

A Method for Large-scale Identification of Protein Arginine Methylation*[§]

Thomas Uhlmann^{†¶||}, Vincent L. Geoghegan^{†¶|}, Benjamin Thomas[§],
Gabriela Ridlova^{‡§}, David C. Trudgian^{§**}, and Oreste Acuto^{‡‡}

The lack of methods for proteome-scale detection of arginine methylation restricts our knowledge of its relevance in physiological and pathological processes. Here we show that most tryptic peptides containing methylated arginine(s) are highly basic and hydrophilic. Consequently, they could be considerably enriched from total cell extracts by simple protocols using either one of strong cation exchange chromatography, isoelectric focusing, or hydrophilic interaction liquid chromatography, the latter being by far the most effective of all. These methods, coupled with heavy methyl-stable isotope labeling by amino acids in cell culture and mass spectrometry, enabled in T cells the identification of 249 arginine methylation sites in 131 proteins, including 190 new sites and 93 proteins not previously known to be arginine methylated. By extending considerably the number of known arginine methylation sites, our data reveal a novel proline-rich consensus motif and identify for the first time arginine methylation in proteins involved in cytoskeleton rearrangement at the immunological synapse and in endosomal trafficking. *Molecular & Cellular Proteomics* 11: 10.1074/mcp.M112.020743, 1489–1499, 2012.

Knowledge of the type, extent and dynamics of post-translational modifications (PTMs)¹ reveals the changing network of protein interactions and the regulation of cellular functions. Methylated arginines (Meth-R) are relatively frequent in cellular proteins (e.g. 0.7–1% of total arginines (1) and this work) and have a very slow turnover (2), likely to confer lasting functional properties to proteins. Meth-Rs are often found at glycine-arginine-rich (GAR) sequences and modulate protein-

protein and protein-nucleic acid interactions by reducing hydrogen-bonding and local hydrophilicity (2). Methylation at arginine can weaken interactions but also enhance the binding of adaptors called Tudor domains to Meth-R-containing sequences (3, 4).

Meth-R occurs in three forms: asymmetric dimethylarginine (ADMA), bearing two methyl groups on one nitrogen of the guanidino group; symmetric dimethylarginine (SDMA), where both nitrogens are singly methylated and monomethylarginine (MMA), a reaction intermediate. In humans, these reactions are catalyzed by a family of protein arginine methyltransferases (PRMTs), with PRMT1, 3, 4, 6, and 8 forming ADMA and PRMT5 and 7 producing SDMA. In mice, PRMT1 and PRMT4 deficiency causes embryonic (5) or Perinatal (6) lethality, respectively, whereas lack of PRMT3 leads to growth defect (7). Arginine methylation has been investigated for its role in the regulation of gene expression at multiple levels and in DNA repair (2). PRMT1, 4, 5, 6, and 7 control the epigenetic code by methylating histone N-terminal tails (H1, H2A, 3, and 4) (4, 8–12) and tune the activity of transcription factors (2, 13–15), coactivators (16, 17) and corepressors (8, 11). Factors involved in mRNA splicing (18, 19), elongation (20), transport and translation (19, 21) often contain Meth-R, suspected to be involved in protein-RNA interaction (2). Moreover, T cell CD28 co-stimulation (22), TNF α (15), TLR4 (15), NGF (23) and NF κ B (15) signaling pathways may be regulated by PRMTs. Protein arginine methylation is implicated in pathogenic processes, including oncogenesis (24), cardiovascular disease (25), autoimmunity and viral infections (26), raising the possibility that abnormally methylated proteins can be disease markers and PRMTs potential therapeutic targets (27).

Only a very limited number of arginine methylated proteins and sites have been identified with certainty to date (see UniProtKB). *In vitro* elegant screenings were developed to search for PRMT substrates (21) but identifying the Meth-R concerned is usually cumbersome. The existence of Meth-R sites is in many cases only suspected by protein similarity such as the presence of GAR motifs, which are substrates for some but not all PRMTs (2). These vexing limitations hamper progress of a very promising field in biology. Previous work developed a powerful approach using heavy methyl-SILAC coupled to MS to identify with high certainty peptides containing Meth-R (28). An anti-ADMA antibody was used to

From the [†]T Cell Signalling Laboratory and [§]Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom

Received May 30, 2012, and in revised form, July 30, 2012

✂ Author's Choice—Final version full access.

Published, MCP Papers in Press, August 3, 2012, DOI 10.1074/mcp.M112.020743

¹ The abbreviations used are: PTMs, post-translational modifications; Meth-R, methylated arginines; ADMA, asymmetric dimethylarginine; GAR, glycine-arginine-rich; SDMA, symmetric dimethylarginine; MMA, monomethylarginine; PRMTs, protein arginine methyltransferases; HILIC, hydrophilic interaction liquid chromatography; SCX, strong cation exchange; SILAC, stable isotope labeling by amino acids in cell culture; PRAM, proline-rich arginine methylation; FDR, false discovery rate.

enrich for Meth-R-containing proteins but only a limited number of Meth-R were identified (28). The relative low abundance of Meth-R-containing proteins and peptides and lack of adequate methods for their enrichment remains therefore a major hurdle for both large- and small-scale discovery of Meth-R sites, leaving the heavy methyl-SILAC method essentially unexploited. Here, we have uncovered distinct physico-chemical properties of Meth-R-containing tryptic peptides and showed that hydrophilic interaction liquid chromatography (HILIC) is a simple method providing excellent enrichment of Meth-R-containing peptides. When coupled to heavy methyl-stable isotope labeling by amino acids in cell culture (SILAC) based MS, HILIC allowed us to identify hundreds of Meth-R sites, providing for the first time a means for “methylome”-scale investigation that vastly outperforms the use of anti-ADMA antibody.

EXPERIMENTAL PROCEDURES

SILAC Labeling and Cell Lysis for MS Analysis—Custom-made RPMI 1640 medium lacking L-Methionine, L-Arginine, and L-Lysine (Invitrogen, Carlsbad, CA) was supplemented with 10% dialyzed fetal calf serum (Invitrogen), L-Arginine and L-Lysine and either a) L-Methionine (CK Gas Products Ltd.), or b) L-Methionine-methyl-¹³C-methyl-D₃ (Sigma, ISOTEC™) at a final concentration of 0.1 mM L-Methionine, 0.29 mM L-Arginine and 0.219 mM L-Lysine and filter sterilized (0.22 μm pore size, Millipore, Billerica, MA). Jurkat cells were grown at 37 °C in a humidified 5% CO₂-containing atmosphere for 5–7 cell doublings in labeling media. Human primary CD4+ lymphocytes were isolated from whole blood of healthy donors by negative selection using a Dynal isolation kit (Invitrogen) according to the manufacturer's protocol. 1×10^6 to 2×10^6 CD4+ lymphocytes in 10 ml of light or heavy medium were then stimulated by plate bound anti-CD3 at 10 μg/ml and soluble CD28.2 (BioLegend, San Diego, CA) at 1 μg/ml. IL-2 (AbDSerotec) was added to a concentration of 50 ng/ml. Cells were grown at 37 °C in a humidified 5% CO₂-containing atmosphere for 7–9 days. Cells were harvested and lysed for either a) 10 min in ice-cold lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1% Nonidet P-40, 0.5% Deoxycholate, protease inhibitor mixture (Roche), or b) 20 min in 8 M Urea in 25 mM Tris pH 8. Lysates were cleared by centrifugation at $14,000 \times g$ for 20 min and mixed in a 1:1 ratio of protein content (measured by Bradford method and absorbance at 280 nm). Lysates were used for a) immunoprecipitation, or b) strong Cation exchange (SCX) chromatography, HILIC and OFFGEL isoelectric focusing.

Immunoprecipitation and In-gel Digestion—Immunoprecipitation was performed on the mixed cell lysates overnight at 4 °C on Sepharose protein G beads (GE Healthcare) pre-incubated with an anti-ADMA mAb 7E6 (Abcam, Cambridge, UK). Beads were washed four times with fresh ice-cold lysis buffer and eluted with 100 mM triethylamine pH 11.5 for 10 min at RT. The eluates were neutralized with $1/20$ volume of 1 M phosphate buffer pH 6.8, boiled in reducing SDS NuPAGE sample buffer (Invitrogen), alkylated with 55 mM iodoacetamide (Sigma) and separated on 4–12% gradient Bis-Tris NuPAGE gels (Invitrogen). The gel was washed with distilled water and lightly stained with Colloidal Blue (Invitrogen). The gel lane was divided into 15 slices and subject to GeLCMS/MS.

SCX, HILIC, and OFFGEL Fractionation—Cleared Urea lysate was acetone precipitated and resuspended in 1.6 M Urea in 25 mM Tris pH8. Overnight digestion at 37 °C was carried out with 12.5 ng/μl trypsin (Proteomics grade, Sigma) or chymotrypsin (Sequencing grade, Promega, Madison, WI). For SCX, 4–6 mg peptides were

acidified with 0.1% trifluoroacetic acid (Sigma) and cleared by centrifuging at $20,000 \times g$ for 10 min. Peptides were applied to a 1 ml SCX column (Resource S, GE Healthcare) using Buffer A (50 mM KH₂PO₄, 20% acetonitrile, pH2.7). Bound peptides were eluted with increasing Buffer B (50 mM KH₂PO₄, 20% acetonitrile, 1 M KCl, pH2.7). In a first region of separation the gradient was stepped up to 7% Buffer B to elute 1+ and most of the 2+ charged peptides. A shallow gradient to 35% Buffer B separated the remaining 2+ from 3+ and higher charged peptides. In the last stage the Buffer B was changed to pH7 and a steep gradient to 100% Buffer B ensured complete elution of bound peptides. Following C18 desalting (Empore Octadecyl C18, Sigma), the fractions were analyzed by LC-MS/MS.

For HILIC 1 mg of peptides were acidified with 0.1% trifluoroacetic acid (TFA) and cleared by centrifuging at 20,000g for 10 min. Peptides were loaded onto a reverse phase C2/C18 column (GE Healthcare) with Buffer A (0.1% formic acid, pH2.7). Elution of peptides was carried out with a steep gradient to 100% Buffer B (0.1% formic acid, 95% acetonitrile, pH2.7). Desalted peptides were then dried under vacuum, re-suspended in Buffer B and applied to a HILIC column (Merck). Bound peptides were eluted with a shallow gradient of Buffer A up to 80%.

For IEF on OFFGEL, 1 mg of peptides were loaded onto an IPG strip (13 cm pH7–11, GE Healthcare) with IPG buffer (pH7–11, GE Healthcare) according to manufacturer's instructions. Peptides were focused in solution with an Agilent 3100 OFFGEL fractionator (Agilent Technologies, Santa Clara, CA) for 20 kWh up to a maximum voltage of 8000 V. Focused peptides were removed and desalted with C18 (Empore Octadecyl C18, Sigma). Peptides were also extracted by immersing the paper wicks from the cathode end in 0.1% TFA for 1–2 h, followed by brief sonication. Peptides from the respective fractions were collected and adjusted to 0.1% TFA, 5% acetonitrile and applied to a homemade C18 stage tip for desalting. Bound peptides were washed with 0.1% TFA, 2% acetonitrile and eluted in 0.1% TFA, 60% acetonitrile. Eluted peptides were dried and sent to the PNAC facility in Cambridge (<http://www.bioc.cam.ac.uk/pnac/aaa.html>) for amino acid analysis on a Biochrome 30 Analyzer.

Mass Spectrometry Data Acquisition—Samples were analyzed on an Ultimate 3000 nano HPLC (Dionex, Camberley, UK) system run in direct injection mode coupled to either a LTQ XL Orbitrap mass spectrometer, or a Q Exactive mass spectrometer (Thermo Electron, Hemel Hempstead, UK). Samples were resolved on a 15 cm by 75 μm inner diameter picotip analytical column (New Objective, Woburn, MA), which was packed in-house with Repronil-Pur C18-AQ phase, 3 μm bead (Dr. Maisch, Germany). A 120 min gradient was used to separate the peptides and each sample was typically injected three times and data merged in order to increase sample coverage. The LTQ XL Orbitrap mass spectrometer was operated in a “Top 5” data dependent acquisition mode and the Q Exactive mass spectrometer was operated in a “Top 10” data dependent acquisition mode. Precursor scans were performed at a resolving power of 60,000, from which either five precursor ions (LTQ XL Orbitrap) or 10 precursor ions (Q Exactive) were selected and fragmented. Charge state +1 ions were rejected.

Mass Spectrometry Data Analysis—MS/MS peak lists were converted to mzXML format using ReAdW version 4.4.1 (LTQ XL Orbitrap data) or to mgf format using ProteoWizard msconvert release 3.0.3535 (Q Exactive data). Both tools were used with default parameters except that zlib compression of spectral data was enabled. Data was uploaded to the central proteomics facilities pipeline (CPFP at: https://cpfp-master.molbiol.ox.ac.uk/cpfp_demo/auth/login) (29). Files were searched using Mascot version 2.3.01 (Matrix science), XITANDEM version 2008.12.01.1 (30) and OMSSA version 2.1.8 (31) against a concatenated target and reversed decoy version of the IPI human sequence database (EBI), versions 3.78 and 3.79 containing

86,702 and 86,635 target protein sequences respectively. Enzyme was set to trypsin (or chymotrypsin for some experiments) allowing for up to 3 missed cleavages. Carbamidomethyl cysteine was set as a fixed modification. For analysis of arginine methylation oxidized methionine, heavy oxidized methionine, monomethylarginine, heavy monomethylarginine, dimethylarginine, and heavy dimethylarginine were set as variable modifications. For analysis of lysine methylation, oxidized methionine, heavy oxidized methionine, monomethyllysine, heavy monomethyllysine, dimethyllysine, heavy dimethyllysine, trimethyllysine, and heavy trimethyllysine were set as variable modifications. Mass tolerances for MS and MS/MS peak identifications were 10 ppm and 0.5 Da respectively (LTQ XL Orbitrap) or 20 ppm and 0.1 Da respectively (Q Exactive). Mass spectrometry data are available at <http://www.peptideatlas.org/PASS/PASS00081>. Assignment of a methylation site required identification by searches in the Central proteomics facility pipeline (CPFP) at 1% FDR and the presence of a confirming peptide in the precursor spectrum of equal intensity separated by a mass difference introduced by the light/heavy methyl group(s). The actual false positive rate of identification using the heavy methyl SILAC criteria (1.2%, [supplemental Fig. S5](#)) was calculated on a representative dataset by analyzing all false positive hits for arginine methylation for the presence of confirming heavy methyl SILAC pairs. This number was then doubled and divided by the total number of hits for arginine methylation. The requirement for the presence of a methylSILAC pair to corroborate each methylated peptide increases confidence and enables even low scoring peptides to be considered (28). Modification localization scoring was performed using the ModLS algorithm (unpublished²) which extends the PTMScore method to incorporate automatic specificity expansion. For each variable modification chosen for the database search all amino acid specificities defined in the Unimod database (www.unimod.org) are considered during localization. This allows the correct localization of e.g. dimethylation to lysine, even if only arginine methylation was chosen for the search. Ambiguous modification localizations are annotated as such in [supplemental Table S5](#). Within CPFP, inference of protein identifications from peptide sequences was performed using ProteinProphet (32). Default parameters were used for protein grouping and the assignment of isoform identifications. Precursor *m/z*, charge, InterProphet probability for all methylated peptides found in this work are listed in [supplemental Table S4](#).

RESULTS

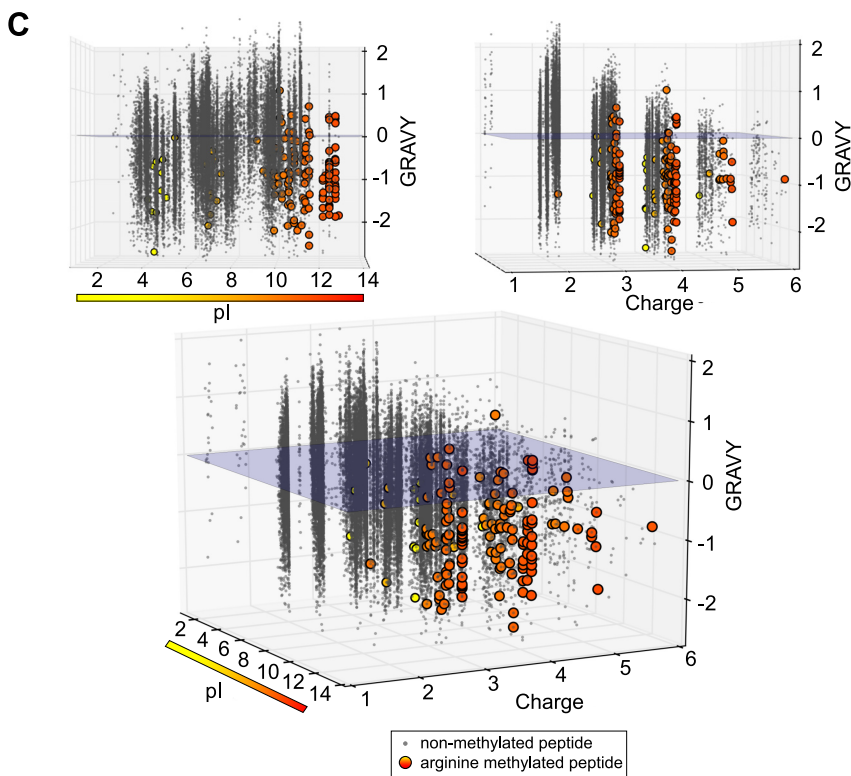
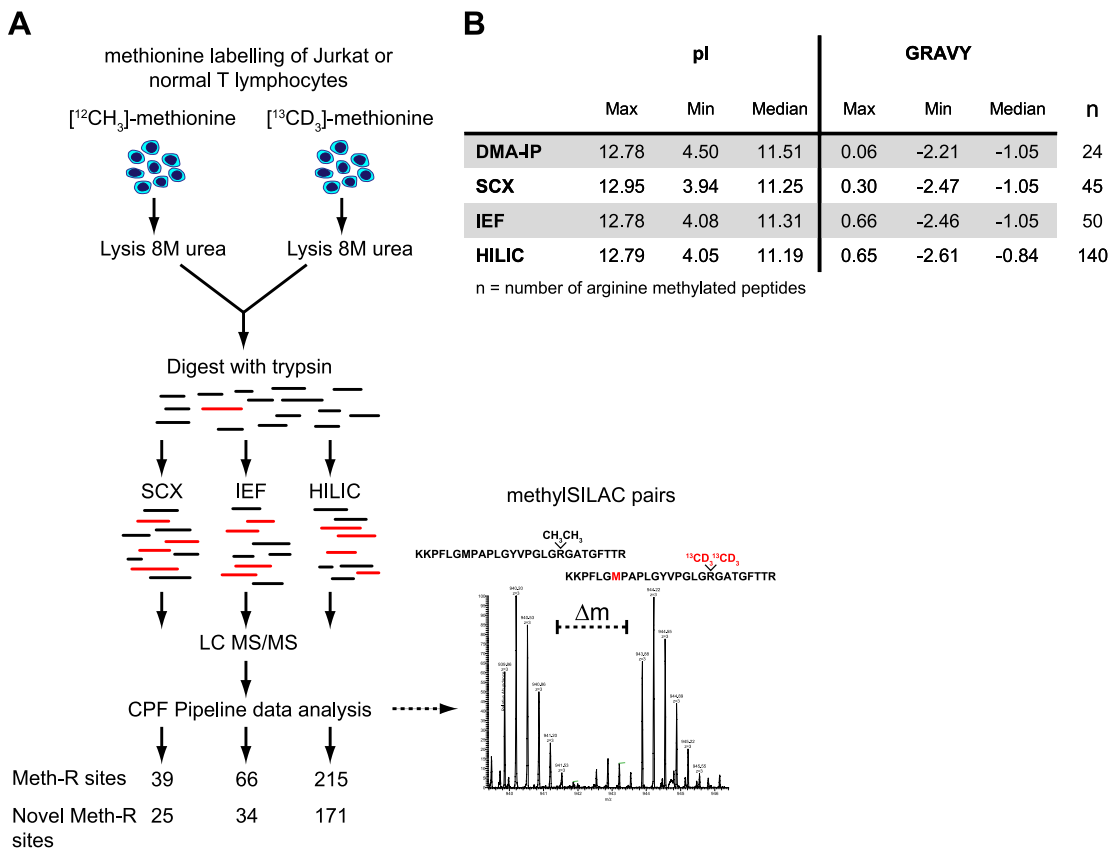
Rationale and Methods for Enrichment of Meth-R Containing Peptides—Biosynthetic labeling with methionine (M0) and [¹³CD₃]-methionine (M4) (heavy methyl-SILAC) (Fig. 1A) enables detection of protein methylation sites by LC-MS/MS with high confidence (28). In this study, we labeled Jurkat T cells by heavy methyl-SILAC and extended this approach to primary cells by showing efficient SILAC labeling of antigen receptor stimulated T-cells ([supplemental Figs. S1, S2](#)). In a previous study, enrichment of Meth-R-containing proteins from HeLa cells with the monoclonal antibody (mAb) 7E6 to MMA and ADMA led to the identification of 59 Meth-R sites (28). By a similar approach (see Methods), we identified 41 Meth-R sites in Jurkat T cells, only 6 of which were novel (e.g. not annotated in the UniProtKB database) ([supplemental Ta-](#)

[ble S1](#)). However, cell proteomes should comprise higher numbers of Meth-R-containing proteins, since 0.7–1.0% of arginine in proteins extracted from cell lines or tissues is methylated (1) (and our unpublished data). Moreover, similar to previous investigations (28), we identified Meth-R sites almost invariably at RG sites ([supplemental Table S1](#)), pointing again at the limited usefulness of 7E6 mAb for proteomic-scale studies. These limitations and the small number of Meth-R sites (137) known to-date (e.g. in UniProtKB), spurred us to develop effective, unbiased approaches that enrich for Meth-R-containing peptides for large-scale identification of Meth-R-containing sites using LC-MS/MS analysis (Fig. 1A).

SCX Chromatography—Approximately 70% of tryptic peptides in proteomes exhibit a +2 net charge at pH 2.7 and can be separated from other peptides according to single charge differences using SCX chromatography (33). Because trypsin cuts poorly at Meth-R, the missed cleavages produce peptides with charge $\geq +3$ at pH 2.7. Thus, most Meth-R peptides should be present in the $\geq +3$ peptide-containing fractions after SCX chromatography. SCX fractions enriched for tryptic peptides of charge $< +3$ from total proteins of Jurkat cell, exhibited $\sim 1.0\%$ of Meth-R (as revealed by amino acid analysis, see Methods and [supplemental Fig. S3A](#)) but no Meth-R peptide could be identified by LC-MS/MS (data not shown). In contrast, in SCX fractions of tryptic peptides of charge $\geq +3$, $\sim 4.0\%$ of arginine was present as Meth-R ([supplemental Fig. S3A](#)) allowing us to identify 39 Meth-R sites from Jurkat and primary human T cells in four independent experiments (Fig. 1A and [supplemental Table S1](#)), of which 25 were new and, interestingly, 4 were not at RG motifs ([supplemental Table S1](#)). SCX fractionation enabled us to discover 10 new Meth-R containing proteins and Meth-R within sequences rarely selected by the 7E6 mAb ([supplemental Table S1](#)). In addition, as predicted by the poor trypsin cleavage at Meth-K, a known Meth-K peptide was also found (histone H4, [supplemental Table S1](#)).

IEF—Histidine-containing tryptic peptides (charge $\geq +3$ at pH 2.7) were the major contaminant (60–70%) within the SCX fractions in which most Meth-R peptides were detected (not shown). This likely contributed to the analyte complexity, limiting the number of Meth-R peptides identified. However, these data brought to our attention key physico-chemical properties of Meth-R tryptic peptides. Arginines prone to be methylated are likely to be fully exposed to solvent and therefore most often embedded within sequences containing hydrophilic and/or neutral amino acid residues. Moreover, we noticed that sequences surrounding Meth-R sites (from UniProtKB and those found by 7E6 Ab and SCX enrichment) infrequently contained negatively charged amino acids. This suggested that tryptic peptides carrying one or more Meth-R should frequently display basic isoelectric points (pIs). This was indeed the case for Meth-R peptides found by 7E6 mAb and SCX enrichment (Fig. 1B) with a median pI > 11.00 . To test this prediction on a larger scale we fractionated tryptic

² David C. Trudgian, Rachele Singleton, Matthew E. Cockman, Peter J. Ratcliffe, and Benedikt M. Kessler, ModLS: Post-translational modification localization scoring with automatic specificity expansion 2012 (submitted).



digests from heavy methyl-SILAC-labeled Jurkat or primary CD4+ lymphocytes by isoelectric focusing on IPG strips (Fig. 1A, see Methods). The results of four independent experiments consistently showed that > 90% of Meth-R peptides displayed a $pI \geq 9$ with a median pI of 11 (Figs. 1B and 1C), indicating that Meth-R tryptic peptides are largely highly basic in nature. Interestingly, at the basic end of the IEF gradient, (pH 9–11 and in the electrode pad itself) the enrichment for Meth-R peptides was substantial, reaching values of $\geq 5\%$ of peptides containing Meth-R as revealed by LC-MS/MS analysis (not shown). However, the lack of IPG strips with a suitable pH range precluded good recovery of most of the highly basic peptides. Using IEF we were able to detect 66 Meth-R sites of which 34 were new (Fig. 1A and supplemental Table S1) and found 20 that did not contain an RG motif (supplemental Table S1). IEF fractionation permitted the identification of 19 proteins previously unknown to contain Meth-R and 3 Meth-K sites.

HILIC—HILIC has recently been shown to be an excellent method to complement reverse-phase fractionation (34) by enriching for highly hydrophilic peptides (35). Because the vast majority of Meth-R-containing tryptic peptides identified after 7E6 mAb, SCX or IEF enrichment were highly hydrophilic with a median GRAVY score (36) of -1.05 (Fig. 1B), we reasoned that HILIC fractionation should also enable selection for Meth-R containing peptides. Trypsin digests of Methyl-SILAC Jurkat or primary CD4+ lymphocytes were loaded onto a HILIC column in 90% acetonitrile. Bound peptides were eluted with increasing concentration of aqueous buffer (supplemental Fig. S3B) and were subjected to amino acid and LC-MS/MS analysis. Fractions containing highly hydrophilic peptides (see supplemental Fig. S3B) were enriched ~ 10 fold for Meth-R, as compared with fractions with fewer hydrophilic peptides, in which no Meth-R-containing peptide could be detected (not shown). Meth-R peptides identified after HILIC separation had a median GRAVY score of -0.84 (Fig. 1B), a value similar to that of Meth-R peptides found by anti-DMA, SCX and IEF. In five independent experiments, LC-MS/MS analysis of the twelve most hydrophilic HILIC fractions led to the identification of 215 Meth-R sites (Fig. 1A and supplemental Table S1). Notably, HILIC enabled also the detection in chymotrypsin-digested peptides of 37 Meth-R sites not found after tryptic digestion (supplemental Table S1).

HILIC alone identified 3 to 5 times higher numbers than 7E6 mAb, SCX and IEF individually and covered $\sim 70\%$ of the 101 unique sites identified by the other three methods together (supplemental Table S1). Fig. 2A illustrates the power of HILIC method to discover Meth-R sites as compared with SCX, 7E6 and IEF individually. Importantly, 171 ($\sim 80\%$) of Meth-R sites found with HILIC were new (Fig. 1A and Table S1), the highest discovery rate among the four different approaches. 87 ($\sim 50\%$) of these new sites were at non-RG motifs. Mono-methylated-Rs were more frequently seen with HILIC than with the three other methods. HILIC alone allowed us to assign 215 Meth-R sites to 115 proteins, 94 of which were not previously known to be substrates of PRMTs. In total, 190 (77%) of the 249 unique Meth-R sites identified by 7E6 mAb, SCX, IEF and HILIC together were new (Fig. 2B and supplemental Table S1). Unique sites were assigned to 131 different proteins, 93 (72%) of which were previously unknown to be methylated on arginine (Fig. 2C, supplemental Table S1 and Fig. 5). Our study brings the total number of Meth-R sites and Meth-R proteins in the human proteome to 327 and 149 (Fig. 2C), respectively, substantially widening, in a single study, our insight into the regulation of cellular functions by PRMTs. HILIC was vastly superior to each of the three other enrichment methods alone or combined for detecting known and new Meth-R sites.

Bioinformatic Analysis of Meth-R Sites and Meth-R Proteins—Heavy methyl-SILAC validation (28) greatly increases the confidence in the identification of Meth-R (and Meth-K) containing peptides. Excluding heavy methyl-SILAC validation, reduced the confidence for Meth-R containing peptides to 67% FDR (despite the fact that for the overall dataset of modified and unmodified peptides the FDR was 1%). However, inclusion of putative Meth-R peptides requiring a corresponding light/heavy methyl-SILAC partner raised the confidence of identifying Meth-R sites to $\sim 1\%$ FDR (supplemental Fig. S5). Therefore we feel highly confident about the Meth-R sites identified in our study, which were all manually validated by inspecting individual light/heavy pairs. This is further supported by the fact that our investigation found 43% of the known Meth-R sites.

Of the 190 new Meth-R sites discovered in our study, 57% occurred at RG sites, at times within GAR motifs (Fig. 3A and supplemental Table S1), thought to be mainly substrates for

FIG. 1. A, Schematic representation of our workflow. Cells were labeled with light or heavy methionine, lysed and combined in a 1:1 protein ratio. Proteins were digested with trypsin or, in some experiments, with chymotrypsin. Arginine methylated peptides (depicted as *red lines*) were enriched through the use of SCX, IEF (in an OFFGEL apparatus) or HILIC and analyzed by mass spectrometry. MS spectra for putative arginine methylated peptides were manually verified for the presence of a 1:1 methyl-SILAC pair. An example methyl-SILAC pair is shown. For each separation method, the total number of Meth-R sites identified as well as the novel ones is indicated; *B*, Meth-R-containing tryptic peptides have distinct physico-chemical properties. The Table indicates maximum, minimum and median values for isoelectric points (pI) and GRAVY scores of Meth-R peptides identified with the indicated enrichment methods. The number of unique peptides (n) from each method used to calculate values is shown. *C*, 3D scatter plot of 33000 non-methylated tryptic peptides (*gray circles*) identified from pH 3–11 IPG strip and 128 arginine methylated tryptic peptides (*yellow-red circles*) identified from anti-DMA-IP, HILIC, SCX and IEF. Three different views of the same plot are shown. The peptides are plotted according to pI , GRAVY score and charge state at pH 2.7, representing the different selection criteria exploited by the enrichment methods. *Yellow* represents a low pI and *red* a high pI .

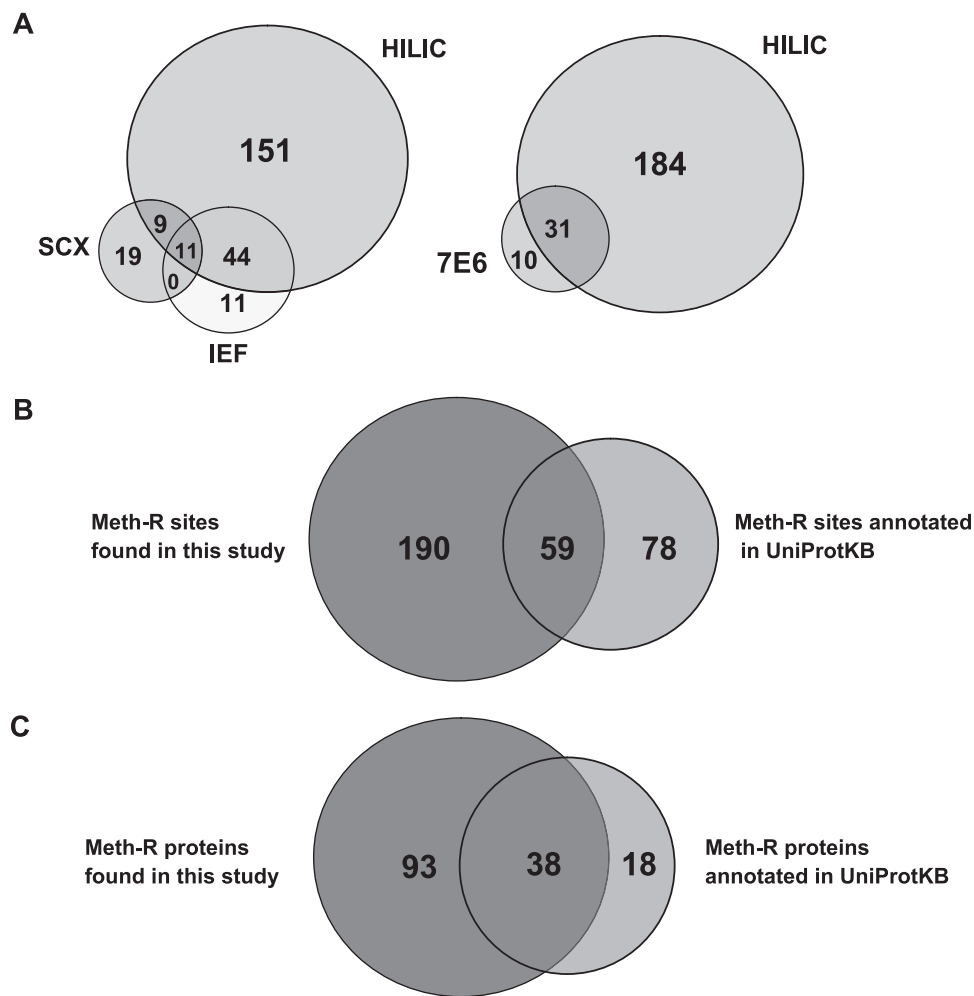


FIG. 2. **A**, Venn diagrams of the number of Meth-R sites identified by HILIC, IEF, and SCX individually and combined (*left*) or unique in HILIC or 7E6 mAb or found by both methods (*right*) in Jurkat cells and primary human CD4⁺ lymphocytes combined. **B**, Venn diagram showing the number of Meth-R sites found only in this study, the Meth-R sites annotated in UniProtKB database and those in common. **C**, Venn diagram of the number of arginine methylated proteins identified in this study, those previously reported in the UniProtKB database and those in common. Circles are approximately scaled relative to number of sites.

PRMT1 (2). The remaining 43% were found at non-RG motifs. This is a considerable increase in the frequency of non-RG containing sequences among Meth-R sites when compared with 25% within the 137 sites known to date ([supplemental Fig. S6](#)). In ~ 50% of non-RG sites found here, proline occurred quite frequently before and/or after R (Fig. 3A). Such a feature was also observable, though less frequently (32%), in previously identified non-GAR containing methylation sites (not shown). Inspecting the set of non-RG sites for the occurrence of proline at a particular position flanking Meth-R (see Methods), revealed that proline at positions -1 (R being 0) and +4, was significantly favored, the sequences containing proline at either or both of these positions (x(P)Rxxx(P)) accounting for 33% of all the methylation sites we found (Fig. 3B). We have provisionally designated this subset of new Meth-R consensus sites as Proline-Rich Arginine Methylation (PRAM) motif. Although glycine and proline often followed the

Meth-R, we found a number of sites where the Meth-R was followed by leucine and alanine (Fig 3C). Clearly, with growing numbers of Meth-R sites to be identified in future proteomics studies, it should be possible to refine consensus site subsets and help assign PRMT specificities.

The newly discovered Meth-R proteins appear to reside in the cytoplasm and/or nucleus with no preference for either compartment Fig. 4A). As previously noticed (2), a high proportion of Meth-R proteins are known to bind nucleic acids and are involved in regulating chromatin structure, transcription, RNA processing and translation. Our study considerably increases the number of such proteins regulated by PRMTs, including five implicated in chromatin structure, 22 participating in the transcription process and 21 in RNA processing (see Figs. 4B and 5, blue circles). Most interestingly, we also found proteins implicated in other cellular functions not previously reported to be subjected to regulation by PRMTs. Among

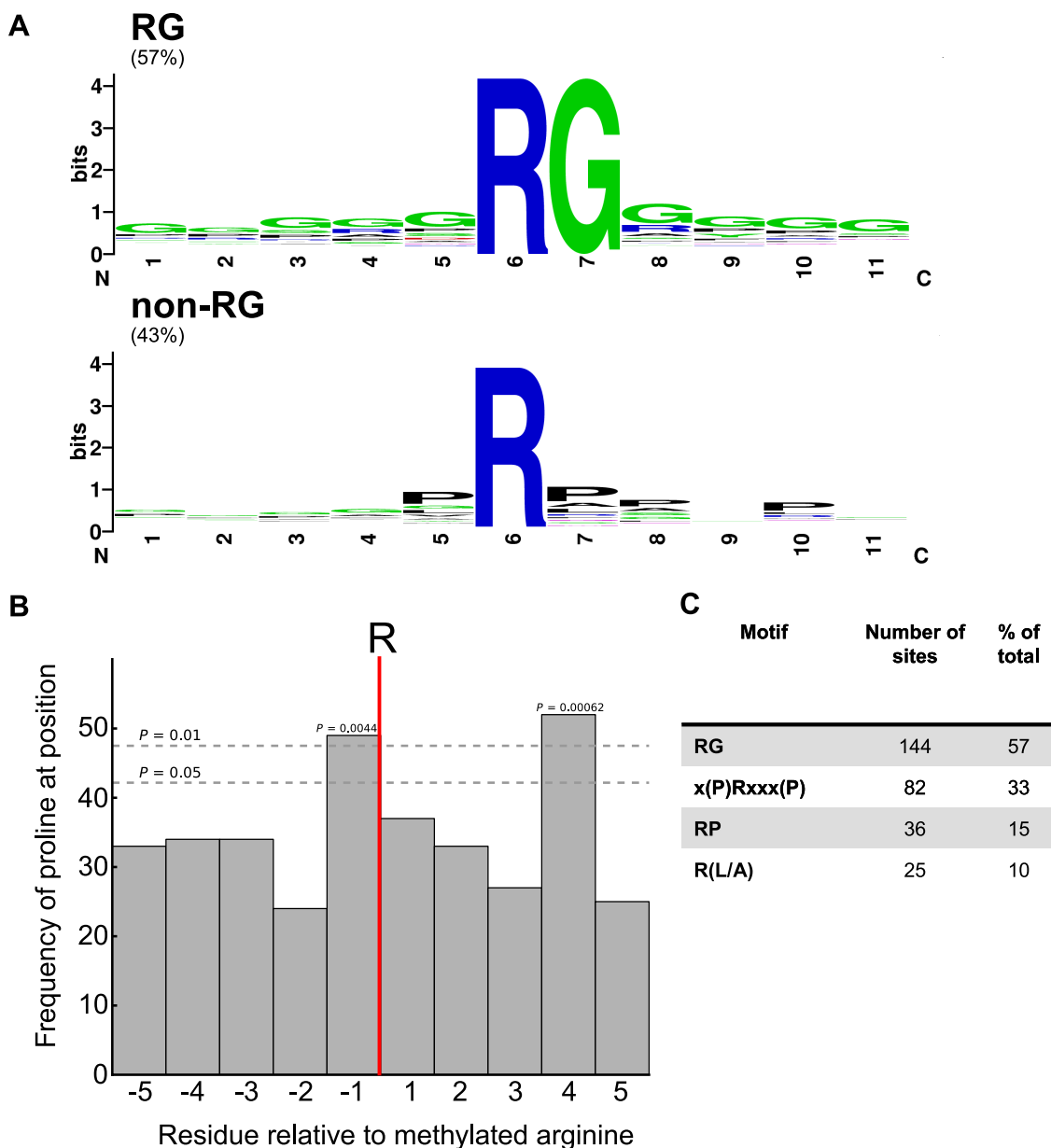


FIG. 3. A, Sequence logo (41) representation of all arginine methylation sites identified in this study. Of methylated arginines, 57% were followed by a glycine and were generally flanked by a glycine rich sequence. The remaining 43% of sites reveal a previously unreported proline-rich arginine methylation (PRAM) motif. **B,** Frequency of proline at each position relative to the methylated arginine. Dashed lines indicate frequency level of proline with a probability of occurrence by chance of 0.05 or 0.01. *p* values for proline frequency at position -1 and +4 are shown. *p* values were calculated by bootstrapping. **C,** Most common motifs found at Meth-R sites. Note that motifs can be composed of combinations of motifs listed.

these (see Fig. 5) were two nuclear pore components (RANBP2 and TPR) and 6 proteins (DNM2, PICALM, SEC24C, GGA3, SNX3, and CLINT1 (see, UniProtKB) implicated in endosomal trafficking and three, WASP/WIP complex (37) and DNM2 (38), involved in actin cytoskeletal re-arrangement at the immunological synapse (IS). A considerable number of Meth-R sites were found in proteins of unknown function (Fig. 5).

Although SCX and HILIC could in principle enrich also for Meth-K proteins, we found only 10 lysine methylation sites

(supplemental Table S1 and Fig. 5). Perhaps Meth-K occurs less frequently than Meth-R in proteins and/or Meth-K- and Meth-R-containing peptides present different physico-chemical properties and thus both would not be enriched by the methods used here. Nevertheless, we found three proteins previously unknown to be Meth-K, including two heat shock proteins, HSPA5 and HSPA8.

Overall, the methods explored here, with HILIC being by far the most effective, allowed us to considerably enrich for

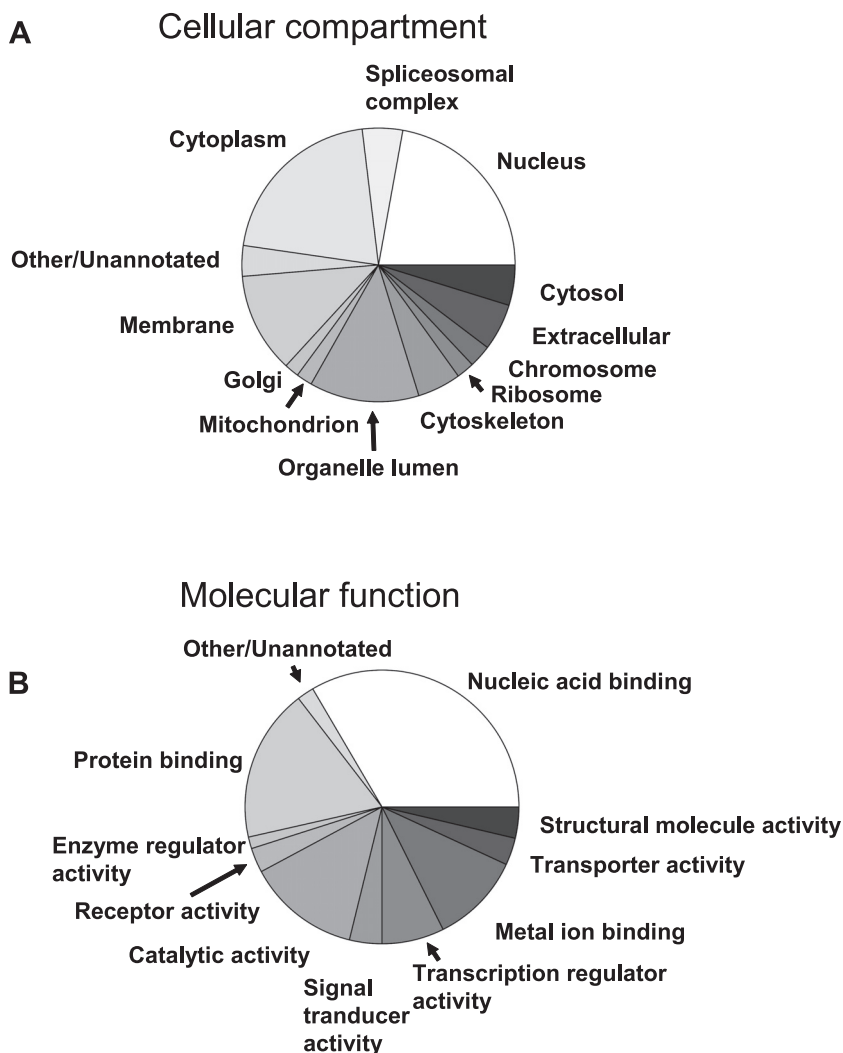


FIG. 4. Breakdown of arginine methylated proteins according to distribution in cellular compartments (A) or by molecular function (B) according to Gene Ontology (GO) annotation.

Meth-R peptides other than those in canonical RG/GAR motifs and uncover a substantial number of new proteins and functions likely to be regulated by PRMTs.

DISCUSSION

To date, no effective and unbiased method for the enrichment of protein and/or peptide containing Meth-R exists. An mAb (7E6) directed at ADMA and MMA permitted the identification of only a handful of Meth-R sites (28) (and this work) almost exclusively at RG sites. Because antibodies are unlikely to effectively discriminate (and with high affinity) the subtle chemical difference between Meth-R and unmodified R, use of pan Meth-R antibodies may not be suitable. A major obstacle in the identification of MS-based Meth-R, namely, being incorrectly mistaken for other isobaric chemical modifications, was recently solved by using a heavy methyl-SILAC approach (28), raising the possibility of accurate large-scale studies of protein arginine methylation. Despite this, a simple and robust method was missing for effective enrichment of Meth-R-containing peptide for small- and proteomic-scale

analysis of arginine methylation. The exploratory phase of this investigation revealed that >90% of Meth-R-containing tryptic peptides displayed high basic pI and high hydrophilicity. Enrichment using IEF was barred by the unavailability of IPG strips with adequate high pH range. We found however that HILIC had excellent capacity to enrich for Meth-R peptides. HILIC alone allowed us to identify from trypsin or chymotrypsin digests of heavy methyl-SILAC cell lysates Jurkat cells and primary CD4+ lymphocytes, 215 Meth-R sites, 80% of which were new and covered almost 90% of the total 190 new Meth-R sites discovered here by all four methods. Our study more than doubles the 137 Meth-R sites known to date and adds 93 Meth-R-containing proteins to the 59 previously known in humans, substantially increasing in a single effort our knowledge of the extent and biological significance of arginine methylation. Although these numbers seem limited in comparison to those reported for other PTMs, it is possible that arginine methylation is more restricted and selective, also considering that in this study we explored only the T cell proteome. Although

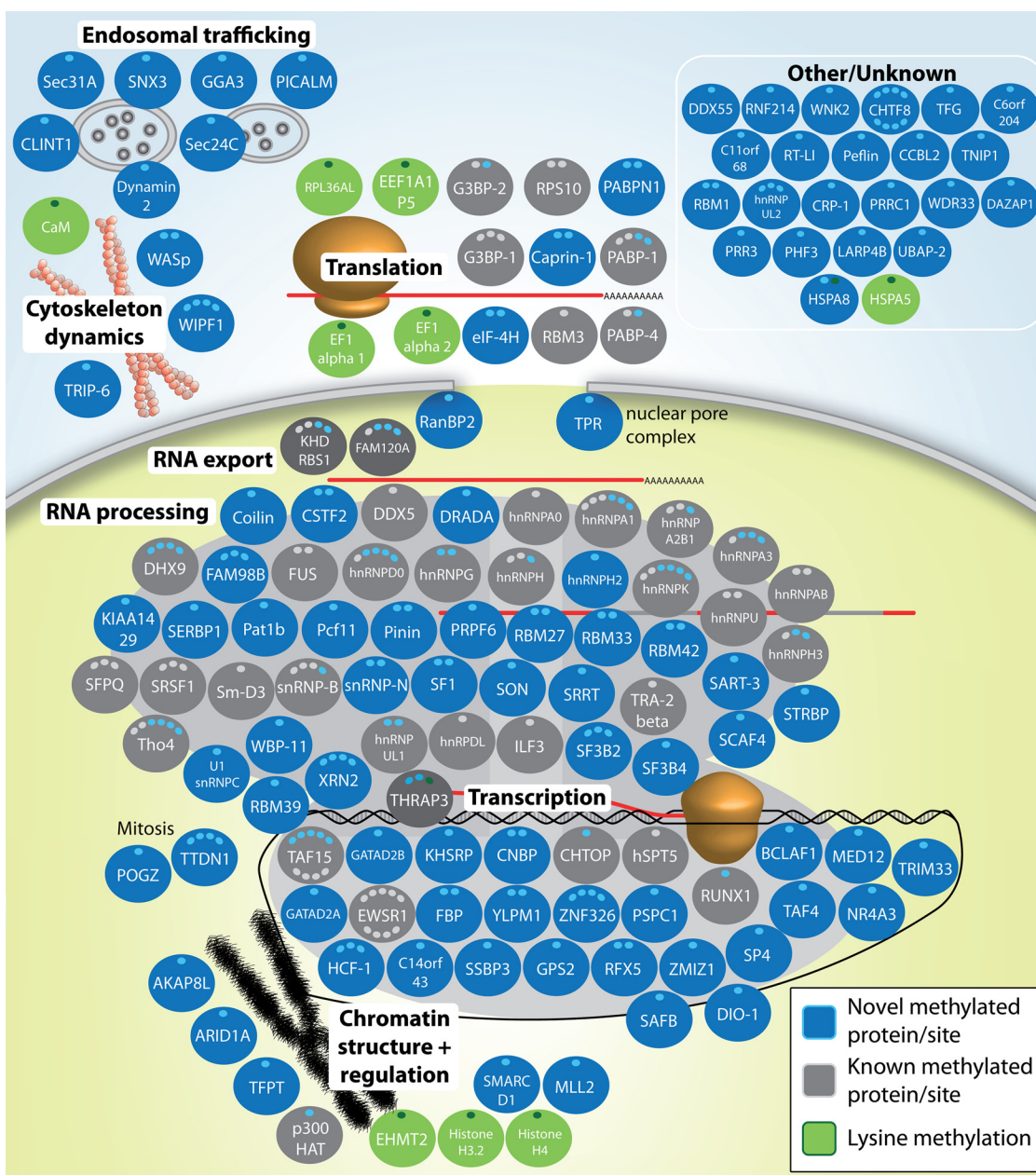


FIG. 5. **Overview of identified proteins containing methylated arginine/lysine.** Proteins are grouped according to annotated function in the UniprotKB database. Proteins newly identified as PRMT substrates in this study are drawn in *dark blue*. Previously known PRMT substrates are shaded in *gray*. Individual arginine methylation sites are marked as *dots*, *blue* representing novel Meth-R sites and *gray* representing previously known Meth-R sites. Green denotes a protein methylated on lysine.

HILIC performance overshadowed SCX and IEF, it missed 40 Meth-R sites identified by the latter two methods, despite a similar number of analysis performed. However, *pl* and GRAVY score of HILIC and SCX/IEF Meth-R peptides did not significantly differ, suggesting that the major cause of missed-hits by HILIC could be a persistent complexity of the peptide mixture in the HILIC fractions. In an effort to discover higher numbers of Meth-R sites after HILIC enrichment, we made use of a Q Exactive Orbitrap (see Methods), a powerful new MS/MS instrument that allows faster frag-

mentation rates, hence higher numbers of peptide identified (39). This allowed us to confirm in a single experiment about half of Meth-R sites identified by several analyses of HILIC fractions on an LTQ XL Orbitrap and led us to discover 11 new ones, indicating that the Q Exactive is largely preferable for proteomics scale discovery of Meth-R. Meth-R proteins detected in Jurkat cells and primary CD4⁺ lymphocytes overlapped only partially (supplemental Fig. S4). Possible differences in protein expression and/or PRMT levels in the two cell types may explain this divergence, considering that

Jurkat is a T cell lymphoma and that cancer cells may express modified repertoires of Meth-R proteins (24). Similarly, differences in tissue expression may explain why our data missed 18 Meth-R proteins in the mammalian proteome.

Only 25% of Meth-R sites annotated to date are at non-RG sites (supplemental Table S2). Some contain prolines and may be CARM1 substrates (21). In contrast to the use of 7E6 mAb, HILIC (as well as SCX and IEF) enrichment seems to be unbiased, as almost half of the sites we identified were at non-RG sites (supplemental Table S1 and Fig. 3A). The relatively low number of Meth-R sites known to date are largely derived from the most abundant proteins implicated in RNA metabolism (supplemental Table S3), as RG sites tend to occur at repeated GAR sequences of RNA binding proteins. However, only 7 such proteins (of 56 total) cover half of the 137 Meth-R sites known to date (see supplemental Tables S2 and S3). Thus, HILIC alone removes the above biases and reveals that PRMTs control proteins implicated in many diverse cellular functions.

Meth-R occurring at GAR sequences of RNA-binding proteins may contribute to protein-RNA interaction (2). The PRAM sequences defined in our study could be part of protein-protein interaction sites and Meth-R may modulate such interactions (40). The PRxxxP motif is predicted to be a poly-proline II helix able to interact with SH3 and other protein domains. Inspection of methylated PRxxxP motif revealed that a hydrophobic amino acid often precedes the +4 proline (PRxxΦP) (supplemental Fig. S7), a signature for poly-proline II helix interacting with SH3 domains and supports a role for PRMTs in modulating protein-protein interactions (40). Interestingly, our data revealed that the WASP/WIP complex and DNM-2, both involved in the formation and maintenance of the IS induced by TCR signaling, contained PRAM sequences known to be part of SH3 binding sites for several binding partners (37, 38). These data, together with the identification of novel Meth-R proteins implicated in endosomal trafficking, chromatin remodeling, transcription, RNA processing and translation open new perspectives to further uncover the role of PRMTs in the T cell activation process.

While HILIC represents a simple and robust method to enrich for Meth-R peptides, a few technical improvements lie ahead for routine “methylome”-scale studies. One immediate possibility to de-complex further peptide mixtures prior to MS analysis is orthogonal combination of two chromatographic methods according to the physico-chemical features of Meth-R-containing peptides highlighted in this study. Large/highly charged peptides might escape detection by MS, a problem that can be partly overcome using proteases other than trypsin. Indeed, chymotrypsin enabled us to identify a sizable number of Meth-R-containing peptides not found in trypsin digests. Moreover, SCX, IEF and, preferentially HILIC, can be readily employed in small-scale formats to discover arginine methylation sites in low complexity peptide mixtures such as

those derived from digests of purified protein complexes or individual proteins, proven or suspected to interact with PRMTs.

The discovery of methods for unbiased enrichment of Meth-R-containing peptides, such as HILIC, will also allow us to ask why and to what extent abnormal protein arginine methylation plays a role in carcinogenesis (e.g. prostate cancer and breast cancer) and several other, important pathologies, including cardiovascular and infectious diseases, contributing perhaps to opening avenues for new therapeutic tools and drugs.

* This work was supported by Wellcome Trust Grant GR076558MA to O.A. V.G. was supported by a PhD studentship from the BBSRC, T.U. was supported by a Swiss National Science Foundation fellowship and a Roche Research Foundation fellowship.

§ This article contains supplemental Figs. S1 to S7 and Tables S1 to S5.

¶ T.U. and V.G. contributed equally to this work.

|| Current address: Dualsystems Biotech AG, Grabenstrasse 11a, 8952 Schlieren, Switzerland.

** Current address: Biochemistry Dept and Proteomics Core UT Southwestern Medical Center at Dallas 6001 Forest Park Rd, Dallas, TX 75390-8816, USA.

‡‡ To whom correspondence should be addressed: T Cell Signaling Laboratory, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom. Tel.: ++44 1865 275615; Fax: ++44 1865 275515; E-mail: oreste.acuto@path.ox.ac.uk.

REFERENCES

1. Bulau, P., Zakrzewicz, D., Kitowska, K., Wardega, B., Kreuder, J., and Eickelberg, O. (2006) Quantitative assessment of arginine methylation in free versus protein-incorporated amino acids in vitro and in vivo using protein hydrolysis and high-performance liquid chromatography. *Bio-Techniques* **40**, 305–310
2. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13
3. Sprangers, R., Groves, M. R., Sinning, I., and Sattler, M. (2003) High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J. Mol. Biol.* **327**, 507–520
4. Yang, Y., Lu, Y., Espejo, A., Wu, J., Xu, W., Liang, S., and Bedford, M. T. (2010) TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol. Cell* **40**, 1016–1023
5. Yu, Z., Chen, T., Hébert, J., Li, E., and Richard, S. (2009) A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Mol. Cell. Biol.* **29**, 2982–2996
6. Kim, J., Lee, J., Yadav, N., Wu, Q., Carter, C., Richard, S., Richie, E., and Bedford, M. T. (2004) Loss of CARM1 results in hypomethylation of thymocyte cyclic AMP-regulated phosphoprotein and deregulated early T cell development. *J. Biol. Chem.* **279**, 25339–25344
7. Swiercz, R., Person, M. D., and Bedford, M. T. (2005) Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochem. J.* **386**, 85–91
8. Ancelin, K., Lange, U. C., Hajkova, P., Schneider, R., Bannister, A. J., Kouzarides, T., and Surani, M. A. (2006) Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat. Cell Biol.* **8**, 623–630
9. Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlantz, H., Lüscher, B., and Amati, B. (2007) Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449**, 933–937
10. Iberg, A. N., Espejo, A., Cheng, D., Kim, D., Michaud-Levesque, J., Richard, S., and Bedford, M. T. (2008) Arginine methylation of the histone H3 tail impedes effector binding. *J. Biol. Chem.* **283**, 3006–3010

11. Jelinic, P., Stehle, J. C., and Shaw, P. (2006) The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.* **4**, e355
12. Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., Cerruti, L., Curtis, D. J., Patel, D. J., Allis, C. D., Cunningham, J. M., and Jane, S. M. (2009) PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* **16**, 304–311
13. Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Nimer, S. D. (2008) Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev.* **22**, 640–653
14. Jansson, M., Durant, S. T., Cho, E. C., Sheahan, S., Edelmann, M., Kessler, B., and La Thangue, N. B. (2008) Arginine methylation regulates the p53 response. *Nat. Cell Biol.* **10**, 1431–1439
15. Covic, M., Hassa, P. O., Sacconi, S., Buerki, C., Meier, N. I., Lombardi, C., Imhof, R., Bedford, M. T., Natoli, G., and Hottiger, M. O. (2005) Arginine methyltransferase CARM1 is a promoter-specific regulator of NF- κ B-dependent gene expression. *EMBO J.* **24**, 85–96
16. Chen, D., Ma, H., Hong, H., Koh, S. S., Huang, S. M., Schurter, B. T., Aswad, D. W., and Stallcup, M. R. (1999) Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174–2177
17. Fathman, J. W., Gurish, M. F., Hemmers, S., Bonham, K., Friend, D. S., Grusby, M. J., Glimcher, L. H., and Mowen, K. A. (2010) NIP45 controls the magnitude of the type 2 T helper cell response. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3663–3668
18. Liu, Q., and Dreyfuss, G. (1995) In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* **15**, 2800–2808
19. Ohkura, N., Takahashi, M., Yaguchi, H., Nagamura, Y., and Tsukada, T. (2005) Coactivator-associated arginine methyltransferase 1, CARM1, affects pre-mRNA splicing in an isoform-specific manner. *J. Biol. Chem.* **280**, 28927–28935
20. Kwak, Y. T., Guo, J., Prajapati, S., Park, K. J., Surabhi, R. M., Miller, B., Gehrig, P., and Gaynor, R. B. (2003) Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol. Cell* **11**, 1055–1066
21. Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* **25**, 71–83
22. Blanchet, F., Cardona, A., Letimier, F. A., Hershfield, M. S., and Acuto, O. (2005) CD28 costimulatory signal induces protein arginine methylation in T cells. *J. Exp. Med.* **202**, 371–377
23. Cimato, T. R., Tang, J., Xu, Y., Guarnaccia, C., Herschman, H. R., Pongor, S., and Aletta, J. M. (2002) Nerve growth factor-mediated increases in protein methylation occur predominantly at type I arginine methylation sites and involve protein arginine methyltransferase 1. *J. Neurosci. Res.* **67**, 435–442
24. Cheung, N., Chan, L. C., Thompson, A., Cleary, M. L., and So, C. W. (2007) Protein arginine-methyltransferase-dependent oncogenesis. *Nat. Cell Biol.* **9**, 1208–1215
25. Wang, Z., Tang, W. H., Cho, L., Brennan, D. M., and Hazen, S. L. (2009) Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1383–1391
26. Bedford, M. T., and Richard, S. (2005) Arginine methylation an emerging regulator of protein function. *Mol. Cell* **18**, 263–272
27. Copeland, R. A., Solomon, M. E., and Richon, V. M. (2009) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug discovery* **8**, 724–732
28. Ong, S. E., Mittler, G., and Mann, M. (2004) Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1**, 119–126
29. Trudgian, D. C., Thomas, B., McGowan, S. J., Kessler, B. M., Salek, M., and Acuto, O. (2010) CPFP: a central proteomics facilities pipeline. *Bioinformatics* **26**, 1131–1132
30. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467
31. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–964
32. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
33. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12130–12135
34. Jian, W., Edom, R. W., Xu, Y., and Weng, N. (2010) Recent advances in application of hydrophilic interaction chromatography for quantitative bioanalysis. *J. Separation Sci.* **33**, 681–697
35. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **7**, 971–980
36. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132
37. Zhang, J., Dong, B., and Siminovitch, K. A. (2009) Contributions of Wiskott-Aldrich syndrome family cytoskeletal regulatory adapters to immune regulation. *Immunol. Rev.* **232**, 175–194
38. Gomez, T. S., Hamann, M. J., McCarney, S., Savoy, D. N., Lubking, C. M., Heldebrandt, M. P., Labno, C. M., McKean, D. J., McNiven, M. A., Burkhardt, J. K., and Billadeau, D. D. (2005) Dynamin 2 regulates T cell activation by controlling actin polymerization at the immunological synapse. *Nat. Immunol.* **6**, 261–270
39. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11**, 10.1074/mcp.M111.013722
40. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J. Biol. Chem.* **275**, 16030–16036
41. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100