not unambiguously localized according to the definition in the Methods section, all others are unambiguous and the methylated arginine is surrounded by fragment spectra on each side.* ***A****,* $_{407}$*YNGNHNNNNGNFR(me1)GSNR*$_{423}$*; Psp2;* ***B****,* $_{424}$*YR(me1)GGPNGSSYK*$_{434}$*; Psp2;* ***C****,* $_{535}$*NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR*$_{553}$*; Psp2;* ***D****,* $_{539}$*GNYNSSGMNGGSRGR(me1)*$_{553}$ *or* $_{539}$*GNYNSSGMNGGSR(me1)GR*$_{553}$*;* ***E****, Psp2;* $_{31}$*NSYNDRPQGGNYR(me1)GGFGGR*$_{49}$*; Dbp2;* ***F****,* $_{505}$*YGGGR(me1)GGR(me1)GGYGR*$_{517}$*; Dbp2;* ***G****,* $_{518}$*R(me1)GGYGGGR*$_{525}$*; Dbp2;* ***H****,* $_{519}$*GGYGGGR(me1)GGYGGNR*$_{532}$*;* ***I****,* $_{380}$*GGRGGFR(me1)PSGSGANTAPLGR*$_{399}$*; Nsr1;* ***J,*** $_{85}$*SGQGAFGNMC(carbam.)R(me1)GGR*$_{98}$*; Rpl4A/B; K,* $_{338}$*GGYGGSR(me1)GGYDSPR(me1)*$_{351}$*; (Npl3). (Plank et al. 2015)*

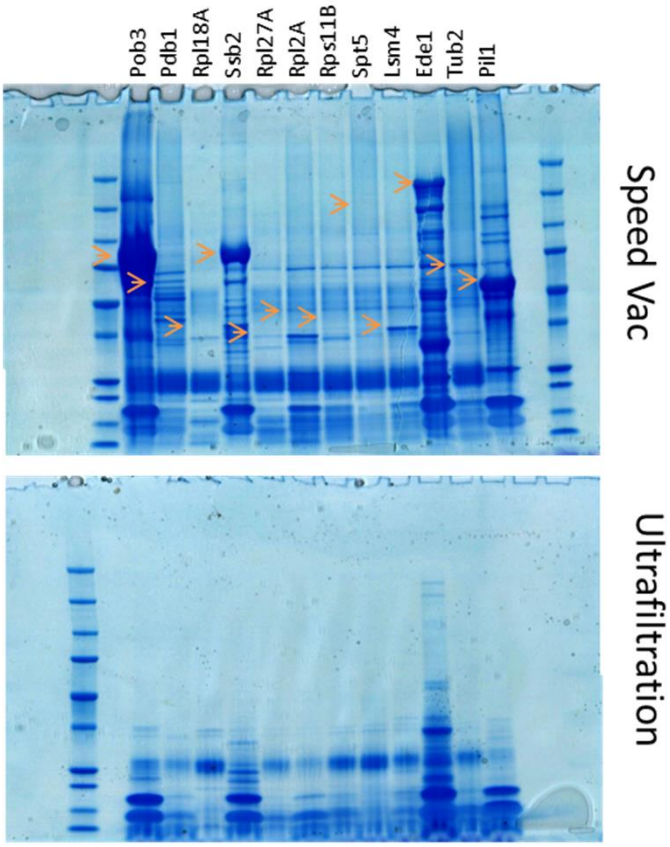



***Figure A3.1: Concentration of proteins by vacuum centrifugation and ultrafiltration methods.*** *A panel of twelve proteins of the MORF library were overexpressed in S. cerevisiae, purified by column chromatography via there ZZ-tag with elution in 1 ml 0.5 M acetic acid pH 3.4. The eluate was either dried in a vacuum centrifuge and the resulting pellet resuspended in 15 µl 8 M urea (top) or concentrated by ultrafiltration through 5 kDa molecular weight cut-off filters (bottom) before being subjected to SDS PAGE.*

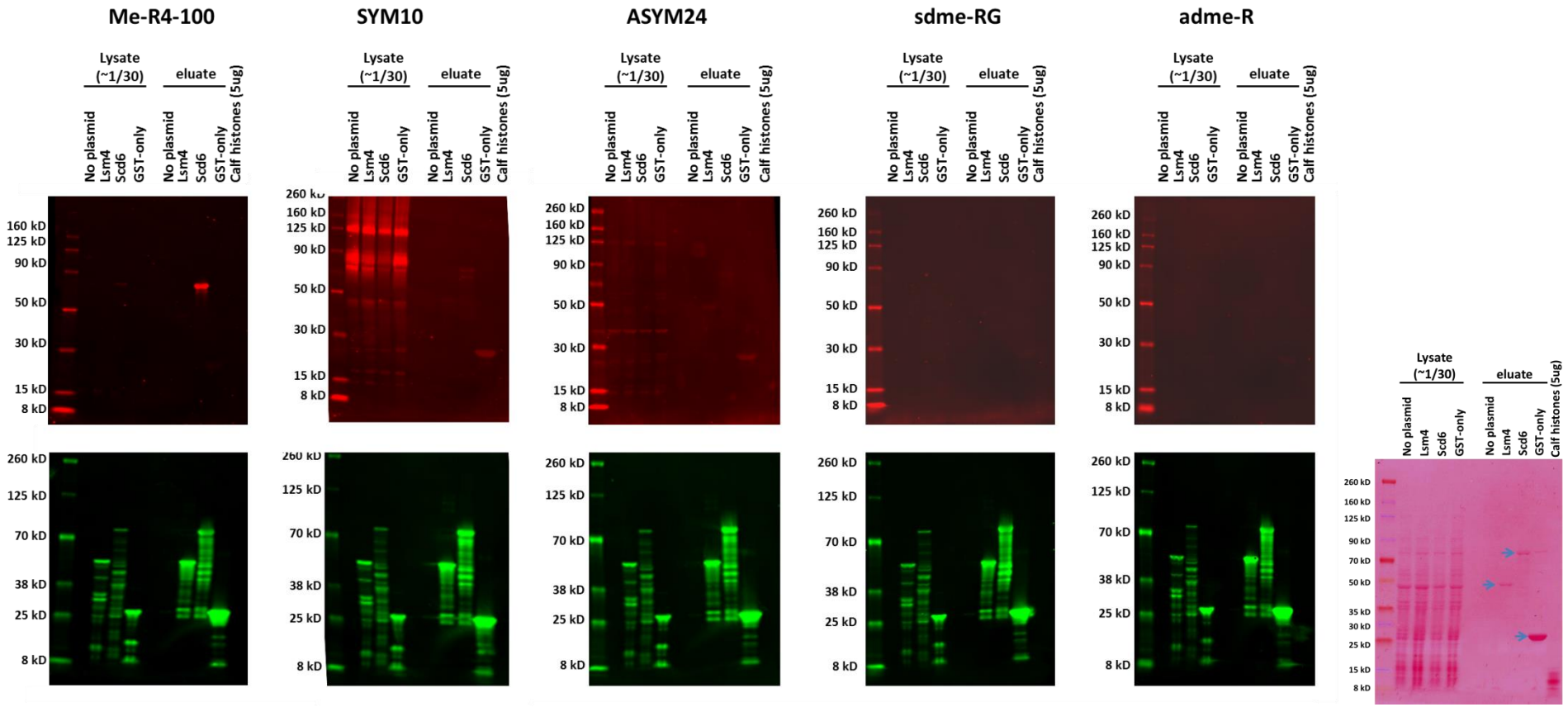



***Figure A4.1: Arginine methylation levels of purified proteins.*** *Lysates of untransformed cells and transformed with plasmids expressing GST-tagged Lsm4, Scd6 and GST-alone, eluates after purification with glutathione sepharose beads and calf histones were subjected to western blotting with antibodies Me-R4-100, SYM10, ASYM24, sdme-RG, adme-R (red) and anti-GST (green). A representative, corresponding Ponceau-stained membrane is shown on the right.*

## Appendix Tables

**Table A2.1:** Antibodies used in this study.

| **Antibody Number (Manufacturer, Name)** | **Description** | **Target / Sequence Motif** | **Reference** |
|---|---|---|---|
| 8015 (CST, Me-R4-100) | According to the manufacturer`s description, this antibody recognizes proteins that are mono-methylated on arginine without sequence preferences and does not cross-react with di-methylated or unmethylated arginine. The antibody was produced from rabbits immunized with a library of synthetic peptides containing mono-methyl arginine surrounded by degenerate amino acids and is a mixture of four monoclonal antibodies. | R(me1) | Dhar, S. et al. (2013) Sci Rep 3, 1311. Guo, A. et al. (2014) Mol Cell Proteomics 13, 372-87. |
| 8711 (CST, D5A12) | According to the manufacturer`s description, this monoclonal antibody recognizes proteins that are mono-methylated on arginine with preference for R(me1)GG motifs and does not cross-react with di-methylated or unmethylated arginine. The antibody was produced from rabbits immunized with a library of synthetic mono-methyl arginine peptides containing R(me1)GG motifs. | R(me1)GG | Dhar, S. et al. (2013) Sci Rep 3, 1311. Guo, A. et al. (2014) Mol Cell Proteomics 13, 372-87. |
| 13222 (CST, sdme-RG) | According to the manufacturer`s description, this antibody recognizes proteins that are symmetrically di-methylated on arginine and does not cross-react with mono- or asymmetrically di-methylated arginine or methylated lysine. The antibody was produced from rabbits immunized with a synthetic peptide library containing the symmetric dimethyl-arginine [sdme-RG] motif and is a mixture of two monoclonal antibodies. | R(me2sym)G | Dhar, S. et al. (2013) *Sci Rep* 3, 1311. |
| 13522 (CST, adme-R) | According to the manufacturer`s description, this antibody recognizes proteins that are asymmetrically di- | R(me2as) | Dhar, S. et al. (2013) *Sci Rep* 3, |

| | | | |
|---|---|---|---|
| | methylated on arginine and does not cross-react with mono- or symmetrically di-methylated arginine or methylated lysine. The antibody was produced from rabbits immunized with a synthetic peptide library containing the asymmetric dimethyl-arginine [adme-R] motif and is a mixture of four mono-clonal antibodies. | | 1311. Guo, A. et al. (2014) *Mol Cell Proteomics* 13, 372-87. |
| SYM10 (Millipore) | In the original publication, the antibody did not recognize non-methylated or asymmetrically di-methylated peptides. It did not recognize peptides with a single RG-repeat. It reacted more strongly with peptides in which RG(me2sym) were non-contiguous. The antibody was produced by immunizing rabbits with a KLH-coupled nonapeptide containing four sDMA-glycine repeats. It is commercially available as a rabbit polyclonal antibody. | RG(me2sym)(..) RG(me2sym) | Boisvert, F.M. et al. (2002) *J. Cell Biol.*, 159: 957-69 |
| ASYM24 (Millipore) | In the original publication, the antibody showed no reactivity against the unmethylated or symmetrically di-methylated version of the immunogenic peptide. The antibody only recognized a subset of Rme2 containing peptides. KLH-coupled peptide KGRGRGRGRGPPPPPRGRGRGRG with all arginines asymmetrically di-methylated was used as an immunogen. It is commercially available as a rabbit polyclonal antibody. | R(me2as)G with selectivity to sequence context | Côté, J et al. (2003) *Mol. Biol. Cell*, 14: 274-87 |
| ICP0801 (Immunechem) | According to manufacturer`s description this antibody was produced by immunisation with R(me1) directly coupled to KLH. | R(me1) | |
| ICP0810 (Immunechem) | According to manufacturer`s description this antibody was produced by immunisation with R(me2as) directly coupled to KLH and does not cross-react with acetylated or mono-methylated proteins. | R(me2as) | |
| AP09328PU-N(Acris) | According to manufacturer`s description this antibody was produced by immunisation with methylated lysine directly coupled to KLH. It is available as a rabbit monoclonal antibody. | Kme1, Kme2, Kme3 | |

**Appendix Table B:** GO_fat terms enriched among proteins found methylated in this study and previously known methylated proteins (Plank et al. 2015).

| Category | Term | Count | % | P-Value | Genes | List Total | Pop Hits | Pop Total | Fold Enrichment | Bonferroni | Benjamini | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GOTERM_MF_FAT | GO:0003723~RNA binding | 21 | 63.64 | 2.7E-12 | SBP1, GAR1, HRP1, RPL4A, HRB1, NSR1, RPS2, LHP1, RPS3, DED1, RPL12A, NOP1, YRA1, GUS1, NAB2, SNP1, DBP2, LSM4, PAB1, GBP2, NPL3 | 31 | 504 | 4190 | 5.6 | 2.1E-10 | 2.1E-10 | 2.8E-09 |
| GOTERM_BP_FAT | GO:0006396~RNA processing | 16 | 48.48 | 1.1E-07 | ELP4, HRP1, GAR1, NSR1, AIR2, RPS2, LHP1, DED1, NOP1, NAB2, SNP1, DBP2, LSM4, PAB1, NPL3, THO2 | 32 | 515 | 4870 | 4.7 | 3.5E-05 | 3.5E-05 | 1.5E-04 |
| GOTERM_CC_FAT | GO:0030529~ribonucleoprotein complex | 16 | 48.48 | 1.4E-07 | SBP1, HRP1, RPL4A, GAR1, HRB1, RPS2, LHP1, RPS3, STM1, RPL12A, NOP1, SNP1, LSM4, PAB1, NPL3, GBP2 | 28 | 590 | 4595 | 4.5 | 1.6E-05 | 1.6E-05 | 1.6E-04 |
| GOTERM_BP_FAT | GO:0050657~nucleic acid transport | 9 | 27.27 | 3.1E-07 | YRA1, NAB2, HRB1, PAB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 112 | 4870 | 12.2 | 9.6E-05 | 4.8E-05 | 4.2E-04 |
| GOTERM_BP_FAT | GO:0051236~establishment of RNA localization | 9 | 27.27 | 3.1E-07 | YRA1, NAB2, HRB1, PAB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 112 | 4870 | 12.2 | 9.6E-05 | 4.8E-05 | 4.2E-04 |
| GOTERM_BP_FAT | GO:0050658~RNA transport | 9 | 27.27 | 3.1E-07 | YRA1, NAB2, HRB1, PAB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 112 | 4870 | 12.2 | 9.6E-05 | 4.8E-05 | 4.2E-04 |
| GOTERM_BP_FAT | GO:0006403~RNA localization | 9 | 27.27 | 1.0E-06 | YRA1, NAB2, HRB1, PAB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 131 | 4870 | 10.5 | 3.2E-04 | 1.1E-04 | 1.4E-03 |
| GOTERM_BP_FAT | GO:0015931~nucleobase, nucleoside, nucleotide and nucleic acid transport | 9 | 27.27 | 1.2E-06 | YRA1, NAB2, HRB1, PAB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 134 | 4870 | 10.2 | 3.8E-04 | 9.5E-05 | 1.7E-03 |
| GOTERM_BP_FAT | GO:0006405~RNA export from nucleus | 8 | 24.24 | 1.3E-06 | YRA1, NAB2, HRB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 93 | 4870 | 13.1 | 4.1E-04 | 8.3E-05 | 1.8E-03 |
| GOTERM_BP_FAT | GO:0010608~posttranscriptional regulation of gene expression | 10 | 30.30 | 1.8E-06 | RPL12A, SBP1, RPL4A, GUS1, NAB2, DBP2, PAB1, RPS2, IMD4, STM1 | 32 | 192 | 4870 | 7.9 | 5.6E-04 | 9.3E-05 | 2.4E-03 |
| GOTERM_CC_FAT | GO:0043228~non-membrane-bounded organelle | 19 | 57.58 | 3.8E-06 | SBP1, GAR1, RPL4A, NSR1, AIR2, RPS2, CDC11, LHP1, STM1, RPS3, RPL12A, POB3, NOP1, LSM4, PAB1, SNF2, NPL3, GBP2, CSE4 | 28 | 1120 | 4595 | 2.8 | 4.4E-04 | 2.2E-04 | 4.3E-03 |
| GOTERM_CC_FAT | GO:0043232~intracellular non-membrane-bounded organelle | 19 | 57.58 | 3.8E-06 | SBP1, GAR1, RPL4A, NSR1, AIR2, RPS2, CDC11, LHP1, STM1, RPS3, RPL12A, POB3, NOP1, LSM4, PAB1, SNF2, NPL3, GBP2, CSE4 | 28 | 1120 | 4595 | 2.8 | 4.4E-04 | 2.2E-04 | 4.3E-03 |
| GOTERM_BP_FAT | GO:0051168~nuclear export | 8 | 24.24 | 9.3E-06 | YRA1, NAB2, HRB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 124 | 4870 | 9.8 | 2.9E-03 | 4.1E-04 | 1.2E-02 |
| GOTERM_BP_FAT | GO:0051028~mRNA transport | 7 | 21.21 | 1.0E-05 | YRA1, NAB2, HRB1, PAB1, NPL3, GBP2, THO2 | 32 | 82 | 4870 | 13.0 | 3.1E-03 | 3.9E-04 | 1.3E-02 |
| GOTERM_BP_FAT | GO:0006417~regulation of translation | 9 | 27.27 | 1.1E-05 | RPL12A, SBP1, RPL4A, GUS1, DBP2, PAB1, RPS2, IMD4, STM1 | 32 | 180 | 4870 | 7.6 | 3.5E-03 | 3.9E-04 | 1.5E-02 |
| GOTERM_CC_FAT | GO:0031981~nuclear lumen | 13 | 39.39 | 2.2E-05 | SBP1, HRP1, GAR1, NSR1, AIR2, RPS2, LHP1, POB3, NOP1, LSM4, SNF2, NPL3, THO2 | 28 | 545 | 4595 | 3.9 | 2.5E-03 | 8.5E-04 | 2.5E-02 |
| GOTERM_BP_FAT | GO:0032268~regulation of cellular protein metabolic process | 9 | 27.27 | 2.6E-05 | RPL12A, SBP1, RPL4A, GUS1, DBP2, PAB1, RPS2, IMD4, STM1 | 32 | 201 | 4870 | 6.8 | 7.8E-03 | 7.9E-04 | 3.4E-02 |
| GOTERM_BP_FAT | GO:0006913~nucleocytoplasmic transport | 8 | 24.24 | 3.0E-05 | YRA1, NAB2, HRB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 148 | 4870 | 8.2 | 9.1E-03 | 8.3E-04 | 4.0E-02 |
| GOTERM_BP_FAT | GO:0051169~nuclear transport | 8 | 24.24 | 3.0E-05 | YRA1, NAB2, HRB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 148 | 4870 | 8.2 | 9.1E-03 | 8.3E-04 | 4.0E-02 |
| GOTERM_MF_FAT | GO:0003727~single-stranded RNA binding | 4 | 12.12 | 5.3E-05 | NAB2, PAB1, NPL3, DED1 | 31 | 11 | 4190 | 49.1 | 4.2E-03 | 2.1E-03 | 5.6E-02 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0006406~mRNA export from nucleus | 6 | 18.18 | 5.5E-05 | YRA1, NAB2, HRB1, NPL3, GBP2, THO2 | 32 | 67 | 4870 | 13.6 | 1.7E-02 | 1.4E-03 | 7.3E-02 |
| GOTERM_BP_FAT | GO:0034660~ncRNA metabolic process | 11 | 33.33 | 9.9E-05 | ELP4, NOP1, GAR1, GUS1, DBP2, LSM4, NSR1, AIR2, RPS2, NPL3, LHP1 | 32 | 393 | 4870 | 4.3 | 3.0E-02 | 2.3E-03 | 1.3E-01 |
| GOTERM_CC_FAT | GO:0005730~nucleolus | 9 | 27.27 | 1.0E-04 | NOP1, SBP1, GAR1, LSM4, NSR1, AIR2, RPS2, NPL3, LHP1 | 28 | 269 | 4595 | 5.5 | 1.2E-02 | 3.0E-03 | 1.2E-01 |
| GOTERM_BP_FAT | GO:0034470~ncRNA processing | 10 | 30.30 | 1.6E-04 | ELP4, NOP1, GAR1, DBP2, LSM4, NSR1, AIR2, RPS2, NPL3, LHP1 | 32 | 335 | 4870 | 4.5 | 4.9E-02 | 3.5E-03 | 2.2E-01 |
| GOTERM_BP_FAT | GO:0016071~mRNA metabolic process | 9 | 27.27 | 2.1E-04 | HRP1, NAB2, SNP1, DBP2, LSM4, AIR2, PAB1, NPL3, THO2 | 32 | 270 | 4870 | 5.1 | 6.2E-02 | 4.3E-03 | 2.8E-01 |
| GOTERM_MF_FAT | GO:0003729~mRNA binding | 5 | 15.15 | 2.8E-04 | HRP1, NAB2, SNP1, PAB1, NPL3 | 31 | 46 | 4190 | 14.7 | 2.2E-02 | 7.5E-03 | 3.0E-01 |
| GOTERM_MF_FAT | GO:0008143~poly(A) RNA binding | 3 | 9.09 | 7.3E-04 | NAB2, PAB1, NPL3 | 31 | 6 | 4190 | 67.6 | 5.7E-02 | 1.5E-02 | 7.7E-01 |
| GOTERM_MF_FAT | GO:0070717~poly-purine tract binding | 3 | 9.09 | 7.3E-04 | NAB2, PAB1, NPL3 | 31 | 6 | 4190 | 67.6 | 5.7E-02 | 1.5E-02 | 7.7E-01 |
| GOTERM_BP_FAT | GO:0016072~rRNA metabolic process | 8 | 24.24 | 7.5E-04 | NOP1, GAR1, DBP2, LSM4, NSR1, AIR2, RPS2, NPL3 | 32 | 248 | 4870 | 4.9 | 2.1E-01 | 1.4E-02 | 1.0E+00 |
| GOTERM_CC_FAT | GO:0043233~organelle lumen | 13 | 39.39 | 7.9E-04 | SBP1, HRP1, GAR1, NSR1, AIR2, RPS2, LHP1, POB3, NOP1, LSM4, SNF2, NPL3, THO2 | 28 | 783 | 4595 | 2.7 | 8.7E-02 | 1.8E-02 | 8.9E-01 |
| GOTERM_CC_FAT | GO:0070013~intracellular organelle lumen | 13 | 39.39 | 7.9E-04 | SBP1, HRP1, GAR1, NSR1, AIR2, RPS2, LHP1, POB3, NOP1, LSM4, SNF2, NPL3, THO2 | 28 | 783 | 4595 | 2.7 | 8.7E-02 | 1.8E-02 | 8.9E-01 |
| GOTERM_BP_FAT | GO:0031123~RNA 3'-end processing | 5 | 15.15 | 1.1E-03 | NOP1, HRP1, NAB2, AIR2, THO2 | 32 | 74 | 4870 | 10.3 | 3.0E-01 | 2.0E-02 | 1.5E+00 |
| GOTERM_BP_FAT | GO:0042254~ribosome biogenesis | 9 | 27.27 | 1.2E-03 | RPL12A, NOP1, GAR1, DBP2, LSM4, NSR1, RPS2, NPL3, RPS3 | 32 | 351 | 4870 | 3.9 | 3.1E-01 | 2.1E-02 | 1.6E+00 |
| GOTERM_CC_FAT | GO:0031974~membrane-enclosed lumen | 13 | 39.39 | 1.3E-03 | SBP1, HRP1, GAR1, NSR1, AIR2, RPS2, LHP1, POB3, NOP1, LSM4, SNF2, NPL3, THO2 | 28 | 827 | 4595 | 2.6 | 1.4E-01 | 2.5E-02 | 1.5E+00 |
| GOTERM_BP_FAT | GO:0016973~poly(A)+ mRNA export from nucleus | 3 | 9.09 | 1.4E-03 | NAB2, HRB1, GBP2 | 32 | 9 | 4870 | 50.7 | 3.5E-01 | 2.2E-02 | 1.8E+00 |
| GOTERM_BP_FAT | GO:0006397~mRNA processing | 7 | 21.21 | 1.5E-03 | HRP1, NAB2, SNP1, LSM4, PAB1, NPL3, THO2 | 32 | 204 | 4870 | 5.2 | 3.8E-01 | 2.3E-02 | 2.0E+00 |
| GOTERM_BP_FAT | GO:0022613~ribonucleoprotein complex biogenesis | 9 | 27.27 | 2.7E-03 | RPL12A, NOP1, GAR1, DBP2, LSM4, NSR1, RPS2, NPL3, RPS3 | 32 | 397 | 4870 | 3.5 | 5.6E-01 | 3.9E-02 | 3.5E+00 |
| GOTERM_BP_FAT | GO:0006364~rRNA processing | 7 | 21.21 | 3.4E-03 | NOP1, GAR1, DBP2, LSM4, NSR1, RPS2, NPL3 | 32 | 239 | 4870 | 4.5 | 6.5E-01 | 4.7E-02 | 4.5E+00 |
| GOTERM_CC_FAT | GO:0010494~stress granule | 3 | 9.09 | 3.8E-03 | HRP1, PAB1, GBP2 | 28 | 16 | 4595 | 30.8 | 3.5E-01 | 6.1E-02 | 4.2E+00 |
| GOTERM_MF_FAT | GO:0000166~nucleotide binding | 14 | 42.42 | 5.6E-03 | SBP1, HRP1, HRB1, NSR1, CDC11, DED1, LHP1, YRA1, GUS1, SNP1, DBP2, PAB1, SNF2, GBP2 | 31 | 897 | 4190 | 2.1 | 3.6E-01 | 8.6E-02 | 5.8E+00 |
| GOTERM_CC_FAT | GO:0005840~ribosome | 7 | 21.21 | 8.7E-03 | RPL12A, NOP1, RPL4A, PAB1, RPS2, STM1, RPS3 | 28 | 316 | 4595 | 3.6 | 6.3E-01 | 1.2E-01 | 9.5E+00 |
| GOTERM_CC_FAT | GO:0022626~cytosolic ribosome | 5 | 15.15 | 9.5E-03 | RPL12A, RPL4A, RPS2, STM1, RPS3 | 28 | 145 | 4595 | 5.7 | 6.7E-01 | 1.1E-01 | 1.0E+01 |
| GOTERM_BP_FAT | GO:0043631~RNA polyadenylation | 3 | 9.09 | 1.1E-02 | HRP1, NAB2, AIR2 | 32 | 25 | 4870 | 18.3 | 9.6E-01 | 1.4E-01 | 1.3E+01 |
| GOTERM_CC_FAT | GO:0005732~small nucleolar ribonucleoprotein complex | 3 | 9.09 | 1.3E-02 | NOP1, GAR1, LSM4 | 28 | 30 | 4595 | 16.4 | 7.8E-01 | 1.4E-01 | 1.4E+01 |
| GOTERM_BP_FAT | GO:0000956~nuclear-transcribed mRNA catabolic process | 4 | 12.12 | 1.6E-02 | HRP1, DBP2, LSM4, AIR2 | 32 | 84 | 4870 | 7.2 | 9.9E-01 | 1.8E-01 | 1.9E+01 |
| GOTERM_BP_FAT | GO:0006402~mRNA catabolic process | 4 | 12.12 | 1.7E-02 | HRP1, DBP2, LSM4, AIR2 | 32 | 86 | 4870 | 7.1 | 9.9E-01 | 1.9E-01 | 2.0E+01 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0006399~tRNA metabolic process | 5 | 15.15 | 2.2E-02 | ELP4, GUS1, LSM4, AIR2, LHP1 | 32 | 170 | 4870 | 4.5 | 1.0E+00 | 2.3E-01 | 2.5E+01 |
| GOTERM_BP_FAT | GO:0006401~RNA catabolic process | 4 | 12.12 | 2.5E-02 | HRP1, DBP2, LSM4, AIR2 | 32 | 100 | 4870 | 6.1 | 1.0E+00 | 2.5E-01 | 2.9E+01 |
| GOTERM_BP_FAT | GO:0000377~RNA splicing, via transesterification reactions with bulged adenosine as nucleophile | 4 | 12.12 | 2.6E-02 | SNP1, LSM4, NPL3, DED1 | 32 | 102 | 4870 | 6.0 | 1.0E+00 | 2.5E-01 | 3.0E+01 |
| GOTERM_BP_FAT | GO:0031124~mRNA 3'-end processing | 3 | 9.09 | 2.6E-02 | HRP1, NAB2, THO2 | 32 | 40 | 4870 | 11.4 | 1.0E+00 | 2.5E-01 | 3.0E+01 |
| GOTERM_BP_FAT | GO:0000375~RNA splicing, via transesterification reactions | 4 | 12.12 | 3.1E-02 | SNP1, LSM4, NPL3, DED1 | 32 | 109 | 4870 | 5.6 | 1.0E+00 | 2.8E-01 | 3.5E+01 |
| GOTERM_BP_FAT | GO:0043555~regulation of translation in response to stress | 2 | 6.06 | 3.1E-02 | SBP1, STM1 | 32 | 5 | 4870 | 60.9 | 1.0E+00 | 2.7E-01 | 3.5E+01 |
| GOTERM_CC_FAT | GO:0000346~transcription export complex | 2 | 6.06 | 4.0E-02 | YRA1, THO2 | 28 | 7 | 4595 | 46.9 | 9.9E-01 | 3.5E-01 | 3.8E+01 |
| GOTERM_BP_FAT | GO:0034621~cellular macromolecular complex subunit organization | 7 | 21.21 | 5.1E-02 | RPL12A, POB3, LSM4, NSR1, SNF2, NPL3, CSE4 | 32 | 431 | 4870 | 2.5 | 1.0E+00 | 4.0E-01 | 5.0E+01 |
| GOTERM_MF_FAT | GO:0042802~identical protein binding | 8 | 24.24 | 5.2E-02 | ELP4, GAR1, NAB2, DBP2, UGP1, LSM4, IMD4, CDC11 | 31 | 487 | 4190 | 2.2 | 9.9E-01 | 5.1E-01 | 4.3E+01 |
| GOTERM_BP_FAT | GO:0034728~nucleosome organization | 3 | 9.09 | 5.5E-02 | POB3, SNF2, CSE4 | 32 | 60 | 4870 | 7.6 | 1.0E+00 | 4.1E-01 | 5.3E+01 |
| GOTERM_BP_FAT | GO:0008380~RNA splicing | 4 | 12.12 | 6.0E-02 | SNP1, LSM4, NPL3, DED1 | 32 | 142 | 4870 | 4.3 | 1.0E+00 | 4.3E-01 | 5.7E+01 |
| GOTERM_BP_FAT | GO:0006259~DNA metabolic process | 7 | 21.21 | 6.5E-02 | POB3, LGE1, YRA1, SNF2, GBP2, STM1, THO2 | 32 | 458 | 4870 | 2.3 | 1.0E+00 | 4.5E-01 | 6.0E+01 |
| GOTERM_BP_FAT | GO:0006283~transcription-coupled nucleotide-excision repair | 2 | 6.06 | 6.8E-02 | YRA1, THO2 | 32 | 11 | 4870 | 27.7 | 1.0E+00 | 4.5E-01 | 6.1E+01 |
| GOTERM_BP_FAT | GO:0006446~regulation of translational initiation | 2 | 6.06 | 8.0E-02 | PAB1, STM1 | 32 | 13 | 4870 | 23.4 | 1.0E+00 | 5.0E-01 | 6.7E+01 |
| GOTERM_BP_FAT | GO:0046907~intracellular transport | 8 | 24.24 | 8.3E-02 | YRA1, NAB2, HRB1, RPS2, NPL3, GBP2, RPS3, THO2 | 32 | 608 | 4870 | 2.0 | 1.0E+00 | 5.1E-01 | 6.9E+01 |
| GOTERM_BP_FAT | GO:0051276~chromosome organization | 6 | 18.18 | 9.5E-02 | POB3, LGE1, SNF2, GBP2, STM1, CSE4 | 32 | 387 | 4870 | 2.4 | 1.0E+00 | 5.5E-01 | 7.4E+01 |
| GOTERM_MF_FAT | GO:0016944~RNA polymerase II transcription elongation factor activity | 2 | 6.06 | 9.6E-02 | POB3, NPL3 | 31 | 14 | 4190 | 19.3 | 1.0E+00 | 6.8E-01 | 6.6E+01 |
| GOTERM_CC_FAT | GO:0044445~cytosolic part | 4 | 12.12 | 9.8E-02 | RPL12A, RPL4A, RPS2, RPS3 | 28 | 190 | 4595 | 3.5 | 1.0E+00 | 6.3E-01 | 6.9E+01 |

DOI 10.1002/pmic.201500032 *Proteomics* 2015, *15*, 3232–3243

RESEARCH ARTICLE

# Expanding the yeast protein arginine methylome

*Michael Plank[1,2], Roman Fischer[1], Vincent Geoghegan[3], Philip D. Charles[1], Rebecca Konietzny[1], Oreste Acuto[3], Catherine Pears[4], Christopher J. Schofield[2] and Benedikt M. Kessler[1]*

[1] Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[2] Chemistry Research Laboratory, University of Oxford, Oxford, UK
[3] Sir William Dunn School of Pathology, University of Oxford, Oxford, UK
[4] Department of Biochemistry, University of Oxford, Oxford, UK

Protein arginine methylation is a PTM involved in various cellular processes in eukaryotes. Recent discoveries led to a vast expansion of known sites in higher organisms, indicating that this modification is more widely spread across the proteome than previously assumed. An increased knowledge of sites in lower eukaryotes may facilitate the elucidation of its functions. In this study, we present the discovery of arginine mono-methylation sites in *Saccharomyces cerevisiae* by a combination of immunoaffinity enrichment and MS/MS. As detection of methylation is prone to yield false positives, we demonstrate the need for stringent measures to avoid elevated false discovery rates. To this end, we employed MethylSILAC in combination with a multistep data analysis strategy. We report 41 unambiguous methylation sites on 13 proteins. Our results indicate that, while substantially less abundant, arginine methylation follows similar patterns as in higher eukaryotes in terms of sequence context and functions of methylated proteins. The majority of sites occur on RNA-binding proteins participating in processes from transcription and splicing to translation and RNA degradation. Additionally, our data suggest a bias for localization of arginine methylation in unstructured regions of proteins, which frequently involves Arg-Gly-Gly motifs or Asn-rich contexts.

Received: January 22, 2015
Revised: March 27, 2015
Accepted: June 2, 2015

**Keywords:**
MethylSILAC / Protein arginine methylation / Systems biology / Tandem mass spectrometry / Yeast

Additional supporting information may be found in the online version of this article at the publisher's web-site

## 1 Introduction

PTMs on proteins play important roles in regulating the catalytic activity, interactions, stability, and subcellular localization of proteins. More than 400 types of covalent modifications to proteins, including phosphorylation, acetylation, ubiquitinylation, and methylation, have been described [1]. Protein arginine methylation was discovered more than 40 years ago and can occur as the addition of a single methyl-group to one of the ω-amino groups of arginine (mono-methyl arginine; MMA) or with two methyl-groups added to the terminal amino groups, either both on the same ω-amino group (asymmetric di-methyl arginine; aDMA) or one on each (symmetric di-methyl arginine; sDMA) [2]. Additionally, mono-methylation of the δ-position (δ-MMA) has been reported in *Saccharomyces cerevisiae*, but not in mammals [3]. Methyl groups are transferred to arginine residues by protein arginine methyltransferases, which use S-adenosyl methionine as a donor.

Functionally, arginine methylation is linked to transcriptional regulation, splicing, nuclear translocation, DNA repair, and signalling, while mechanistically it may inhibit or facilitate protein–protein and protein-RNA interactions [4]. In particular, the methylation of certain arginine residues is required for the binding of tudor-domain containing proteins [5].

**Correspondence**: Prof. Benedikt M. Kessler, Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford OX3 7FZ, UK
**E-mail**: benedikt.kessler@ndm.ox.ac.uk

**Abbreviations**: **DMA**, di-methyl arginine; **MMA**, mono-methyl arginine; **PSM**, peptide spectrum match

**Colour Online**: See the article online to view Fig. 1 in colour.

## Significance of the study

Recent reports have led to a surge in the number of discovered protein arginine methylation sites in higher eukaryotes. At the same time, our understanding of the functions of these sites has not grown proportionally. Insights into the biology of arginine methylation will rely on studies in lower eukaryotes that are more readily experimentally accessible.
Here, we apply peptide immuno-affinity enrichment for the identification of arginine methylation sites in *Saccharomyces cerevisiae*. Our work underscores a significant technical point in regards to inherent challenges in the confident identification of protein methylation sites. We show how these can be addressed by MethylSILAC and provide an evaluation of the performance of this experimental strategy.
Our results indicate that, while arginine methylation is less frequent in yeast than in human cells, it is located in similar protein families and sequence contexts. Arginine methylation was found predominantly in unstructured protein regions, which frequently occurs in, but does not strictly require, a glycine-arginine-rich environment. Our data also suggest that asparagine rich regions also provide a sequence context for arginine methylation.

While the number of known protein methylation sites remained low for several decades, recent studies employing tandem MS have led to a surge in the number of discovered sites in humans and mice, bringing the total up to the thousands [6–9]. The precise biological functions, as well as the enzymatic activities responsible for establishing the modification at many of these sites, however, remain unknown. In contrast, the number of known nonhistone methylated proteins in lower eukaryotic model-organisms such as yeast, is low. For instance, 18 non-histone target proteins of arginine methyltransferases in *S. cerevisiae* have been summarized in a recent review [10] and the sites of methylation have only been defined for a subset of these proteins. In a more recent study, 90 proteins in *S. cerevisiae* were found in a pull-down with an anti-MMA antibody, on which 15 methylation sites could be mapped to five of the proteins by MS/MS [11]. Additional 21 sites (mono- or di-methyl) on five proteins were subsequently detected by the same group by coexpressing the target proteins with a methyltransferase in *E. coli* [12]. A computational re-analysis of peptide-mass finger printing spectra of *S. cerevisiae* proteins predicted 38 arginine methylation sites, only a few of which however have been verified experimentally to date [13].

While only a limited number of studies aimed at discovering methylation in *S. cerevisiae*, a better knowledge of sites in this organism would be beneficial, as due to its lower genetic redundancy and ease of manipulation, it would lend itself to studies of both the downstream function and upstream mediators of these modifications.

The identification of protein methylation by MS/MS poses a challenge, as the mass difference of 14 Da introduced by a methyl group is isobaric to several amino acid substitutions (e.g. Asp-Asn, Gly-Ala, Val-Leu/Ile). A strategy in which proteins with unlabelled and metabolically heavy-labelled methyl-groups are mixed and therefore display a characteristic heavy-light pair (in the following referred to as 'MethylSILAC') has been employed on several occasions to increase the confidence in identified methylation sites in human cells [7, 8, 14].

Here, we report the use of MethylSILAC in combination with immunoaffinity enrichment to reveal 41 arginine monomethylation sites on 13 proteins in *S. cerevisiae*, including six proteins that have not yet been described to be methylated. Our results strengthen the association between arginine methylation and RNA processing. They indicate that the previously observed prevalence of arginine methylation sites in RG-rich sequences extends to *S. cerevisiae* and suggest that they are predominantly found in unstructured regions, especially Asn-rich sequences.

## 2 Materials and methods

### 2.1 Experimental outline

An outline of the experimental workflow is shown in Fig. 1. Cells were grown in media with unlabelled or heavy methionine, proteins extracted, and digested with trypsin and methylated peptides enriched by immunoaffinity purification. Eluted peptides were analysed by MS and methylated peptides identified via their MS2-spectra and the presence of a heavy/light pair on MS1 level.

### 2.2 Yeast and mammalian cell culture and protein preparation

*S. cerevisiae* BY4741 cells were grown at 180 rpm and 30°C in SC medium (Formedium: CYN0410 and DSCK1000) with light methionine or methionine-[methyl$^{13}CD_3$] (Sigma: 299154) and 2% glucose for eight generations and harvested in log-phase. Cells were lysed by four rounds of bead-beating for 30 s at 6000 rpm in 20 mM Tris, pH 8, and 8 M urea and debris was removed by centrifugation at 20 000 *g* for 15 min. The protein concentration was determined by BCA assay (Pierce: 23 227F) and heavy and light samples mixed in a 1:1 protein ratio.

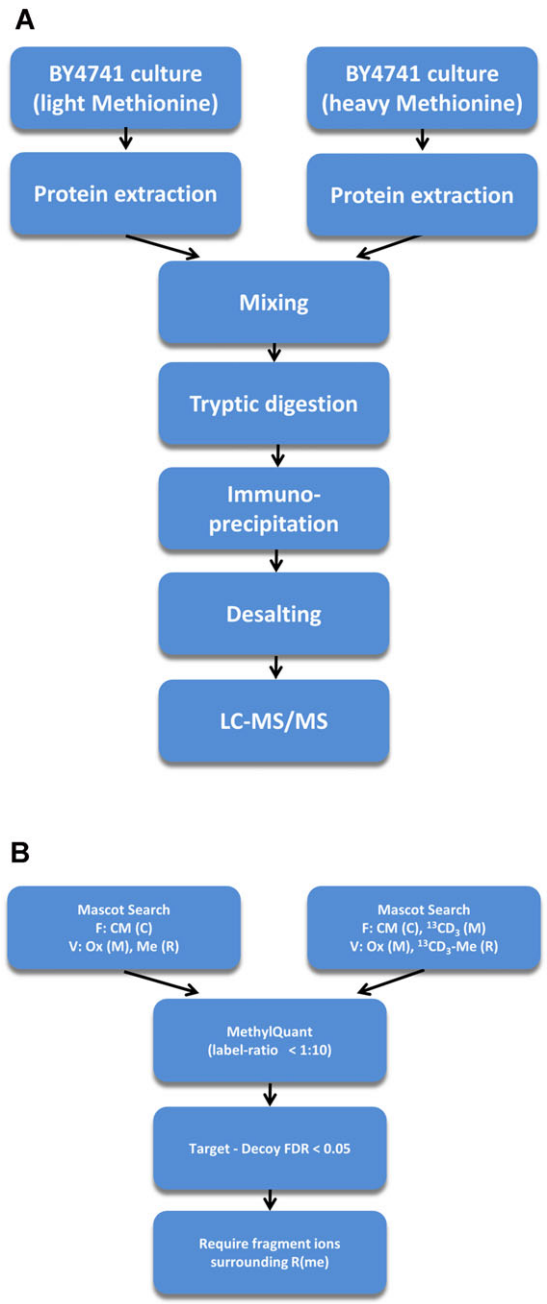



**Figure 1.** Workflow for assessing protein arginine methylation used in this study. (A) Experimental workflow and MS analysis. (B) MS data analysis workflow of MethylSILAC experiments. F, Fixed modification; V, Variable modification; CM, Carbamidomethylation; Ox, Oxidation; Me, Methylation.

The preparation of proteins from Jurkat cells as a control is described in the supplemental material.

### 2.3 Peptide immunoaffinity purification

Cleared lysate corresponding to 25 mg protein per sample from *S. cerevisiae* or Jurkat lysates was subjected to 30 min each of reduction (5 mM DTT), alkylation (15 mM iodoacetamide), and quenching (5 mM DTT) at room temperature. The urea concentration was reduced to 2 M by adding 20 mM Tris, pH 8, and $CaCl_2$ was added to 1 mM. Digestion was carried out with trypsin (Promega: V5111) at a 1:100 mass-ratio and incubation at 37°C overnight. Samples were split into batches corresponding to 12.5 mg protein and TFA added to 0.1%. After an incubation of 10 min at room temperature, insoluble material was removed by centrifugation (2000 *g*, 5 min) and samples desalted by solid phase extraction (Sep-Pak C18 Plus, Waters) with stepwise elution at 10, 20, 30, and 40% ACN in 0.1% TFA, respectively. Eluates were combined, frozen at –80°C and lyophilized. Aliquots of 80 μL protein A agarose bead slurry (Roche: 11719408001) were washed twice in immunoaffinity purification buffer (IAP: 50 mM MOPS, 10 mM $NaH_2PO_4$, 50 mM NaCl) and incubated with a mixture of 100 μL Me-R4-100 and 50 μL D5A12 anti-methyl antibodies (CST: 8015 and 8711; Supporting Information Table 1) for 3h at 4°C, adjusted to a final volume of 500 μL with IAP buffer, then washed three times in IAP buffer. The peptide aliquots equivalent to 12.5 mg protein were resuspended in 1.3 mL of IAP buffer, cleared at 20 000 g for 5 min, added to the antibody-bound beads and rotated overnight at 4°C. The beads were washed three times in IAP buffer, once in water and peptides eluted in 100 μL 0.15% TFA. Peptides were purified by passing through a stage tip with a glass microfiber filter (Whatman: 1820-042) followed by desalting on a C18 filter (Supelco: 66883-U). The latter was washed with 0.1% TFA and peptides eluted in 50 μL 50% acetonitrile, 0.1% TFA. Samples corresponding to 25 mg protein starting material were combined, dried in a Speed Vac, resuspended in 20 μL 2% acetonitrile, 0.1% formic acid, and subjected to LC-MS/MS analysis. The experiment was performed in three technical replicates for *S. cerevisiae* samples by splitting cultures before protein extraction.

The immunoaffinity purification using a mix of both antibodies had been previously employed successfully on human cells [15]. As RGG-motif containing peptides occurred with high frequency in our results, the experiment was repeated with 100 μL Me-R4-100 alone.

### 2.4 Mass spectrometry analysis

Peptides were analysed by LC-MS/MS in data-dependent acquisition mode as described in the Supplementary Methods.

### 2.5 MS data analysis

The raw files obtained from MS analysis were converted to .mgf format using MSConvert [16], keeping the 200 most intense peaks. After confirming that the three replicate samples were similar in terms of the number of both total and methylated peptides identified at a given FDR in an initial search, files were merged and searched with Mascot

(Matrix Science) against the UniProt database limited to entries from *S. cerevisiae* with common contaminants added, specifying tryptic digestion, up to three missed cleavages, 10 ppm precursor- and 0.07 Da fragment-tolerance, selecting b- and y- (inclusive water and ammonium loss), a- and immonium ions for scoring and enabling decoy search. Searches were performed separately with parameters for light (fixed modification: Carbamidomethyl; variable modifications: Oxidation (M), Monomethyl (R)) and heavy labelled (fixed modification: Carbamidomethyl, Label:$^{13}C^2H_3$ (M); variable modifications: Oxidation (M), $^{13}C^2H_3$ Monomethyl (R)) peptides.

The results from the two parameter sets were reformatted to be compatible with the novel software MethylQuant (a description of the algorithm is provided in the Supplementary Methods) merged into one file and analysed with MethylQuant_0_90_64bit, which calculates the peak intensities of both partners of a heavy-light pair. Methylated peptides with H/L ratio of $>0.1$ and $<10$ were accepted. The FDR was estimated by 2*#decoy PSMs/(#target PSMs + #decoy PSMs). Results were filtered at a Mascot *E*-value corresponding to a FDR $<0.05$ and we further required the presence of ions arising from fragmentation directly next to the methylated residue (one if the methylated arginine was C-terminal). Additionally, if the localization percentage of the Mascot site analysis for the first hit was less than three times the value of the following hit, the site assignment is reported as ambiguous.

A separate analysis was performed, based on a Mascot search additionally allowing Dimethyl (R) and $^{13}C_2{}^2H_6$ Dimethyl (R) among the light and heavy variable modifications, respectively. Downstream data analyses on methylated peptides are described in the Supplementary Methods.

## 3 Results

### 3.1 Proteomic approach toward protein arginine methylation discovery

We applied a MS/MS approach to comprehensively survey protein arginine methylation. The experimental workflow included a labelling strategy leading to stable isotope incorporation into methyl groups (MethylSILAC) (Fig. 1). MS-based identifications were required for at least one feature of the heavy/light SILAC pair and then filtered based on a heavy to light peak intensity ratio of $>0.1$ and $<10$ (Fig. 2A, B). The corresponding experiment was performed on peptides from Jurkat cell lysates as a positive control (Fig. 2C, D; Supporting Information Fig. 12). As shown in Fig. 2E, the FDR calculated only on peptide spectrum matches (PSMs) to peptides containing a methyl group is consistently higher than on all PSMs at the same E-value cut-off after the Mascot search. We next asked if the use of the MethylSILAC strategy reduces the number of false positives compared to an experiment without MethylSILAC, and indeed the FDR at a given *E*-value after applying MethylQuant is consistently lower than after the database search alone (Fig. 2F). This indicates an increase in specificity, but might come at a loss of sensitivity. However, the number of PSMs at a given FDR is consistently higher with than without MethylSILAC (Fig. 2G), which confirms that the overall sensitivity at a given FDR is increased when using the MethylSILAC strategy.

### 3.2 Identification of protein arginine methylation sites

After filtering based on the MS1 intensity ratio and MS2 spectral matching, 62 matches to peptides of the *S. cerevisiae* proteome remained. The list comprises 41 unambiguous methylation sites, several of which are either novel or provide evidence for mono-methylation of known di-methylation sites (Table 1, Supporting Information Table 2) [11,12,17–19]. The sites were mapped to 13 proteins, for six of which methylation had not been reported previously to our knowledge, while the remaining seven provide confirmation of previously described methylated proteins. Fragment spectra for each novel site are shown in Fig. 3 and Supporting Information S1–S11.

We unambiguously mapped three to five methylation sites each to Nop1, Nsr1, Sbp1, Dbp2, Ded1, and Psp2 and 12 to Npl3. Several of these sites are covered by more than one peptide and often two methylation sites are found on the same peptide as the sites are in close proximity. An additional methylated peptide matches to Nop1 in three different positions (R16, R34, R52), making it an ambiguous assignment. A further MS/MS spectrum for Psp2 is ambiguous in localizing a methylation to one of the five unambiguous sites (R551) or to an additional site (R553) (Table 1). In the case of Nab2, we could localize one methylation site to R222, while the second methylation in this peptide might be located on R226 or R229. Co-fragmentation of a mixture of peptides with methyl-groups localized at different residues is also possible. One methylation site each was detected on Gar1, Rpl4, Elp4, Lge1 and Lsm4. When searching for di- additionally to mono-methylation, 14 peptides indicated the occurrence of di-methylation at sites reported above and at two additional sites on Psp2 (R443, R447) (Supporting Information Table 3). However, in many cases, the confidence of site localization is low, potentially due to co-fragmentation of peptides with different distributions of methyl-groups. A repetition of the immunoprecipitation with antibody Me-R4-100 (CST: 8711) alone resulted in the discovery of one additional site on protein Pdc1 (R161) (Supporting Information Table 5).

### 3.3 GO of methylated proteins

We used DAVID to elucidate the GO for which arginine methylated proteins are enriched [20]. This was performed separately for the arginine methylation sites identified in this study and for all currently known sites including the ones presented here. The results are shown in Supporting

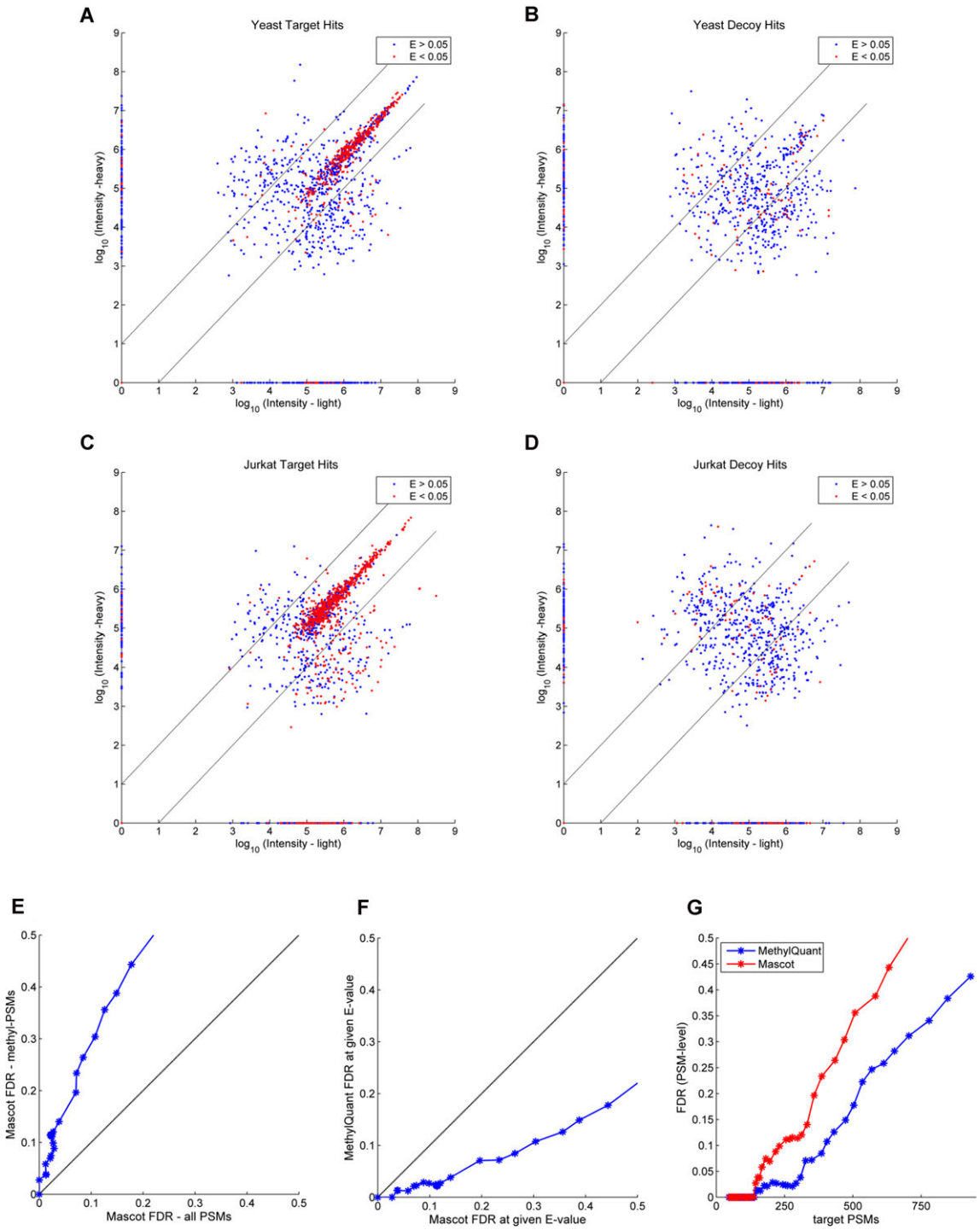



**Figure 2.** Confidence assessment of protein arginine methylation mapping. (A–D) Light versus heavy intensities (log10 transformed; peaks with an intensity of 0 are replaced by 0 on the log-scale) of feature pairs identified as arginine methylated by Mascot search. Identifications with a Mascot *E*-value $<0.05$ are shown in red and $>0.05$ in blue. An arbitrary upper threshold of $E<0.3$ was imposed in the second case. Feature pairs were detected using MethylQuant. The heavy to light ratio thresholds of 0.1 and 10 applied to these data are shown as black lines. Data obtained for *S. cerevisiae* (A, B) and Jurkat cells (C, D) are shown for Mascot target (A, C) and decoy (B, D) hits. (E) FDRs at different *E*-value thresholds after Mascot search for all peptide spectrum matches (PSMs) versus matches to methylated peptides. (F) FDR after Mascot search versus FDR at the same *E*-value cutoffs after processing with MethylQuant and filtering to a heavy-to-light ratio of $>0.1$ and $<10$. (G) FDR at PSM level at different numbers of identified target PSMs for processing with Mascot alone (red) or via the MethylQuant strategy (blue). Plots E–G are based on *S. cerevisiae* samples.

**Table 1.** Protein arginine methylation sites in yeast identified by MS/MS

| Accession | Protein name | Prot. prev. known | Methylated residue | Methyl. site prev. known | Study | Modified peptide |
|---|---|---|---|---|---|---|
| P50109.2 | Psp2 | n | 419 | | | YNGNHNNNNGNFR(me1)GSNR |
| | | | 425 | | | YR(me1)GGPNGSSYK |
| | | | 575 | | | GGR(me1)GSSGNYSNYNNR |
| | | | 538 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR |
| | | | 551 | | | NNGR(me1)GNYNSSGM(ox)NGGSR(me1)GR |
| | | | | | | NNGR(me1)GNYNSSGMNGGSR(me1)GR |
| | | | 551 or 553 | | | GNYNSSGMNGGSRGR(me1) or GNYNSSGMNGGSR(me1)GR |
| P24783.1 | Dbp2 | n | 43 | | | NSYNDRPQGGNYR(me1)GGFGGR |
| | | | 509 | | | YGGGR(me1)GGR(me1)GGYGR |
| | | | 512 | | | YGGGR(me1)GGR(me1)GGYGR |
| | | | 518 | | | R(me1)GGYGGGR |
| | | | 525 | | | GGYGGGR(me1)GGYGGNR |
| P06634.2 | Ded1 | y | 51 | y(me2) | Erce, 2013 | NNSSNYNNNNGGYNGGR(me1)GGGSFFSNNR |
| | | | 62 | y(me2) | Erce, 2013 | R(me1)GGYGNGGFFGGNNGGSR |
| | | | 578 | y | Erce, 2013 | DNSFR(me1)GGSGWGSDSK |
| P15646.1 | Nop1 | y | 70 | y(me1 and me2) | Low, 2013 | GGSR(me1)GGR(me1)GGAAGGAR |
| | | | 73 | y (me1 and me2) | Low, 2013 | GGR(me1)GGAAGGAR |
| | | | | | | GGSR(me1)GGR(me1)GGAAGGAR |
| | | | | | | GGR(me1)GGAAGGAR(me1) |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK |
| | | | 81 | y(me2) | Low, 2013 | GGAAGGAR(me1)GGAK |
| | | | | | | GGR(me1)GGAAGGAR(me1) |
| | | | | | | GGR(me1)GGAAGGAR(me1)GGAK |
| | | | 16 or 34 or 52[a)] | all three: y(me2as - UniProt by Similarity) | | GGSR(me1)GGFGGR |
| P28007.1 | Gar1 | y | 147 | y(me2as- UniProt probable) | | SGAPGGR(me1)GGASMGR |
| P27476.1 | Nsr1 | y | 353 | y(me2as- UniProt probable) | | LDFSSPRPNNDGGR(me1) |
| | | | 379 | y(me2as- UniProt probable) | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |
| | | | 382 | y(me2as- UniProt probable) | | GGR(me1)GGFRPSGSGANTAPLGR |
| | | | | | | GGAR(me1)GGR(me1)GGFRPSGSGANTAPLGR |
| | | | | | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |
| | | | 386 | n | | GGRGGFR(me1)PSGSGANTAPLGR |
| | | | | | | GGAR(me1)GGR(me1)GGFR(me1)PSGSGANTAPLGR |
| P49626.2 / P10664.4 | Rpl4B/Rpl4A | n | 95 | | | SGQGAFGNMCR(me1)GGR |
| Q01560.1 | Nop3 / Npl3 | y | 307 | y(me1 and me2) | Hart-Smith, 2012 | GGFSR(me1)GGFGGPR |

**Table 1.** Continued

| Accession | Protein name | Prot. prev. known | Methylated residue | Methyl. site prev. known | Study | Modified peptide |
|---|---|---|---|---|---|---|
| | | | | | | GGFSR(me1)GGFGGPR(me1) |
| | | | | | | GGFSR(me1)GGFGGPR(me1)GGFGGPR |
| | | | 314 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGFGGPR |
| | | | | | | GGFSR(me1)GGFGGPR(me1) |
| | | | | | | GGFSR(me1)GGFGGPR(me1)GGFGGPR |
| | | | 321 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) |
| | | | 329 | y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1)GGYGGYSR(me1) |
| | | | 337 | y(me1 and me2) | Hart-Smith, 2012 | GGYGGYSR(me1)GGYGGSR |
| | | | 344 | n | | GGYGGSR(me1)GGYDSPR |
| | | | | | | GGYGGSR(me1)GGYDSPR(me1) |
| | | | 351 | y(me1) | Hart-Smith, 2012 | GGYDSPR(me1)GGYDSPR |
| | | | | | | GGYGGSR(me1)GGYDSPR(me1) |
| | | | | | | GGYDSPR(me1)GGYDSPR(me1) |
| | | | 358 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1)GGYSR |
| | | | | | | GGYDSPR(me1)GGYDSPR(me1) |
| | | | | | | GGYDSPR(me1)GGYSR(me1) |
| | | | | | | GGYDSPR(me1)GGYSR(me1)GGYGGPR[b] |
| | | | 363 | y(me1 and me2) | Hart-Smith, 2012 | GGYDSPRGGYSR(me1) |
| | | | | | | GGYSR(me1)GGYGGPR |
| | | | | | | GGYDSPR(me1)GGYSR(me1) |
| | | | | | | GGYDSPR(me1)GGYSR(me1)GGYGGPR[b] |
| | | | 377 | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR |
| | | | 384 | y(me1 and me2) | Hart-Smith, 2012 | GSYGGSR(me1)GGYDGPR |
| | | | | | | GSYGGSR(me1)GGYDGPRGDYGPPR |
| | | | 391 | y (me2) | Hart-Smith, 2012 | GGYDGPR(me1)GDYGPPR |
| | | | 314 or 321[a] | both: y(me1 and me2) | Hart-Smith, 2012 | GGFGGPR(me1) |
| | | | 351 or 358[a] | 351: y(me1); 358: y(me1 and me2) | Hart-Smith, 2012 | GGYDSPR(me1) |
| | | | 377 or 384[a] | y(me1 and me2) | Hart-Smith, 2012 | NDYGPPR(me1)GSYGGSR or NDYGPPRGSYGGSR(me1) |
| P32505.1 | Nab2 | y | 222 | (region: 180–256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) |
| | | | 226 or 229 | (region: 180–256) | Green, 2002 | GGR(me1)GGNR(me1)GGR or GGR(me1)GGNRGGR(me1) |

**Table 1.** Continued

| Accession | Protein name | Prot. prev. known | Methylated residue | Methyl. site prev. known | Study | Modified peptide |
|---|---|---|---|---|---|---|
| P10080.2 | Ssbp1 / Sbp1 | y | 125 | y(me1 and me2) | Hsieh, 2007[c)] | TPGΩM(ox)ΩR(me1) |
| | | | | | | TPGΩMΩR(me1) |
| | | | 145 | y(me2) | Hsieh, 2007[c)] | GGFR(me1)GGYR |
| | | | 165 | y(me2) | Hsieh, 2007[c)] | GGAR(me1)GGFNGΩK |
| | | | | | | GGAR(me1)GGFNGΩKR |
| | | | | | | GR(me1)GGAR(me1)GGFNGΩKR |
| P40070.1 | Lsm4 | n | 119 | | | DSNNNR(me1)GNYNR |
| Q02796.1 | Lge1 | n | 39 | | | GGAGGSYYR(me1)GGNASYGAR |
| Q02884.1 | Elp4 | n | 13 | | | GEILNDR(me1)GSGLR |
| P06169.1 | Pdc1 | n | 161 | | | TTYVTQR(me1)PVYLGLPANLVDLNVPAK[d)] |

a) The identified peptide matches to more than one location in the protein.
b) Only one of the sites was surrounded by fragment ions.
c) Sites of methylation are not given in this study, but could be partially derived from the peptides and number of methyl groups; these sites were found by expressing Sbp1 and Hmt1 in *E. coli*.
d) This identification resulted from a separate IP with antibody Me-R4-100 only.

Information Table 4A and B, respectively. The GO terms most significantly (Benjamini-Hochberg corrected $p < 0.05$) enriched in both cases are 'RNA binding' and 'RNA processing'. 'RNA transport', 'RNA localization' and 'RNA export from nucleus' are among the top categories for the list derived from all known methylation sites, while both lists emphasize RNA metabolism of both mRNA and rRNA. The cellular compartments that showed most significant enrichment of methylated proteins were 'nucleolus' for the sites reported in this study and 'ribonucleoprotein complex' for the complete list of methylated proteins. These findings are in agreement with previous reports describing the association between arginine methylation and RNA-related functions [4, 21].

### 3.4 Sequence context of arginine methylation sites

To determine whether arginine methylation in *S. cerevisiae* is associated with a defined sequence motif, we evaluated the residue frequency surrounding the methylation sites. For the sites identified in this study, a clear increase in the frequency of glycine residues is perceptible, especially at the positions directly following the methylated arginine. The sequence context is similar when adding previously identified sites (including di-methyl arginines) (Supporting Information Fig. 14). Indeed, 31 of the 41 sites are in an RGG context, while nine are followed by a single glycine. R386 in Nsr1 forms an exception and is followed by proline. A spectrum formed by the combination of a CID- and ETD-based fragmentation on the same precursor resulted in complete sequence coverage of the corresponding peptide and therefore further disproved the possibility of a miss-localization of the methylation (Supporting Information Fig. 13).

To ensure that the strong prevalence of RGG-motif peptides in our results was not solely explained by the inclusion of an RGG-biased antibody in the pull-down mix, we also separately performed the experiment with an antibody raised against arginine methylated peptides without sequence bias. In this experiment, eight sites were in a RGG, two in a RG and one in a RP context (Supporting Information Table 5).

The sites on several proteins containing multiple methyl groups (Nop1, Nsr1, Npl3, Dbp2, Ded1, and Psp2) cluster in specific regions (Fig. 4). This clustering together with the strong enrichment of methylation on proteins involved in functions related to RNA metabolism led us to ask whether the methylation sites themselves are frequently found in specific protein motifs or domains, especially RNA-binding domains. For ten of the 13 proteins at least a subset of their methylation sites is located in a RGG-box (defined here as at least two RGG sequences, each separated by a maximum of nine amino acids). Additionally, in PSP2, all five identified sites are localized within (R419, R425, R538) or in close vicinity (R551, R575) to the region designated as Asn-rich in UniProt. Notably, only one of these sites is in an RGG context, while in all other cases the arginine is followed by only one glycine. In Ded1 two sites (R61, R62) and the only

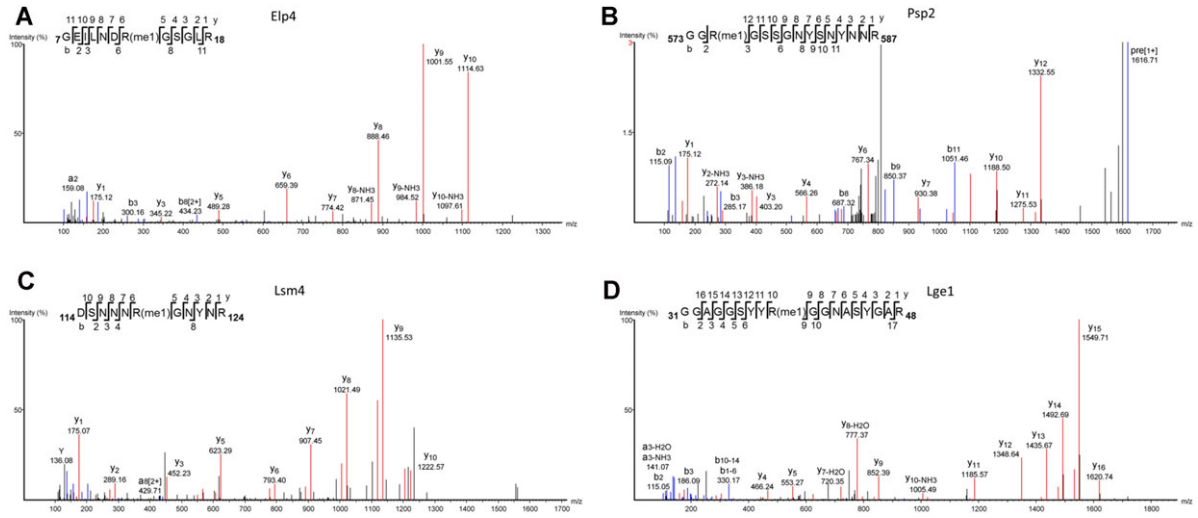



**Figure 3.** Peptide MS/MS spectra demonstrating protein arginine methylation. MS/MS spectrum of peptide 7GEILNDR(me1)GSGLR18 derived from protein Elp4 (A), 573GGR(me1)GSSGNYSNYNNR587 from Psp2 (B), 114DSNNNR(me1)GNYNR124 from Lsm4 (C) and 31GGAGGSYYR(me1)GGNASYGAR48 from Lge1 (D).

site identified in Lsm4 (R119) are also located in an Asn-rich region (Fig. 4).

As no overlap between methylation sites and annotated, structured protein domains could be discovered, we compared the localization of the sites with secondary structure predictions and found that, with the exception of R95 in Rpl4, all sites are located within extended unstructured regions (no secondary structure information was available for Lge1 and Elp4) (Fig. 4).

## 4 Discussion

### 4.1 Challenges in methyl PTM assignment and the necessity for MethylSILAC

Methylation is a PTM that poses a particular challenge to discovery by MS/MS as it is isobaric to several amino acid substitutions. Here, we employed the MethylSILAC strategy described previously, to use the observation of a heavy-light pair as additional evidence for the presence of a methylation [14]. By comparing the FDR obtained with or without this approach at different *E*-value cut-offs we show that this strategy increases the confidence of identification. This in turn allows the *E*-value cut-off to be relaxed and consequently renders the assay more sensitive. To our knowledge, this strategy had not yet been applied in *S. cerevisiae*, nor had enrichment of methylation sites on the peptide level.

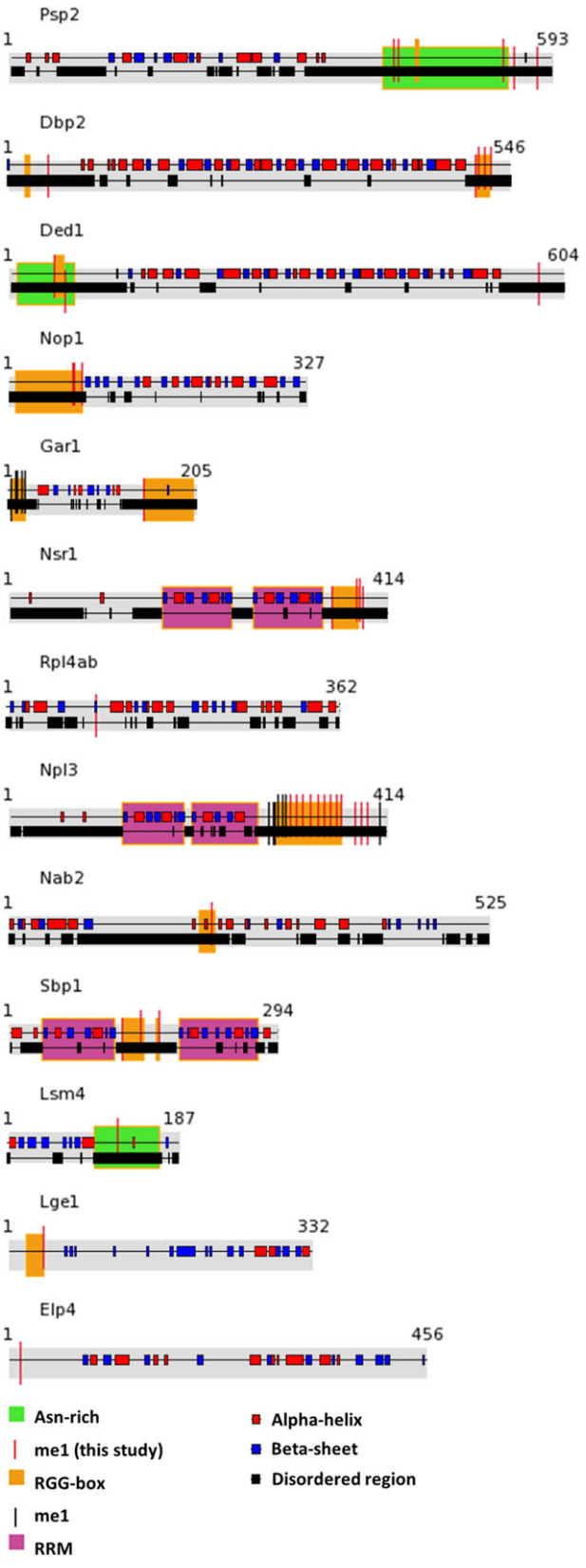
### 4.2 Methylation sites are located in RGG-boxes and unstructured regions

We identified 41 mono-methylation sites on 13 proteins below a FDR of 0.05 from spectra containing fragment ions directly surrounding the methyl-arginine (Table 1). The same experiment was performed with antibodies against di-methylated arginines (symmetric and asymmetric; CST: #13222 and #13522), but no methylation sites could be discovered with the same criteria. The higher success of the peptide immunoaffinity purification approach for mono- as compared to di-methylated peptides is consistent with a previous study and may potentially be due to a higher quality of available antibodies targeting the former modification [9]. While the fact that not all previously known sites were found in this study indicates the restriction of the antibody based approach, our data provide a considerable increase in the number of unambiguous mono-methylation sites.

Our results are in good agreement with previous findings since we confirmed several previously known methylation sites and also found most methylation sites in a RGG-context. For ten of the identified proteins, Psp2, Lge1, Nop1, Gar1, Nsr1, Nab2, Sbp1, Dbp2, Ded1 and Npl3, some or all of the methylation sites are within a RGG-box. RGG-boxes are a highly conserved feature [22] and have been found in several RNA-binding proteins [23]. Interactions of the RGG box with DNA and RNA have been described [24, 25]. The functional importance of arginine methylation of an RGG-box has however been most clearly demonstrated in the context of protein–protein interaction: Sky1, a kinase involved in regulating proteins involved in mRNA transport, phosphorylates Npl3 [26], which in turn promotes mRNA dissociation and re-import of Npl3 into the nucleus after RNA export [27]. The RGG domain of Npl3 binds to Sky1 in unmethylated, but not in methylated form [28].

A pronounced preference for localization of sites within unstructured regions suggests that this may be a general requirement for methylation to occur. Absence of secondary structure elements is consistent with the prevalence of




glycines in the vicinity of methylated arginines and the finding that the only site we found not to be followed by glycine is instead followed by proline. It is also consistent with the well-established occurrence of arginine methylation in histone tails [29]. Additionally, sequences with strong Asn-bias, which were observed for three of the methylated proteins, may be a further way to ensure an unfolded nature. These findings suggest that methylation sites are recognized by their primary structure rather than their fold by the corresponding methyltransferase or co-factor.

### 4.3 Arginine methylated proteins perform RNA-related functions

Several of the proteins found to be methylated here take part in mRNA processing at different points in its life-time: Epl4 is linked to transcription as part of the elongator complex. The RNA-helicase Dbp2 is involved in mRNP formation [12], which includes the binding of export factors such as Npl3 and Nab2, to which methylation sites could be mapped. Ded1, a RNA helicase plays a role in translation initiation [31] and Rpl4 forms part of the ribosome. Finally, Sbp1 is involved in RNA decapping while, depending on its interaction partners, Lsm4 can either function in splicing or mRNA degradation. Three of the methylated proteins, Nop1, Gar1 and Nsr1 are linked to rRNA processing. Lge1 plays a role in transcriptional activation, but its mechanism of action and whether it interacts with DNA or RNA is unclear [35]. Arginine methylation has been previously observed to influence RNA binding [36].

We anticipate that the detection and localization of methylation sites in a simple eukaryote will help understand the role of methylation of their corresponding homologues in higher eukaryotes.

◀ **Figure 4.** Schematic view of domain organization and secondary structures of arginine methylated proteins identified. RGG boxes, Asn-rich motifs, and RRM domains are shown as boxes as indicated in the legend, and unambiguous arginine monomethylation sites as vertical lines in red when found in this study and black when described previously (see also Table 1). An RGG box was defined as at least two RGG sequences, each separated by a maximum of nine amino acids. In the foreground secondary structure elements are displayed on the top row with alpha-helices as red and beta sheets as blue boxes. Disordered regions are shown as black boxes on the bottom row. This information is not available for Lge1 and Elp4. Annotated methylation sites from this and previous studies are at positions: Psp2: 419, 425, 538, 551, 575; Dbp2: 43, 509, 512, 518, 525; Ded1: 51, 62, 578; Nop1: 70, 73, 81; Gar1: 4, 8, 11, 15, 19, 147; Nsr1: 353, 379, 382, 386; Rpl4ab: 95; Npl3: 284, 288, 290, 294, 298, 302, 307, 314, 321, 329, 337, 344, 351, 358, 363, 377, 384, 391, 404; Nab2: 222; Sbp1: 125, 145, 165; Lsm4: 119; Lge1: 39; Elp4: 13. The protein domains were generated using the drawing tool available at http://domaindraw.imb.uq.edu.au/ and secondary structure depictions are from http://www.yeastrc.org/.

*We thank the members of the Mellor lab and S. Mohammed for helpful advice, and M. Bentley and K. di Gleria for their support. M.P. is funded by an Engineering and Physical Sciences Research Council (EPSRC) DTC studentship. B.M.K. is supported by grants from CRUK and the Biomedical Research Centre (NIHR), Oxford, UK. R.F. and B.M.K. are supported by a Kennedy Research Institute Trust Fund.*

*The authors have declared no conflict of interest.*

## 5 References

[1] Magrane, M., Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. *Database J. Biol. Databases Curation* 2011, *2011*, bar009.

[2] Paik, W. K., Kim, S., Protein methylase I. Purification and properties of the enzyme. *J. Biol. Chem.* 1968, *243*, 2108–2114.

[3] Zobel-Thropp, P., Yang, M. C., Machado, L., Clarke, S., A novel post-translational modification of yeast elongation factor 1A. Methylesterification at the C terminus. *J. Biol. Chem.* 2000, *275*, 37150–37158.

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

[5] Côté, J., Richard, S., Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.* 2005, *280*, 28476–28483.

[6] Boisvert, F.-M., Côté, J., Boulanger, M.-C., Richard, S., A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics MCP* 2003, *2*, 1319–1330.

[7] Uhlmann, T., Geoghegan, V. L., Thomas, B., Ridlova, G. et al., A method for large-scale identification of protein arginine methylation. *Mol. Cell. Proteomics MCP* 2012, *11*, 1489–1499.

[8] Bremang, M., Cuomo, A., Agresta, A. M., Stugiewicz, M. et al., Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol. Biosyst.* 2013, *9*, 2231–2247.

[9] Guo, A., Gu, H., Zhou, J., Mulhern, D. et al., Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol. Cell. Proteomics MCP* 2014, *13*, 372–387.

[10] Low, J. K. K., Wilkins, M. R., Protein arginine methylation in *Saccharomyces cerevisiae*. *FEBS J.* 2012, *279*, 4423–4443.

[11] Low, J. K. K., Hart-Smith, G., Erce, M. A., Wilkins, M. R., Analysis of the proteome of *Saccharomyces cerevisiae* for Methylarginine. *J. Proteome Res.* 2013, *12*, 3884–3899.

[12] Erce, M. A., Abeygunawardena, D., Low, J. K. K., Hart-Smith, G., Wilkins, M. R., Interactions affected by arginine methylation in the yeast protein-protein interaction network. *Mol. Cell. Proteomics MCP* 2013, *12*, 3184–3198.

[13] Pang, C. N. I., Gasteiger, E., Wilkins, M. R., Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. *BMC Genomics* 2010, *11*, 92.

[14] Ong, S.-E., Mittler, G., Mann, M., Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* 2004, *1*, 119–126.

[15] Sylvestersen, K. B., Horn, H., Jungmichel, S., Jensen, L. J., Nielsen, M. L., Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. *Mol. Cell. Proteomics MCP* 2014, *13*, 2072–2088.

[16] Chambers, M. C., Maclean, B., Burke, R., Amodei, D. et al., A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 2012, *30*, 918–920.

[17] Hart-Smith, G., Low, J. K. K., Erce, M. A., Wilkins, M. R., Enhanced methylarginine characterization by post-translational modification-specific targeted data acquisition and electron-transfer dissociation mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2012, *23*, 1376–1389.

[18] Green, D. M., Marfatia, K. A., Crafton, E. B., Zhang, X. et al., Nab2p is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. *J. Biol. Chem.* 2002, *277*, 7752–7760.

[19] Hsieh, C.-H., Huang, S.-Y., Wu, Y.-C., Liu, L.-F. et al., Expression of proteins with dimethylarginines in *Escherichia coli* for protein-protein interaction studies. *Protein Sci. Publ. Protein Soc.* 2007, *16*, 919–928.

[20] Huang, D. W., Sherman, B. T., Lempicki, R. A., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009, *37*, 1–13.

[21] Liu, Q., Dreyfuss, G., In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* 1995, *15*, 2800–2808.

[22] Kiledjian, M., Dreyfuss, G., Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* 1992, *11*, 2655–2664.

[23] Burd, C. G., Dreyfuss, G., Conserved structures and diversity of functions of RNA-binding proteins. *Science* 1994, *265*, 615–621.

[24] Hanakahi, L. A., Sun, H., Maizels, N., High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* 1999, *274*, 15908–15912.

[25] Mears, W. E., Rice, S. A., The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *J. Virol.* 1996, *70*, 7445–7453.

[26] Siebel, C. W., Feng, L., Guthrie, C., Fu, X. D., Conservation in budding yeast of a kinase specific for SR splicing factors. *Proc. Natl. Acad. Sci. USA* 1999, *96*, 5440–5445.

[27] Gilbert, W., Siebel, C. W., Guthrie, C., Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA* 2001, *7*, 302–313.

[28] Lukasiewicz, R., Nolen, B., Adams, J. A., Ghosh, G., The RGG domain of Npl3p recruits Sky1p through docking interactions. *J. Mol. Biol.* 2007, *367*, 249–261.

[29] Byvoet, P., Shepherd, G. R., Hardin, J. M., Noland, B. J., The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem. Biophys.* 1972, *148*, 558–567.

[30] Yu, M. C., Bachand, F., McBride, A. E., Komili, S. et al., Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes Dev.* 2004, *18*, 2024–2035.

[31] Iost, I., Dreyfus, M., Linder, P., Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J. Biol. Chem.* 1999, *274*, 17677–17683.

[32] Lipson, R. S., Webb, K. J., Clarke, S. G., Rmt1 catalyzes zinc-finger independent arginine methylation of ribosomal protein Rps2 in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 2010, *391*, 1658–1662.

[33] Brahms, H., Raymackers, J., Union, A., de Keyser, F. et al., The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J. Biol. Chem.* 2000, *275*, 17122–17129.

[34] Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., Lührmann, R., Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA N. Y. N* 2001, *7*, 1531–1542.

[35] Zhang, X., Kolaczkowska, A., Devaux, F., Panwar, S. L. et al., Transcriptional regulation by Lge1p requires a function independent of its role in histone H2B ubiquitination. *J. Biol. Chem.* 2005, *280*, 2759–2770.

[36] Wei, H.-M., Hu, H.-H., Chang, G.-Y., Lee, Y.-J. et al., Arginine methylation of the cellular nucleic acid binding protein does not affect its subcellular localization but impedes RNA binding. *FEBS Lett.* 2014, *588*, 1542–1548.

[37] Lott, K., Li, J., Fisk, J. C., Wang, H. et al., Global proteomic analysis in trypanosomes reveals unique proteins and conserved cellular processes impacted by arginine methylation. *J. Proteomics* 2013, *91*, 210–225.