Protein Lysine methyltransferase G9a acts on non-histone targets

Philipp Rathert, Arunkumar Dhayalan, Marie Murakami, Xing Zhang, Raluca Tamas, Renata Jurkowska, Yasuhiko Komatsu, Yoichi Shinkai, Xiaodong Cheng, Albert Jeltsch

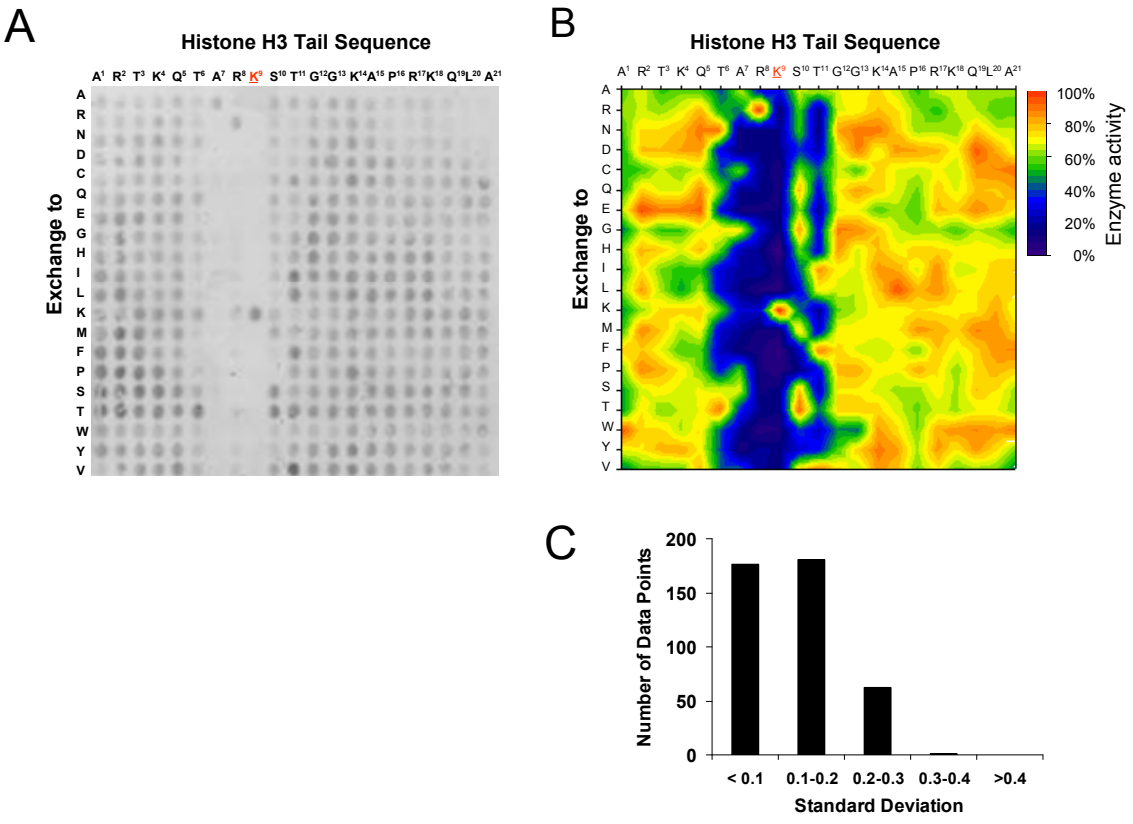
Supplementary information table of contents

Supplementary Figure 1: Specificity analysis of human G9a SET domain.

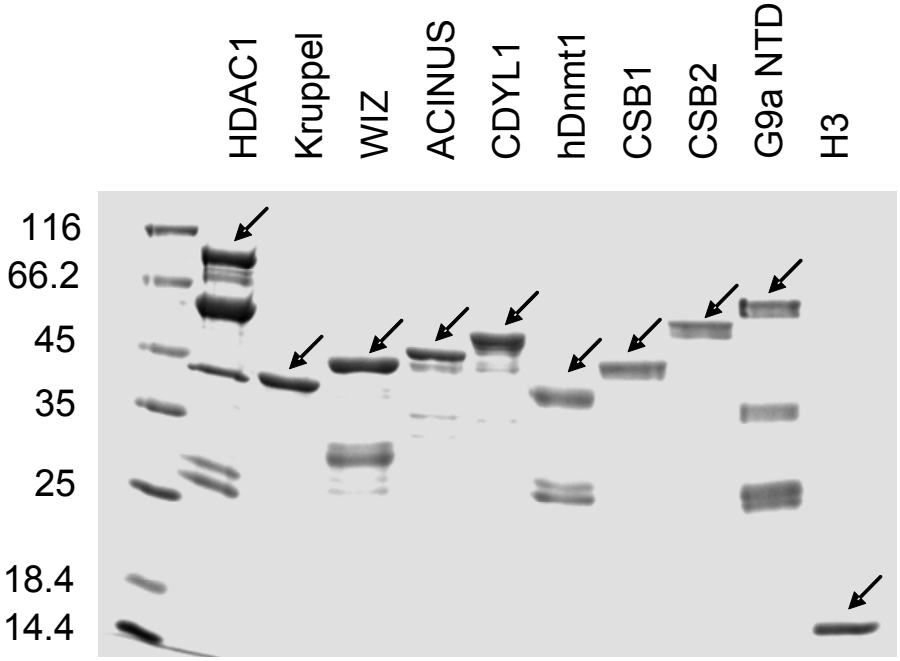
A) Example of one 20 x 21 peptide array methylated by G9a. The sequence of the H3 tail is given on the horizontal axis. Each residue was exchanged against all 20 natural amino acids (as indicated on the vertical axis) and the relative efficiency of methylation by G9a catalytic SET domain analyzed.

B) Average results obtained from three independent membrane arrays that were methylated with G9a after normalization.

C) Distribution of standard errors for the average methylation of each peptide using the three independent membranes as shown in Fig. 1B.

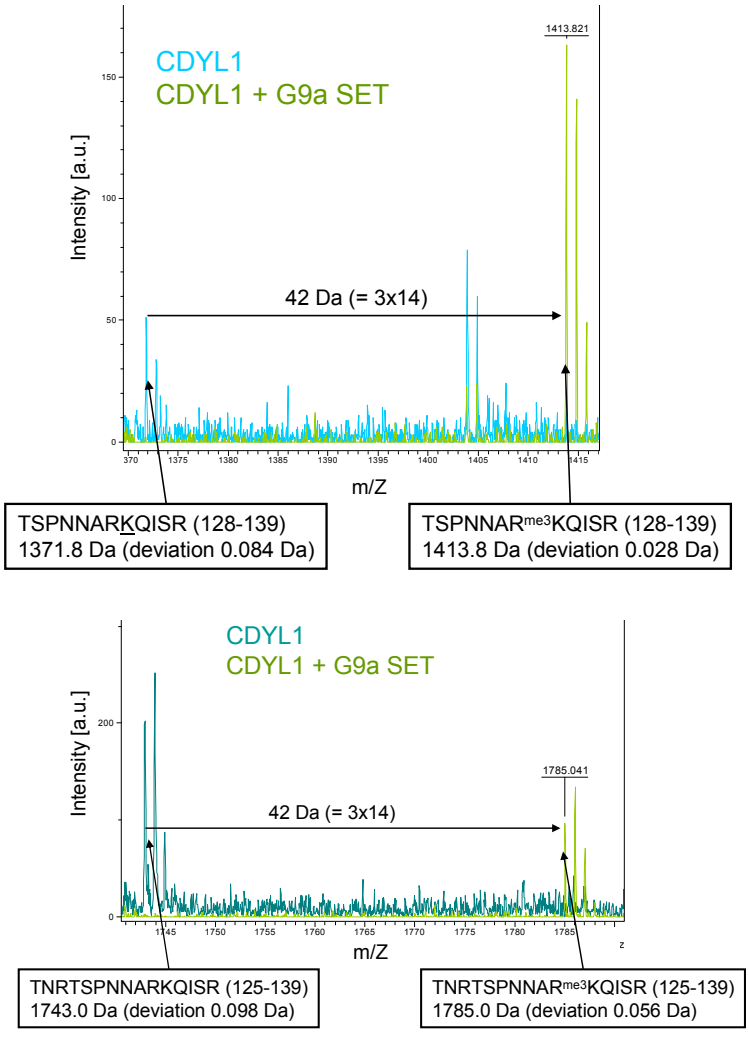


Supplementary Figure 2: Coomassie stained gel of purified GST-tagged G9a target protein domains and histone H3. The proteins shown here were used for the detection of protein methylation by transfer of radioactively labeled methyl-groups after incubation with G9a catalytic SET domain in the presence of radioactively labeled [methyl-3H]-AdoMet **1** (shown in Fig. 2A). N-terminal domain of G9a is abbreviated as G9a NTD.

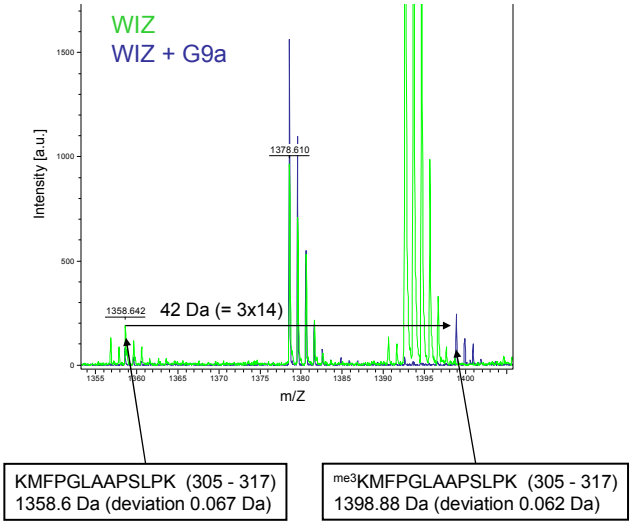


Supplementary Figure 3: Detection of protein methylation of non-histone G9a target protein domains by mass spectrometry after in vitro methylation with purified G9a catalytic SET domain. GST-tagged non-histone G9a target proteins were incubated for 5h with G9a catalytic SET domain in the presence of unlabelled AdoMet **2** and digested with different proteases as indicated.

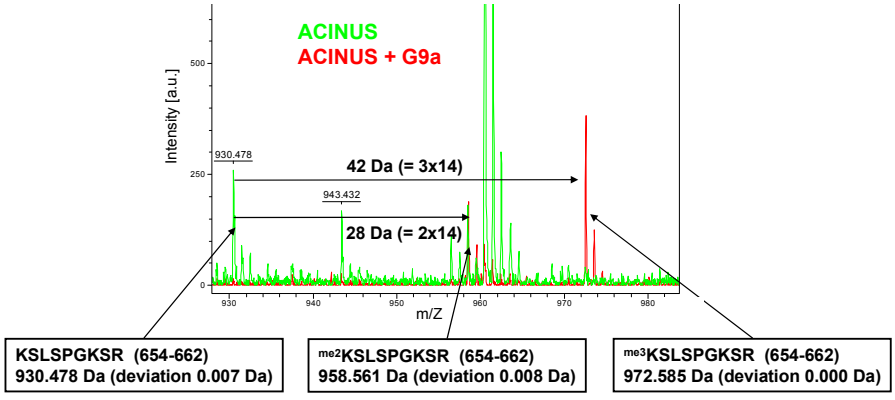
A) Detection of methylation of CDYL1 after ArgC digestion. Trimethylation of K135 in the fragment (residues 128-139) gave rise to the peak at 1413.8 Da with a mass shift of 42 Da (3x14 Da). Partial digestion with ArgC led to the fragment (residues 125-139) which is shifted from 1743.04 Da to 1785.0 Da by 42 Da.



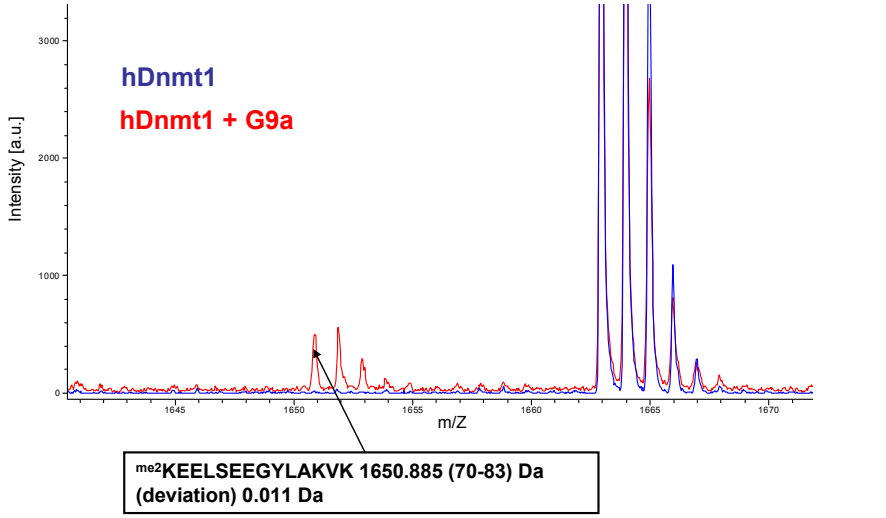
B) Detection of methylation of WIZ after tryptic digestion. Trimethylation of K305 in the fragment (residues 305-317) gave rise to the peak at 1398.9 Da. The additional peaks detected in this mass range correspond to WIZ fragments 1044-1054 (theoretical mass 1380.6 Da), 1127-1139 (theoretical mass 1392.6 Da) and G9a SET domain 937-948 (theoretical mass 1358.6 Da).



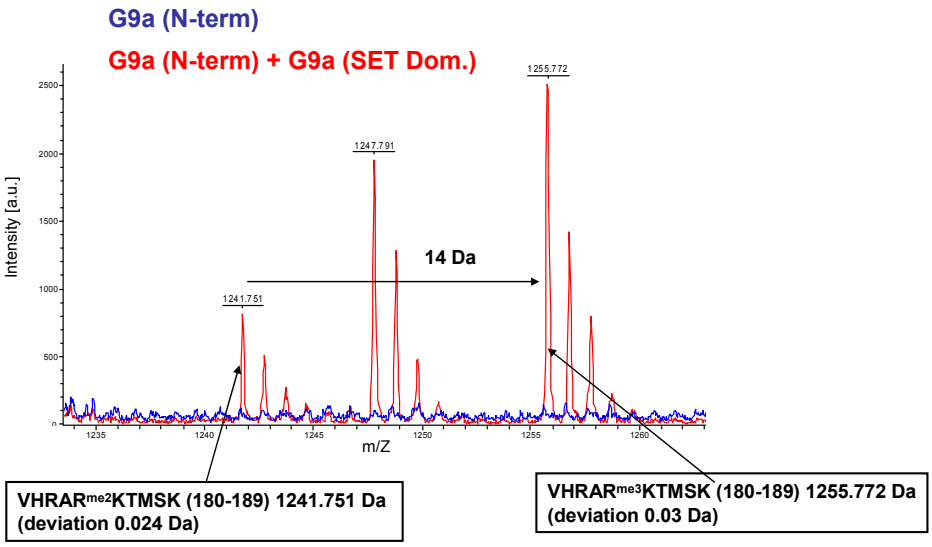
C) Detection of methylation of ACINUS after tryptic digestion. Dimethylation of K654 in the fragment (residues 654-662) gave rise to the peak at 958.6 Da and trimethylation led to a shift of the mass of the fragment to 972.6 Da. The additional peaks detected in this mass range correspond to ACINUS fragments 611-618 (theoretical mass 943.4 Da), 517-524 (theoretical mass 960.5 Da) and to GST fragment 79-87 (theoretical mass 956.4 Da).



D) Detection of methylation of human Dnmt1 after tryptic digestion. Dimethylation of K70 in the fragment (residues 70-83) gave rise to the peak at 1650.9 Da. The additional peak detected in this mass range corresponds to Dnmt1 fragment 10-25 (theoretical mass 1663.0 Da).

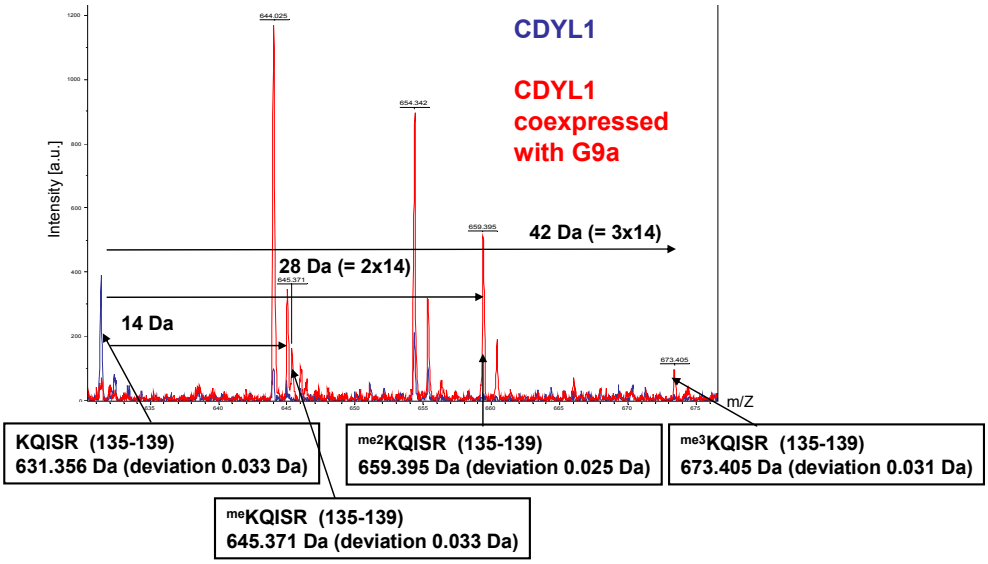


E) Detection of methylation of G9a in the N-terminal domain (G9a NTD) after tryptic digestion. Di- and tri-methylation of K185 in the fragment (residues 180-189) gave rise to the peak at 1241.8 Da and 1255.8 Da, respectively. The additional peak detected in this mass range corresponds to G9a SET domain fragment 829-838 (theoretical mass 1247.8 Da).

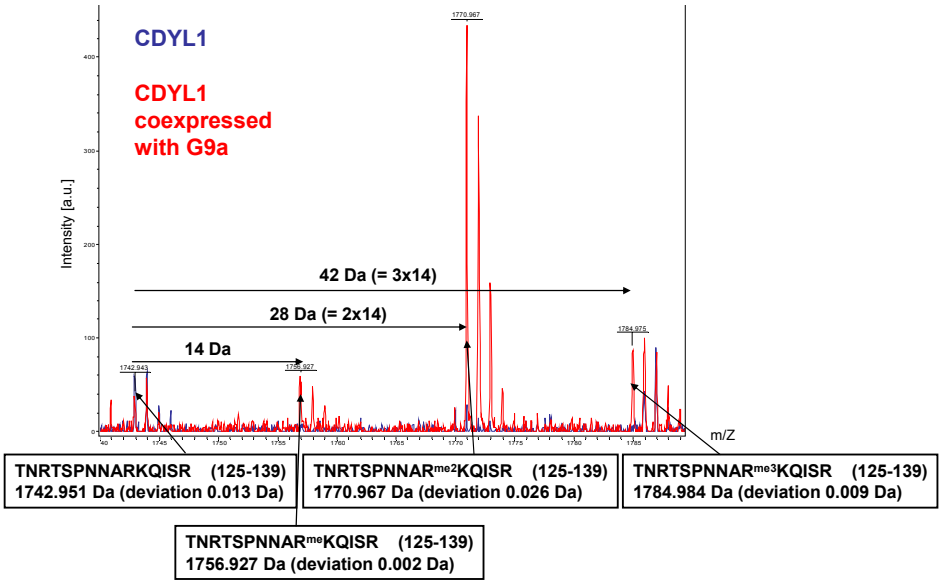


Supplementary Figure 4: Detection of protein methylation by mass spectrometry of CDYL1, WIZ and G9a N-terminal domain after co-expression with G9a SET domain in *E. coli*. GST-tagged non-histone G9a target proteins were purified and digested with different proteases as indicated.

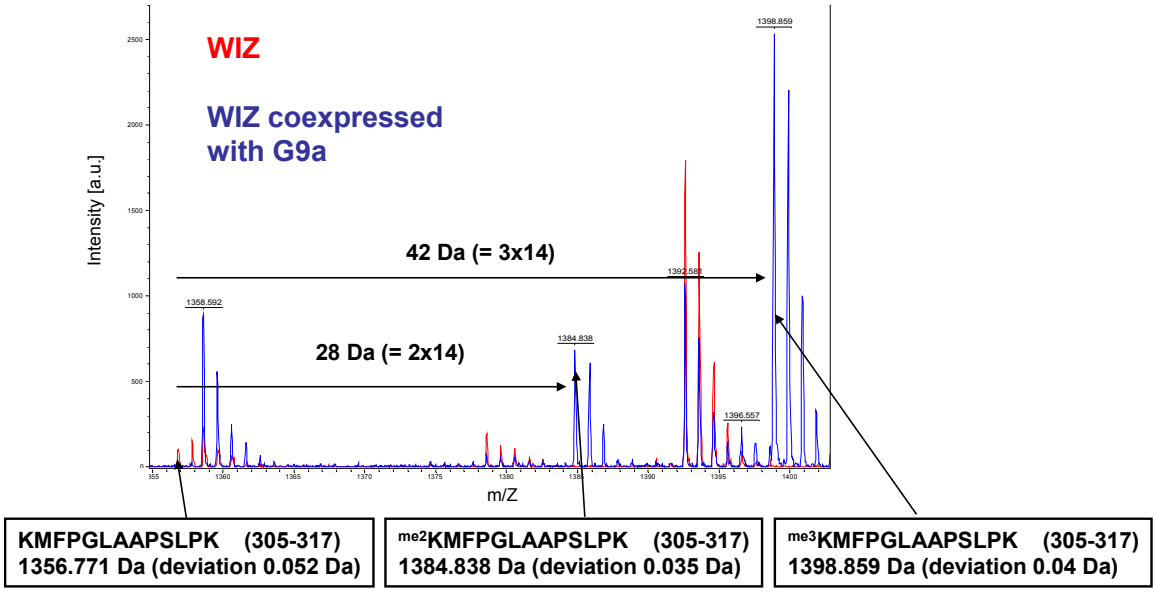
A) Detection of methylation in CDYL1 after complete digestion with ArgC protease. Methylation of K135 in the peptide fragment 135-139 (631.4 Da) gave rise to peaks at 645.4 Da (corresponding to monomethylation), 659.4 Da (corresponding to dimethylation) and 673.4 Da (corresponding to trimethylation). These peaks are absent in unmethylated CDYL1 preparations.



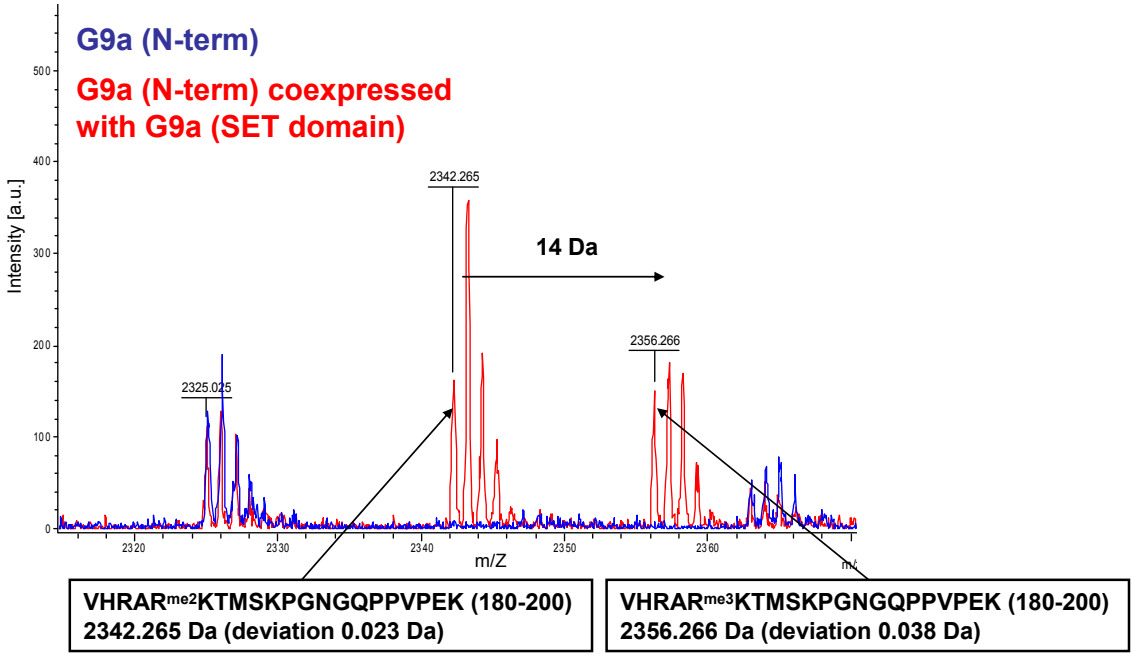
B) Detection of methylation in CDYL1 after incomplete digestion with ArgC protease. Methylation of K135 in the peptide fragment 125-139 (1743.0 Da) gave rise to peaks at 1757.0 Da (corresponding to monomethylation), 1771.0 Da (corresponding to dimethylation) and 1785.0 Da (corresponding to trimethylation). These peaks are absent in unmethylated CDYL1 preparations.



C) Detection of methylation in WIZ after tryptic digestion. Methylation of K305 in the 305-317 fragment (1356.8 Da) gave rise to peaks at 1384.8 Da (corresponding to dimethylation) and 1398.8 Da (corresponding to trimethylation). These peaks are absent in unmethylated WIZ preparations. Methylation of K305 (and not K317) in these fragments was confirmed by MALDI MS/MS analysis. The additional peaks detected in this mass range correspond to G9a SET domain fragment 937-948 (theoretical mass 1358.6 Da) and to WIZ fragment 323-335 (theoretical mass 1392.6 Da).



D) Detection of methylation of G9a in the N-terminal domain (G9a NTD) after tryptic digestion. Methylation of K185 in the fragment (residues 180-200) gave rise to peaks at 2342.3 Da (corresponding to dimethylation) and 2356.3 Da (corresponding to trimethylation). These peaks are absent in unmethylated preparations.



Supplementary Table 1: Identification of target proteins for G9a. In the upper part of the table results of methylation experiments with the histone peptides H3K9, H3K27, H1bK25 and H4K20 are displayed. In the lower part, the methylation of several new non-histone peptides is shown.

Target Protein	Acc. No.	Sequence	aa pos.	signal
Histone H3 K9	CAB02546	ARTKQTA**RK**STGGKAPRKQL	1-20	
Histone H3 K27	CAB02546	QLATKAA**RK**SAPATGGVKKP	19-38	
Histone H1b K25	P10412	EKTPVKKKA**RK**SAGAAKRKA	15-34	
Histone H4 K20	P62805	KGGAKRH**RK**VLRDNIQGITK	12-31	
WIZ	NP_067064	PTASPPPTA**RK**MFPGLAAPS	295-314	
CDYL1	Q9Y232	TNRTSPNNA**RK**QISRSTNSN	125-144	
CSB	Q03468	QAATSRDIN**RK**LDSVKRQKY	160-179	
		KKQGCNKRAA**RK**APAPVTPP	286-305	
		ASVGEGGGGG**RK**VGRYRDDG	437-456	
		QPAFGADHDVPK**RK**KFPASN	1041-1060	
G9a automethylation	Q96KQ7	GQPKVHRA**RK**TMSKPGNGQP	176-195	
Dnmt1	P26358	LCDLETKL**RK**EELSEEGYLA	61-80	
Kruppel	NP_009180	QRRSESPDS**RK**RRIHRCDFE	303-322	
ACINUS	Q9UKV3	RSRSASSNS**RK**SLSPGVSRD	644-663	
HDAC1	CAG46518	SDSEEEGEGG**RK**NSSNFKKA	421-440	

The Chromodomain Y-like protein 1 (CDYL1) is a direct interaction partner of G9a [1]. The Widely interspaced zinc finger motifs protein (WIZ) is part of a multiprotein complex containing G9a [2]. The Cockayne syndrome group B protein (CSB) interacts directly with G9a [3]. ACINUS, histone deacetylase-1 (HDAC1), DNA methyltransferase-1 (Dnmt1) and Kruppel-like factor-12 (Kruppel) are all known to be in complex with G9a [4-6].

Supplementary Discussion

Summary of literature data on methylation and demethylation of non-histone proteins at lysine residues.

SET7/9 methylates p53 and TAF10 [7,8]. Symd2 and Set8 methylate p53 [9,10] The *S. cerevisia* COMPASS complex methylates non-histone targets during chromosome segregation [11]. The Ezh2 complex methylates even cytosolic targets [12]. LSD1 demethylates p53 [13]. In addition, cytochrome C and Rubisco are methylated at lysine residues as well [14].

Examples of reading domains with preference for mono- or dimethyllysine residues.

Reading domains with preferential interaction with mono- or dimethylated lysine residues, which are candidates for an interaction with G9a methylated non-histone proteins, include L3MBTL1 and ankyrin repeats. L3MBTL1 has been shown to preferentially interact with mono- or dimethyllysine [15,16]. The ankyrin repeats of G9a and Ga9-like protein (GLP) are H3K9me1 and H3K9me2 binding modules [17]. Hence the ankyrin repeats of G9a and GLP bind the epigenetic mark that is generated by G9a and GLP catalytic SET domains *in vivo*. It is tempting to speculate that the dual roles of methylating and methyl-binding contribute to the co-regulator function of G9a-containing complexes.

Supplementary Methods

Cloning and protein purification

A fragment encoding the C-terminal 280 residues of human G9a (the catalytic SET domain) was made as described [18]. The sequence encoding human HDAC1, Kruppel-like factor-12 (residues 249-402), ACINUS (residues 513-697), CDYL1 (residues 64-283), WIZ (residues 179-359), Dnmt1 (residues 1-148), Cockayne syndrome group B domain 1 (CSB1) (residues 81-246) and domain 4 (CSB4) (residues 1024-1254) and G9a N-terminal fragment (residues 112-352) were amplified from cDNA derived from HEK293T cells and cloned into pGEX-6P-2 vector (GE Healthcare) using BamHI/XhoI sites, except WIZ was cloned using BamHI/SalI sites. For expression, *E. coli* BL21 cells (Novagen) carrying the corresponding plasmid (or plasmids when co-expression with the G9a SET domain) were grown in Luria-Bertani medium at 37 °C to OD600 of 0.6, then shifted to 22 °C for 20 min and induced overnight with 1 mM IPTG. The collected cells were resuspended in 20 mM HEPES (pH 7.5), 500 mM KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol and disrupted by sonication. The supernatants were passed through glutathione Sepharose 4B resin (Amersham Biosciences) and the bound proteins were eluted by 40 mM glutathione and dialyzed against 20 mM HEPES (pH 7.5), 200 mM KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol for 2 h then overnight against the same buffer with 60% glycerol. The murine HP1ß protein (accession No. NP_031648) was expressed and purified as described [19]. Histone H3 (H3.1) was purchased from New England Biolabs.

Synthesis of peptide SPOT arrays

Peptide arrays were synthesized using the SPOT synthesis method [20,21]. Each spot had diameters of 2 mm and contained approximately 9 nmol of peptide (Autospot Reference Handbook, Intavis AG). Successful synthesis of each peptide was confirmed by bromphenol blue staining of the membranes.

Methylation of peptide arrays

The membranes containing 420 peptide spots were washed for 20 min in methylation buffer containing 50 mM Tris/HCl, pH 9.0, 5 mM $MgCl_2$, 4 mM DTT and afterwards incubated at ambient temperature for 45 min in methylation buffer containing 20 nM G9a catalytic SET domain and 0.35 µM labeled [methyl-^{3}H]-AdoMet **1** (2.93×10^{15} Bq/mol) (NEN Life Sciences). The membranes were washed four times with 50 mM NH_4HCO_3, dried between whatman papers (Whatman GmbH, Dassel, Germany) and rinsed with Amplify NAMP100V solution (GE Healthcare, Munich, Germany). The membranes were incubated on Hyperfilm^{TM} high performance autoradiography films (GE Healthcare, Munich, Germany) in the dark for 3-7 days. The films were developed using AGFA Curix 60 developing machine (Agfa Deutschland Vertriebsgesellschaft mbH & Co. KG, Cologne, Germany). All methylation experiments on peptide arrays were carried out at least in triplicate.

Data analysis by a discrimination factor

Using Dim-5 as an example, we showed that the kinetics of peptide methylation on arrays were linear with respect to time and enzyme concentration [22]. Incubation times and G9a concentration were chosen not to have full turnover such that the amount of methylation of each spot reflects the rate of the enzyme with this particular substrate. One inherent advantage of the method is that all peptide spots are methylated in competition which ensures that equal amounts of active enzyme and cofactor are available for all substrates. Under the experimental conditions used here, the relative rates of methylation correspond to ratios of k_{cat}/K_d values for the respective peptide, the latter representing an established parameter for quantification of enzyme specificity [23].

To compare the accuracy of recognition of each residue in the substrate quantitatively, the relative contribution of each amino acid i at position x for peptide recognition was calculated by a discrimination factor D:

$$D = \frac{\overline{v_{j \neq i}}}{v_i} - 1$$

where v_i is the rate of modification of peptide carrying amino acid i and $\overline{v_{j \neq i}}$ is the average rate of methylation of all 19 peptides carrying a different amino acid $j \neq i$ at position x (including the wild type sequence). For example in Fig. 1D, the discrimination factor of 6 for a glycine at position 7 indicates that the peptide with glycine at that position is methylated 6 times faster than the average of all peptides carrying any of the other amino acids at this site. Since the detection limit of the experiments was at about 3% of the full activity, the lower limit of the discrimination factors for R8 and K9, which both could not be replaced by any other residue, was 30.

Identification of potential non-histone targets

A Scansite search [24] with the G9a substrate specificity profile [NTGS] [GCS] [R] [K] [TGQSVMA] [FVILA] resulted in 92 human proteins containing potential target sites. Eighteen of them that have known nuclear localization were selected for further analysis. In addition, we retrieved proteins known to interact with G9a from Human Protein Reference Database [25] and literature and scanned for RK sequences.

Methylation of purified domains

Protein domain methylation was performed in methylation buffer supplemented with either 0.35 µM labeled [methyl-^{3}H]-AdoMet **1** or with 1 mM unlabeled AdoMet **2** for MALDI experiments (see below). Target protein (2 µM) was incubated with 200 nM G9a catalytic SET domain for 5 h at 21 °C. An aliquot of 10 µl of the reaction mixture were loaded on SDS PAGE, separated and methylation detected by autoradiography.

MALDI analysis

Methylated samples were dialyzed against 100 mM NH_4HCO_3, pH 8.5 and digested with 0.04 mg Trypsin or ArgC (Sigma) at 37 ˚C for 8 h. MPT Anchor-ChipTM target plate (Bruker Daltonic GmbH, Bremen, Germany) was pre-spotted with a-cyano-4-hydroxycinnamic acid (HCCA) matrix and 2 µL of different dilutions of the digests in TA (mixture of 30% Acetonitrile and 70% of a solution of 0.1 % trifluoracetic acid in water) were added. The sample mixtures were removed after 30 s and the spots were washed with 10 mM NH_4HCO_3 and air dried. A Bruker Autoflex II machine (Bruker Daltonic GmbH, Bremen, Germany) was used and calibrated with the Bruker Daltonic peptide calibration standard II.

Reading domain binding to peptides arrays containing modified residues

After synthesis of the array, the cellulose membrane was blocked by immersing in TTBS buffer (10 mM Tris/HCl pH 8.3, 0.05% Tween-20 and 150 mM NaCl) containing 5% non-fat dried milk at 4 °C overnight. After washing in TTBS buffer, the membrane was incubated with purified GST-tagged HP1ß or CDYL1 (10 nM) at room temperature for 1 h in buffer containing 100 mM KCl, 20 mM HEPES pH 7.5, 1 mM EDTA, 0.1 mM DTT and 10% glycerol. After washing in TTBS buffer, the membrane was incubated with goat anti-GST antibody (GE Healthcare #27-4577-01, 1:10000 dilution) in TTBS buffer for 1 h at room temperature. After additional washing with TTBS, the membrane was incubated with horseradish peroxidase conjugated anti-Goat antibody (Invitrogen #81-1620 1:12000) in TTBS for 1 h at room temperature. Finally, the membrane was submerged in ECL developing solution (Amersham Pharmacia Biotech, Little Chalfont, Bucks., U.K.) and autoradiographed.

Detection of CDYL1 and WIZ methylation in vivo

Full length mouse G9a and flag-tagged full-length mouse Wiz or eGFP-NLS-human CDYL1 (residues 64-283) expression vectors (pCAGGS-G9aS and mG9aS, pCAGGS-lox-flag-WizmWiz or pEGFP-CDYL1) were transfected into HEK293T cells using a TransIT-LT1 lipofection reagent (Mirus). After 40 h

transfection, total nuclear lysates were prepared using high salt extraction buffer containing 420 mM NaCl, 0.1% NP-40, 20 mM HEPES pH 7.5, and 1.5 mM $MgCl_2$. Flag-tagged Wiz or eGFP-CDYL1 was immunoprecipitated by anti-Flag antibody (M2, Sigma) or anti-GFP antibody (Roche), respectively, and subjected to Western blot analysis. To detect lysine methylation and G9a, anti-pan methyl-lysine antibodies (MEK3D7 and MEK5B11, [26]) and anti-G9a (A8620A, Perseus Proteomics, Japan) were used.

Supplementary References

1. Shi, Y. et al. *Nature* **422**, 735-8 (2003).
2. Ueda, J., Tachibana, M., Ikura, T. & Shinkai, Y. *J. Biol. Chem.* **281**, 20120-8 (2006).
3. Yuan, X., Feng, W., Imhof, A., Grummt, I. & Zhou, Y. *Mol. Cell* **27**, 585-95 (2007).
4. Duan, Z., Zarebski, A., Montoya-Durango, D., Grimes, H.L. & Horwitz, M. *Mol. Cell. Biol.* **25**, 10338-51 (2005).
5. Esteve, P.O. et al. *Genes Dev.* **20**, 3089-103 (2006).
6. Rual, J.F. et al. *Nature* **437**, 1173-8 (2005).
7. Chuikov, S. et al. *Nature* **432**, 353-60 (2004).
8. Kouskouti, A., Scheer, E., Staub, A., Tora, L. & Talianidis, I. *Mol. Cell* **14**, 175-82 (2004).
9. Huang, J. et al. *Nature* **444**, 629-32 (2006).
10. Shi, X. et al. *Mol. Cell* **27**, 636-46 (2007).
11. Martin, C. & Zhang, Y. *Nat. Rev. Mol. Cell. Biol.* **6**, 838-49 (2005).
12. Su, I.H. et al. *Cell* **121**, 425-36 (2005).
13. Huang, J. et al. *Nature* **449**, 105-8 (2007).
14. Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L. & Hurley, J.H. *Cell* **111**, 91-103 (2002).
15. Li, B., Carey, M. & Workman, J.L. *Cell* **128**, 707-19 (2007).
16. Min, J. et al. *Nat. Struct. Mol. Biol.* **14**, 1229-30 (2007).
17. Collins, R.E. et al. *Nature Structural and Molecular Biology* (2008).
18. Collins, R.E. et al. *J. Biol. Chem.* **280**, 5563-70 (2005).
19. Botuyan, M.V. et al. *Cell* **127**, 1361-73 (2006).
20. Wenschuh, H. et al. *Biopolymers* **55**, 188-206 (2000).
21. Frank, R. *J. Immunol. Methods* **267**, 13-26 (2002).
22. Rathert, P., Zhang, X., Freund, C., Cheng, X. & Jeltsch, A. *Chem. Biol.* **15**, 5-11 (2008).

23. Fersht, A. *Structure and Mechanism in Protein Science*, (W.H. Freeman and Company, New York, 1998).
24. Obenauer, J.C., Cantley, L.C. & Yaffe, M.B. *Nucleic Acids Res.* **31**, 3635-41 (2003).
25. Mishra, G.R. et al. *Nucleic Acids Res.* **34**, D411-4 (2006).
26. Iwabata, H., Yoshida, M. & Komatsu, Y. *Proteomics* **5**, 4653-64 (2005).