

VIP Very Important Paper



Analysis of the Substrate Specificity of the SMYD2 Protein Lysine Methyltransferase and Discovery of Novel Non-Histone Substrates

Sara Weirich,^[a] Maren Kirstin Schuhmacher,^[a] Srikanth Kudithipudi,^[a] Cristiana Lungu,^[a] Andrew D. Ferguson,^[b] and Albert Jeltsch^{*,[a]}

The SMYD2 protein lysine methyltransferase methylates various histone and non-histone proteins and is overexpressed in several cancers. Using peptide arrays, we investigated the substrate specificity of the enzyme, revealing a recognition of leucine (or weaker phenylalanine) at the -1 peptide site and disfavor of acidic residues at the +1 to +3 sites. Using this motif, novel SMYD2 peptide substrates were identified, leading to the discovery of 32 novel peptide substrates with a validated target site. Among them, 19 were previously reported to be methylated at the target lysine in human cells, strongly suggesting that SMYD2 is the protein lysine methyltransferase re-

sponsible for this activity. Methylation of some of the novel peptide substrates was tested at the protein level, leading to the identification of 14 novel protein substrates of SMYD2, six of which were more strongly methylated than p53, the best SMYD2 substrate described so far. The novel SMYD2 substrate proteins are involved in diverse biological processes such as chromatin regulation, transcription, and intracellular signaling. The results of our study provide a fundament for future investigations into the role of this important enzyme in normal development and cancer.

Introduction

In recent years, it has been discovered that lysine methylation, which was initially identified on histone proteins, occurs on a wide range of non-histone proteins, where it plays essential regulatory roles in various cellular processes.^[1-4] Lysine residues can be mono-, di- and trimethylated and the biological outcome of the modification depends on the site and degree of methylation in target proteins. Lysine methylation is introduced by protein lysine methyltransferases (PKMTs)^[5,6] including enzymes of the SET and MYND-containing protein (SMYD) family.^[6-8] The biological function of SMYD PKMTs is diverse including gene regulation, chromatin remodeling, transcription, signal transduction, cell cycle control, and DNA damage response.^[6-8] The SMYD PKMT family consists of five members,

called SMYD1 to 5. They share their structural domain architecture with other SET domain PKMTs, where a catalytically active SET domain containing a SET-I insertion involved in target peptide interaction is followed by a Post-SET domain. Different from other SET domain PKMTs, the SET domain of SMYD family enzymes is further split by the insertion of a MYND (Myeloid-Nerve-DEAF-1) domain between the initial part of the SET domain (the S-sequence) and the SET-I part. The MYND domain is a zinc-finger domain that mediates protein-protein interactions. Different from other SET-domain PKMTs, the SMYD family enzymes do not contain a Pre-SET domain, but they carry an additional SMYD family specific C-terminal domain (CTD).

The SMYD2 enzyme (also called KMT3C) was initially discovered as an H3K36 dimethyltransferase and SMYD2 activity had been shown to lead to repression of transcribed genes.^[9] Later it was found that SMYD2 also methylates H3K4 in the presence of the HSP90 protein.^[10] In addition to the methylation of histone targets, SMYD2 was reported to methylate several non-histone proteins like p53^[11] and the retinoblastoma protein (RB).^[12] TP53 is one of the most studied tumor suppressor genes and it has important biological functions in the regulation of cell cycle arrest, apoptosis and the DNA damage response. Four lysine residues in the C-terminal part of p53 can be methylated by different PKMTs leading to distinct biological outputs.^[13] Methylation studies with p53 lysine to arginine mutants revealed that K370 is the main methylation site of SMYD2 on p53.^[11] Methylation of K370 results in repression of p53 regulated transcription and p53 mediated apoptosis in cancer cells due to decreased occupancy of methylated p53 at

[a] Dr. S. Weirich, Dr. M. K. Schuhmacher, Dr. S. Kudithipudi, Dr. C. Lungu, Prof. Dr. A. Jeltsch
Department of Biochemistry
Institute of Biochemistry and Technical Biochemistry
University of Stuttgart, Allmandring 31, 70569 Stuttgart (Germany)
E-mail: albert.jeltsch@ibt.uni-stuttgart.de

[b] Dr. A. D. Ferguson
AstraZeneca R&D Boston
35 Gatehouse Drive, Waltham, MA 02451 (USA)

Supporting information and the ORCID identification numbers for the authors of this article can be found under <https://doi.org/10.1002/cbic.201900582>.

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

This article is part of a Special Issue commemorating the 20th Anniversary of ChemBioChem. To view the complete issue, visit our homepage

its target genes.^[11,12] Different studies reported that methylation of p53 by SMYD2 was higher than methylation of the other identified histone targets.^[14–16] The retinoblastoma tumor suppressor protein (RB) involved in the regulation of cell cycle progression and apoptosis is another non-histone substrate of SMYD2 with a strong cancer connection. SMYD2 methylation of RB occurs at lysine 860, leading to the progression of the cell cycle via different molecular pathways.^[17] Moreover, SMYD2 also methylates the estrogen receptor alpha (ER α) protein at lysine 266 repressing ER α transactivation activity.^[18] Details of the substrate peptide recognition of SMYD2 have been identified by two structures of the enzyme with the p53 and the ER α peptide substrates.^[14,19] In addition, several other non-histone targets were identified in the last couple of years, including PARP1,^[20] MAPKAPK3^[21] and PTEN.^[22] SMYD2 is essential for normal organismal development^[6–8] and dysregulation of SMYD2 was found in cardiovascular disease and cancer.^[6–8] For example, SMYD2 is significantly overexpressed in many cancers including esophageal squamous cell carcinoma (ESCC), bladder and gastric cancer,^[23,24] and in triple negative breast cancer it was shown to have a tumor promoting effect.^[8] In some cases, the molecular mechanisms connecting the methylation of SMYD2 non-histone substrates and oncogenic effects have been discovered and, based on these, inhibitors of SMYD2 have been tested as therapeutic agents.^[24]

Better understanding of the biology of the SMYD2 protein lysine methylation signaling network critically depends on the identification of the substrate profile of SMYD2, but in spite of progress in this direction, the identification of PKMT substrates is not trivial.^[25] General proteome-wide screenings were undertaken aiming to identify SMYD2 substrates in an unbiased way. By combining proteomic lysine methylation data with SMYD2 knock-down, 34 SMYD2 dependent methylation events were discovered in human cell lines, two of which were validated by in vitro methylation experiments (AHNAK and AHNAK2).^[26] In an approach combining enzyme specificity and modelling, four novel SMYD2 substrates (SIX1, SIX2, SIN3B and DHX15) were discovered and validated.^[27] Another bioinformatic approach combining different data sources like known substrates, proteomic data, protein interaction, gene ontology and structural data has led to the identification of six novel substrates (MAPT, CCAR2, EEF2, NCOA3, STUB1 and UTP14A).^[28] In the current study, we also applied an unbiased search for novel SMYD2 substrates. In our workflow, the substrate specificity of SMYD2 is analyzed by methylation of peptide arrays to identify additional non-histone targets in a systematic manner. Different studies already showed that the analysis of the specificity profiles of PKMTs is a powerful method to investigate their substrate recognition and based on this discover novel substrates.^[25,29–35] With this approach, we identified 32 novel peptide substrates of SMYD2 with validated methylation sites. Among them, 14 novel non-histone protein substrates of SMYD2 were discovered, six of which were more strongly methylated than p53, the best SMYD2 substrate described so far. The novel SMYD2 substrates are involved in diverse cellular processes like chromatin interaction and modification, gene regulation, DNA repair, actin filament dynamics, and Ras signaling.

Results and Discussion

As described above, it is a big challenge to link protein lysine methyltransferases with lysine methylation events of specific substrates. In this study, we wanted to analyze the substrate sequence specificity of SMYD2 in detail using a peptide array approach. Afterward, we used the specificity information to search for and validate additional non-histone substrates of SMYD2 that will give deeper insight into the wide substrate spectrum of SMYD2 and the connected biological functions.

Protein expression and catalytic activity of SMYD2 on p53-K370

Full-length SMYD2 and p53 were expressed as GST-fusion proteins in *Escherichia coli* cells and purified by affinity chromatography (Figure 1A). The circular dichroism (CD) spectrum of the purified SMYD2 showed minima around 208 and 220 nm characteristic for a folded protein (Figure S1A in the Supporting Information) and the obtained melting temperature of 55.9 °C (Figure S1B) documented stable folding. SMYD2 activity was tested using peptide SPOT arrays probing several well-studied histone methylation marks together with p53-K370. To this end, 20-residue peptides were synthesized on a cellulose membrane using the SPOT technology.^[36,37] The resulting SPOT peptide arrays were incubated with SMYD2 using radioactively labelled AdoMet as cofactor and the transfer of radioactively labelled methyl groups to p53 was detected by autoradiography (Figure 1B). Note that the different peptide sequences could yield slightly different background signal levels due to unspecific binding of AdoMet. This experiment confirmed the methylation of H3K36 as shown by the clear methylation signal of the H3 28–48 peptide, which was completely lost after the K36A exchange. In the case of the H3 1–20 peptide, a methylation signal was observed, which was decreased after the H3K4 exchange but not affected by the K9A exchange, indicating methylation of H3K4. However, the strongest methylation was observed on the p53 peptide, which was almost completely lost after the K370A exchange demonstrating a strong and specific methylation of p53-K370 by SMYD2. The methylation activity of SMYD2 was next tested on the p53 protein using radioactively labeled AdoMet. The methylated reaction mixture was separated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the transfer of radioactively labelled methyl groups to p53 was detected by autoradiography (Figure 1C). The result revealed a strong methylation of p53, whereas no signal could be detected in the absence of p53.

Substrate specificity analysis of SMYD2

For specificity analysis, larger SPOT peptide libraries were synthesized by using the p53 sequence (363–377) as template, which was selected based on the strong methylation signal observed in Figure 1B. The peptide arrays consisted of 300 individual peptides in each of which one amino acid of the original sequence was exchanged against one of the 20 natural

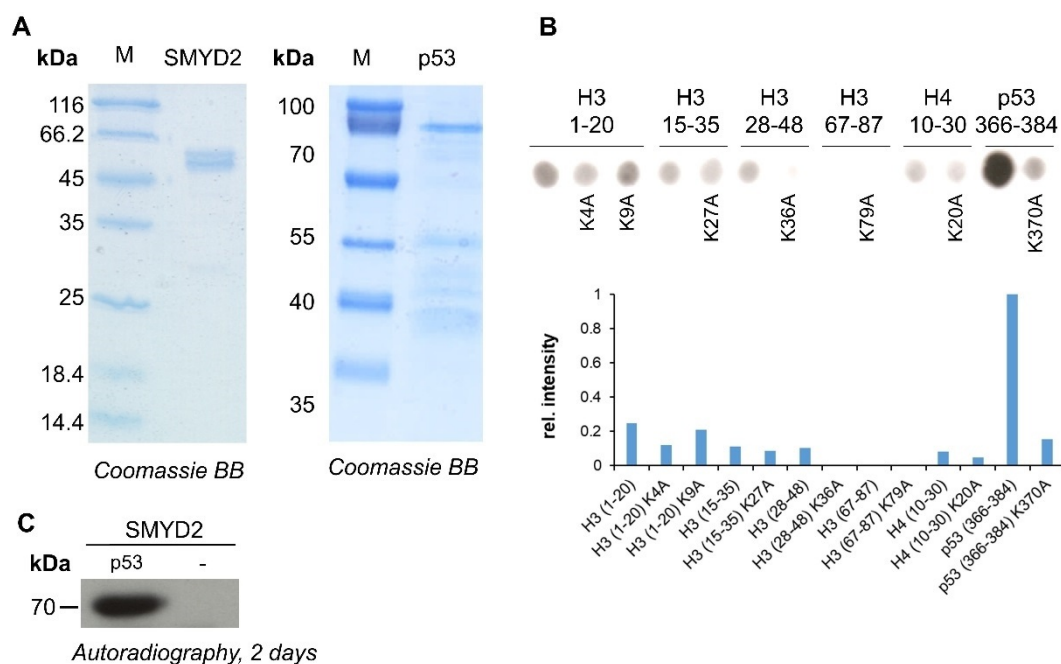


Figure 1. Protein purification and activity validation of SMYD2. A) Full-length SMYD2 and p53 were cloned as GST-fusion proteins and purified by affinity chromatography. B) Methylation of peptide SPOT arrays containing the previously identified methylation sites of SMYD2 on H3, some additional well studied histone methylation sites and the p53 methylation substrate. K-to-A peptides were included to identify the lysine residues that were methylated in the corresponding wild-type peptide. The lower part shows a quantification of the spot intensities. C) Methylation of the purified p53 protein with radioactively labeled AdoMet. Methylated samples were separated by SDS-PAGE and the methyl group transfer was detected by autoradiography. After two days of exposure methylation of p53 could be detected, whereas in absence of p53 no methylation was observed.

amino acids generating all possible single amino acid exchanges of the template sequence. With this approach, the influence of each single amino acid at every position of the substrate peptide on the enzymatic activity of SMYD2 can be investigated in detail. The peptide arrays were incubated with methylation buffer containing SMYD2 and radioactively labeled AdoMet as cofactor. The transfer of methyl groups to the immobilized peptides was detected by autoradiography (Figure 2A). The experiment was conducted two times. Afterward, each array was quantified and the results of the two independent experiments were normalized and averaged (Figure 2B). Based on these data, the standard deviations (SD) were calculated for the methylation activity on each single spot to evaluate the quality of the peptide array methylation experiments.

As shown in Figure 2C, the SPOT peptide array methylation data were highly reproducible, because 87% of the peptides showed an SD smaller than 20%, around 97% of the peptides showed an SD smaller than 30% and only seven peptides had an SD larger than 30% (Figure 2C). Our data show that SMYD2 is specific toward a leucine at position -1 of the substrate peptide sequence (considering the target lysine as 0). It only weakly tolerates phenylalanine at the same site. SMYD2 did not exhibit specificity for amino acids N-terminal to position -1 . At the $+1$ to $+3$ positions, polar uncharged and basic residues are preferred, but acidic residues (aspartate and glutamate), cysteine and large hydrophobic as well as aromatic residues were not tolerated to a different degree leading to the following specificity profile: [LF]-K-[ARNGHLKFSTV]-[ARNGKST]-[ARNGKSTV].

Several of the known non-histone targets, including p53 (SHLK³⁷⁰SKK, the template sequence of the peptide array), ER α (RMLK²⁶⁶HKR), Rb (RVLK⁸⁶⁰RSA), PARP1 (LTLK⁵²⁸GGA) and MAPKAPK3 (KDLK³⁵⁵TSN) fit to the derived sequence motif, as well as the K36 methylation site (GGVK³⁶KPH) on histone H3. The preference of SMYD2 for leucine and phenylalanine at the -1 site of the target peptide can be explained by the structure of the enzyme in complex with the p53 (PDB ID: 3S7F)^[14] or ER α peptides (PDB ID: 4O6F).^[19] In both structures, the corresponding target peptide leucine residue at the -1 site is in close proximity to methyl groups of SMYD2 residues T105, L108, V179, and T185 and further contacted by the C α of N180 and G183 and C β of S196, altogether creating a hydrophobic pocket large enough to accommodate a leucine residue (Figure 3A). The disfavor for acidic residues at the $+1$ to $+3$ sites in the target peptide can be explained by the close proximity of two acidic side chains from SMYD2 (E187 and D242) and the general acidic nature of the peptide binding pocket (Figure 3B). Overall SMYD2 showed a relatively weak peptide sequence specificity with surprisingly little sequence specific readout. This finding suggests that the methylation specificity of this enzyme very likely is also controlled by substrate binding specificity at the protein level.

Methylation of non-histone peptide substrates by SMYD2

Since we were interested to discover novel biological relevant substrates of SMYD2, the newly identified sequence motif was used to screen for potential novel non-histone targets in the

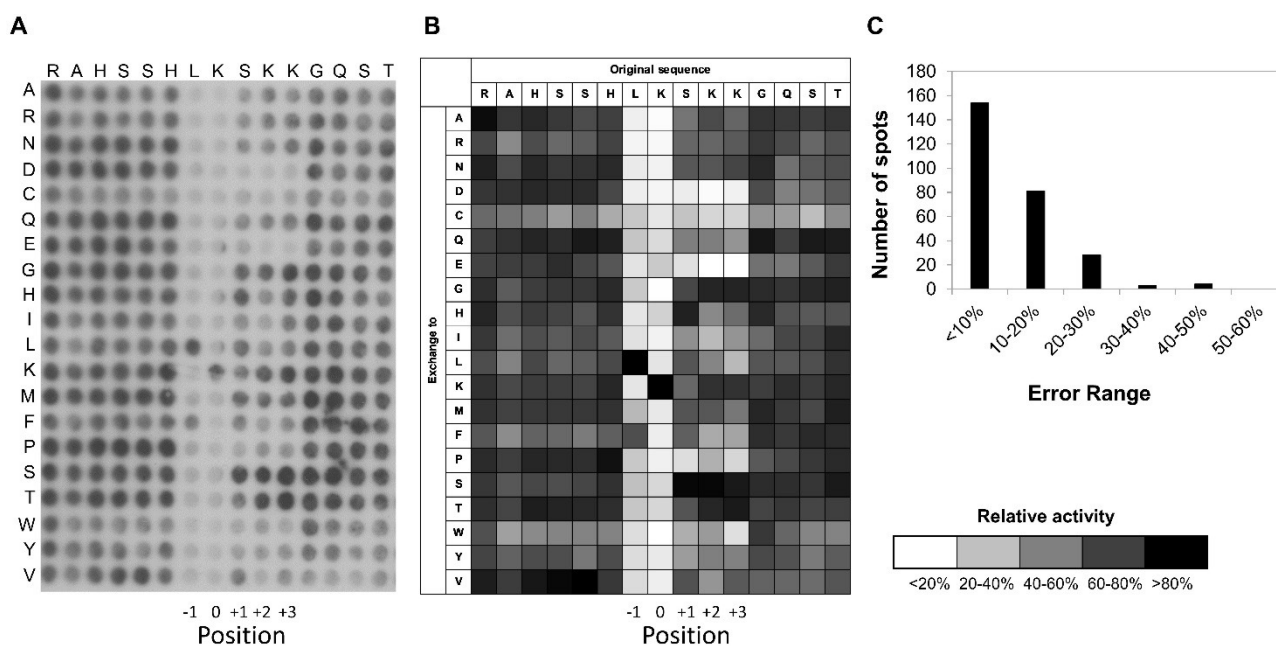


Figure 2. Substrate sequence specificity analysis of SMYD2. A) Example of a substrate sequence specificity SPOT array methylated by SMYD2. Arrays of 15-residue peptides were synthesized by using p53 (363–377) as template sequence, represented in the horizontal axis. Each residue was exchanged against all 20 natural amino acid residues, shown in the vertical axis. For methylation, the membrane was incubated with SMYD2 in the presence of radioactively labelled AdoMet and the transfer of methyl groups visualized by autoradiography. The exposition time of the film was three days. B) Two independent peptide array methylation experiments were performed and the data were normalized, averaged and the methylation signals presented as greyscale heatmap. Black color represents strong methylation, while light grey represents weak methylation. C) Distribution of the standard deviations of SMYD2 activity on all peptides tested in two repetitions.

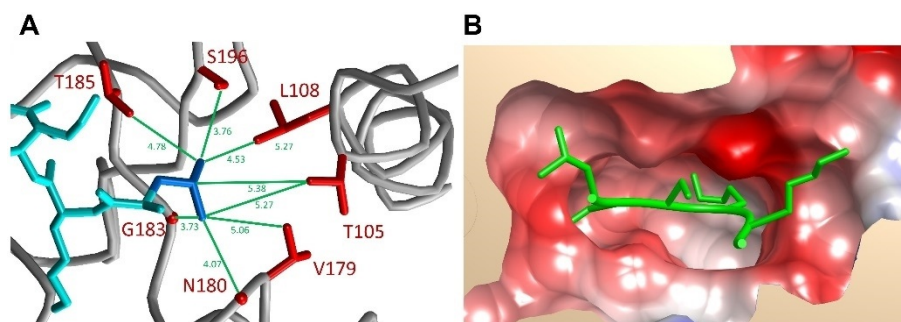


Figure 3. Structural details of the SMYD2-p53 complex (3SF7).^[14] A) Residues in the vicinity of the Leu at the –1 position. The peptide is shown in cyan with the side chain of the Leu residue in blue. SMYD2 is shown in ribbon view (grey) with the interacting residues in red. Distances are indicated in Å. B) Electrostatic surface view of SMYD2 illustrating the dominance of acidic residues in the binding pocket. The bound peptide is shown in green.

human proteome using Scansite (<https://scansite4.mit.edu/>).^[38] Among the almost 8000 hits, 124 proteins with nuclear localization and important biological functions were selected for further analysis (Table S1). To check if SMYD2 can methylate these predicted targets, 15-residue peptides were synthesized on a cellulose membrane using the SPOT synthesis method. The p53 peptide and an artificial peptide with an optimized SMYD2 target sequence were included as positive controls and their corresponding K-to-A mutants were used as negative controls. To compare the SMYD2 activity with the activity at the histone methylation sites, H3K4, H3K36 and the corresponding K-to-A mutant peptides were included on the array as well. Figure 4A shows that SMYD2 can methylate about 40 peptides with same or higher activity than p53 indicating that about $\frac{1}{3}$ of

the predicted targets were methylated. The other peptides were either weakly or not methylated at all. Out of these 40 substrates, 14 were selected for further investigation based on their high methylation intensity.

To further develop the detection of possible novel substrates, the PhosphoSite Plus database (<https://www.phosphosite.org/>)^[39] was used, which allows to identify proteins with known methylation within the broadened SMYD2 sequence motif [LF]-[K]. With this approach, 155 methylation sites in non-histone proteins were identified which are methylated in cells at potential SMYD2 target sites (Table S2). All potential targets were synthesized as 15 amino acid long peptides with the target lysine in the center on a SPOT peptide array and treated as explained previously (Figure 4B). As before, this

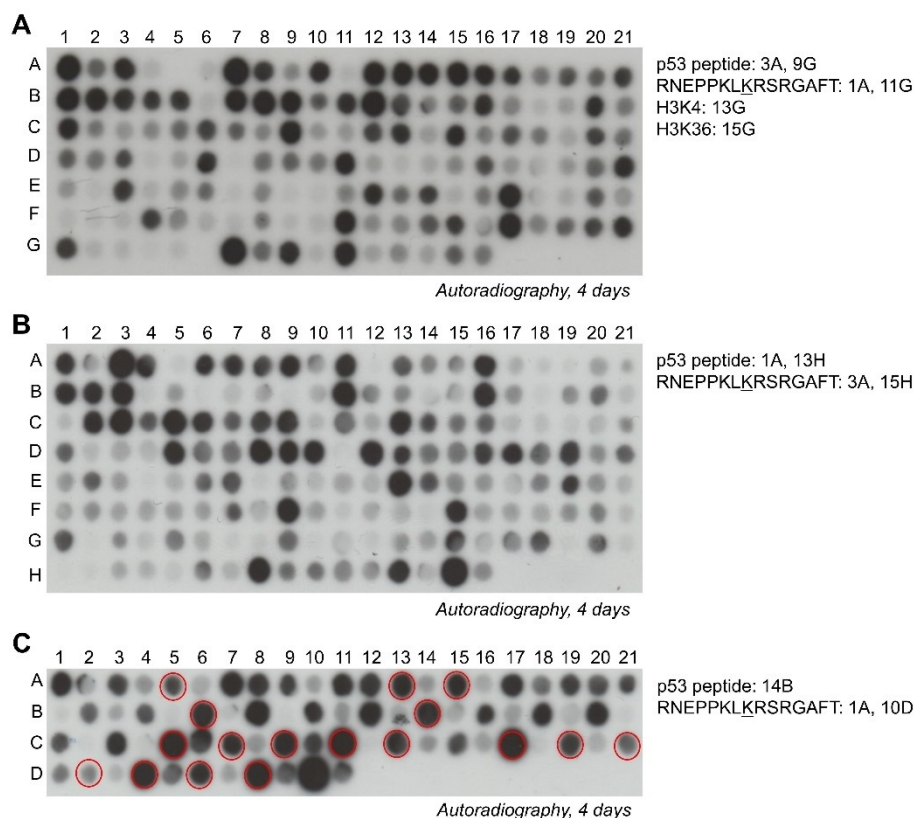


Figure 4. Screening of SMYD2 non-histone peptide substrates. A) SPOT array methylation of candidate SMYD2 targets taken from the Scansite search database (Table S1). B) SPOT array methylation of candidate SMYD2 targets taken from the PhosphoSite Plus database (Table S2). C) Selected peptides found to be methylated in panel A and B were synthesized on an additional SPOT peptide array together with their corresponding K-to-A mutants (Table 1). On each array, the artificial peptide and p53 peptide were included as positive controls. As negative controls, the corresponding K-to-A mutants of the artificial peptide and p53 peptide were synthesized next to the corresponding wild-type spots. Spots marked with red circles in (C) were selected for further investigation of protein methylation.

array contained the sequence of p53 and the artificial peptide together with their K-to-A mutants as controls. Out of the 155 potential non-histone targets, 19 peptides were methylated equally strong as p53 or even stronger and were selected for further investigation. Interestingly, the more global search applied here, which did not include the information about the +1 to +3 sites, led to a significantly lower hit rate of only 12% as compared to 33% found in the screen based on the full motif.

The 33 selected non-histone targets from both screens were synthesized on an additional peptide SPOT array together with their K-to-A mutants to confirm that the predicted target lysine is methylated by SMYD2 (Figure 4C). Strikingly, for 32 out of the 33 tested peptides a complete loss or strong decrease in the methylation signal was detected with the alanine mutant relative to the wild-type peptides (Table 1). This shows that the target lysine is indeed methylated by SMYD2 indicating an excellent power of the previous screens. Among the peptides with validated methylation sites, 24 were selected for further protein work based on their strong methylation and the known important biological functions of the corresponding proteins.

Methylation of non-histone protein substrates by SMYD2

After confirming the methylation of 32 peptide substrates, further methylation analysis of the validated substrates followed at protein level. Methylation of the targets at the protein level can differ from peptide level, because the target lysine may not be accessible in the context of the folded protein. The selected targets were cloned as GST fusion proteins, over-expressed and purified by affinity chromatography. For 18 of them purification was possible at sufficient quality for the follow-up work (Figure S2). Sequencing results of NFKBID revealed mutations and this target was excluded from further analysis. Roughly equal protein amounts were incubated with SMYD2 in methylation buffer containing radioactively labeled AdoMet. After methylation, the samples were separated on a 16% SDS gel and the methyl group transfer was detected by autoradiography. As positive control, the p53 protein was included. The results shown in Figure 5 demonstrate that six proteins (AFF1, CENPW, TAF1, PHF20, FGD5 and RAPH1, marked with a red asterisk) were stronger methylated than p53. The methylation signals of three proteins (UHRF2, RAD18 and CAP1, marked with blue asterisk) were similar to p53. For five proteins (TMUB1, CHD3, AGAP2, PHF2, and ELAC1, marked

Table 1. List of potential novel non-histone targets identified in Scansite search (Figure 4A) and PhosphoSite Plus database (Figure 4B) for which target site methylation was investigated in the SPOT array shown in Figure 4C. The target lysine residue is highlighted in boldface. Column 1: Swissprot protein number; column 2: abbreviation; column 3: target lysine; column 4: original screen leading to the discovery of this methylation site (Figure 4A or B); column 5: position of the spot in Figure 4C; column 6: position of K-to-A variant spot in Figure 4C; column 7: target site methylation validated; column 8: methylation sites studied at protein level are indicated by “+”.

1	2	Name	Sequence	3	4	5	6	7	8
		artificial peptide	RNEPPKLRSRGAF T	8	1A	2A			
Q9NW08	POLR3B	DNA-directed RNA polymerase III subunit RPC2	PVYYQKLKHMVLDK M	1013	B	3A	4A	+	
P30876	POLR2B	DNA-directed RNA polymerase II subunit RPB2	PTYQQRLKHMVDD KI	1052	B	5A	6A	+	+
P07305	H1F0	histone H1.0	PKKSVAFK KT KEIK	107	B	7A	8A	+	
O94761	RECQL4	ATP-dependent DNA helicase Q4	PDYGQRLKANLKG TL	110	B	9A	10A	+	
Q09666	AHNAK	neuroblast differentiation-associated protein AHNAK	KLKGP K FKMP EM HF K	806	B	11A	12A	–	
Q6ZNL6	FGD5	FYVE, RhoGEF and PH domain-containing protein 5	DGCFG EL KKR GRA VP	1301	B	13A	14A	+	+
Q70E73	RAPH1	Ras-associated and pleckstrin homology domains-containing protein 1	TLKHG TL KLGLSS SN	134	B	15A	16A	+	+
Q8IWQ1	LIN9	TGS2	HRGGQ PL KKR RG SS K	143	B	17A	18A	+	
P35749	MYH11	myosin-11	GREVN AL KS KLR RG N	1925	B	21A	1B	+	
P38117	ETFB	electron transfer flavoprotein subunit β	ATLP NIM KAK KK K IE	200	B	2B	3B	+	
Q09161	NCBP1	nuclear cap-binding protein subunit 1	AN TES YL KRR Q K TH V	204	B	4B	5B	+	
Q01518	CAP1	adenylyl cyclase-associated protein 1	THKN PAL KA Q SG P VR	286	B	6B	7B	+	+
Q9UBF8	PI4KB	phosphatidylinositol 4-kinase β	ISL SS NL KRT AS NPK	290	B	8B	9B	+	
A6NCV1	OR6C74	olfactory receptor 6C74	KQ V KD V FK HT V K K IE	300	B	10B	11B	+	
Q71F23	CENPU	centromere protein U	SQ ML T NL KR KNA K MI	303	B	12B	13B	+	
P04637	p53	cellular tumor antigen p53	RA HSS HL KS KG Q ST	370	B	14B	15B	+	+
Q9UPS6	SETD1B	histone-lysine N-methyltransferase SETD1B	LMID PAL KK G HH K LY	41	B	16B	17B	+	
Q2Q1W2	TRIM71	E3 ubiquitin-protein ligase TRIM71	KATGD GL K RAL Q G K V	495	B	18B	19B	+	
Q8TEW8	PARD3B	partitioning defective 3 homologue B	AGL GV SL KGN KS RET	514	B	20B	21B	+	
Q8WXG6	MADD	MAP kinase-activating death domain protein	AT PF PS L KG N R R AL V	884	B	1C	2C	+	
Q8NI38	NFKBID	NF- κ B inhibitor δ	EG LR QL L K RS R V AP P	454	A	3C	4C	+	
P51825	AFF1	AF4/FMR2 family member 1	K PA K PAL K RS R RE AD	883	A	5C	6C	+	+
Q5EE01	CENPW	centromere protein W	H V L AA A K V IL K K S R G	84	A	7C	8C	+	+
Q12873	CHD3	chromodomain-helicase-DNA-binding protein 3	P V R T KK L K R G R P G R K	348	A	9C	10C	+	+
Q99490	AGAP2	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	E PP AP GL K R G R E G G R	329	A	11C	12C	+	+
Q96PU4	UHRF2	E3 ubiquitin-protein ligase UHRF2	S R G K T PL K NG S S C K R	166	A	13C	14C	+	+
Q96PU4	UHRF2	E3 ubiquitin-protein ligase UHRF2	V V K A G ER L K M S K K K A	407	A	15C	16C	+	
Q9BVT8	TMUB1	transmembrane and ubiquitin-like domain-containing protein 1	H D T I G S L K R T Q F P G R	129	A	17C	18C	+	+
O75151	PHF2	lysine-specific demethylase PHF2	AG K R LL K RA K NS V DL	847	A	19C	20C	+	+
Q9NYY6	RRN3	RNA polymerase I-specific transcription initiation factor RRN3	P F D P C V L K R S K K F ID	567	A	21C	1D	+	+
Q13618	CUL3	cullin-3	L F I D D K L K G V K G L T	397	A	2D	3D	+	+
P21675	TAF1	transcription initiation factor TFIID subunit 1	S K K ES S L K K S R L L G	557	A	4D	5D	+	+
Q9BVI0	PHF20	PHD finger protein 20	K G C E V PL K R P R L D K N	299	A	6D	7D	+	+
Q9H777	ELAC1	zinc phosphodiesterase ELAC protein 1	Q L M K S QL K A G R I T K I	51	A	8D	9D	+	+

with black asterisk) methylation was detected, but weaker than for p53. Only three of the proteins did not show detectable methylation (RRN3, CLU3 and Polr2B), indicating an excellent overall success rate of the validation of peptide substrates in this analysis of 82%.

Because we were mostly interested in targets with strong methylation, verification of the predicted target lysine methylation was performed for nine selected protein substrates (AFF1, CENPW, UHRF2, RAD18, TAF1, PHF20, CAP1, FGD5 and RAPH1; Table 2). To this end, site-directed mutagenesis was performed to exchange the target lysine against arginine. The mutated proteins were overexpressed, purified by affinity chromatography and methylated as described before (Figure 6). All K-to-R mutant non-histone targets (except RAPH1) exhibited a loss or strong decrease in methylation relative to their corresponding wild-type proteins. In the case of FGD5, mutation of the predicted target lysine only led to a gradual decrease in methylation signal. The sequence of this protein contains two additional LK motif in close proximity to the predicted target lysine, but mutation of these two lysine residues did not lead

to a further decrease in the methylation signal (data not shown). Further investigation is necessary to detect which additional lysine residue of FGD5 is methylated by SMYD2. In the case of RAPH1, the initial mutation of the target lysine K134 did not lead to a decrease in methylation. This protein contains additional LK motifs and mutation of K129 to arginine led to a decrease in methylation. Complete loss of methylation was achieved by combination of the K129R and K134R mutations indicating that SMYD2 methylates the predicted target lysine K134 but also K129.

In summary, methylation of the lysine residue predicted from peptide studies was validated for all nine tested proteins. Three of them (CAP1, FGD5 and RAPH1) had been shown previously in proteomics screens to be methylated in cells at the SMYD2 target lysine identified here, but the PKMT responsible to introduce these methylations was not known. Our data strongly suggest that these methylation events are catalyzed by SMYD2. The newly discovered SMYD2 substrates all have important biological functions in chromatin regulation (CENPW, PHF20, UHRF2, and RAD18), transcription (AFF1,

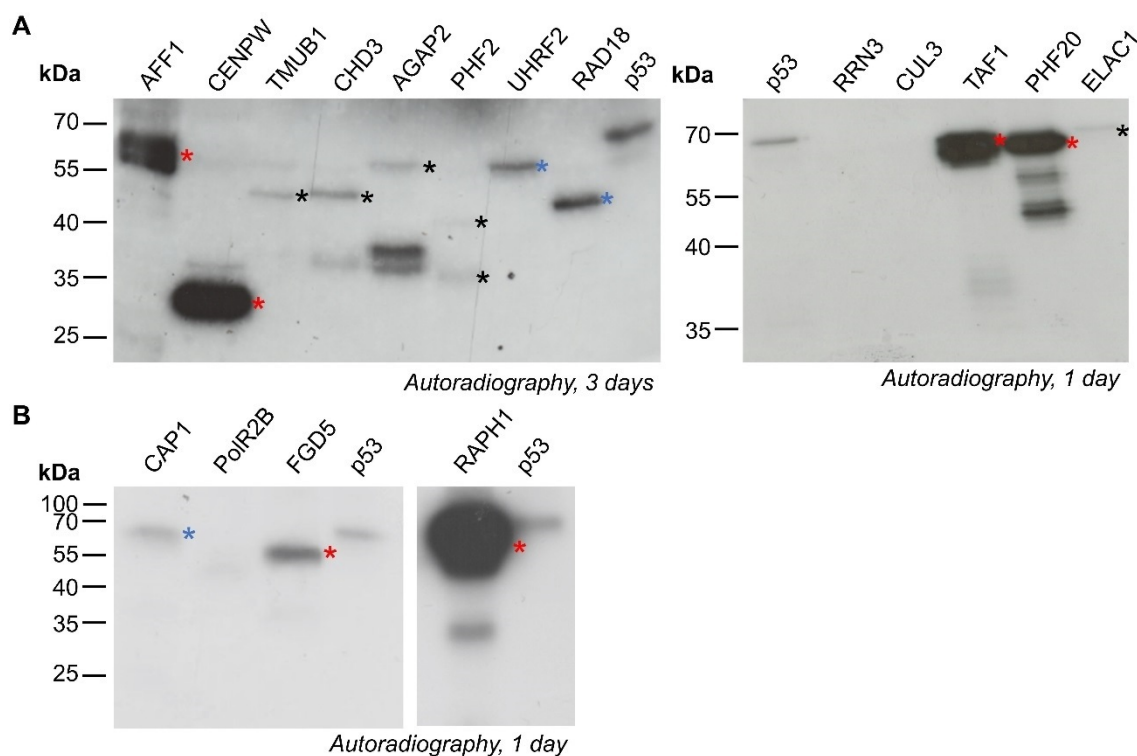


Figure 5. Methylation of SMYD2 non-histone targets at protein level. The selected candidate non-histone proteins were purified by affinity chromatography and similar protein amounts (Figure S2) were incubated with radioactively labeled AdoMet and SMYD2. As positive control, p53 was included in each experiment. After methylation, the samples were separated by SDS-PAGE and the transfer of methyl groups was detected by autoradiography. A) Methylated non-histone target protein domains coming from the Scansite search. B) Methylated non-histone target protein domains coming from the PhosphoSite Plus database search for methylated proteins. Red asterisks indicate substrates with stronger methylation than p53. Targets marked with blue or black asterisks were methylated similarly to p53 or weaker than p53, correspondingly.

Table 2. List of newly discovered SMYD2 protein substrates with validated target lysine methylation. The target lysine residue is highlighted in boldface. Column 1: Swissprot protein number; column 2: abbreviation; column 3: target lysine position; column 4: boundaries of the cloned domains used in this study; column 5: approximate methylation level estimated from the autoradiographic images in Figure 5 indicated by ++ or +, if the methylation was stronger than p53 or similar to p53.

1	2	Name	Sequence	3	4	5
Q6ZNL6	FGD5	FYVE, RhoGEF and PH domain-containing protein 5	DGCFGELKKRGRAVP	1301	1207–1462	++
Q70E73	RAPH1	Ras-associated and pleckstrin homology domains-containing protein 1	TLKHGTLKGLSSSSN	129, 134	1–260	++
Q9BVI0	PHF20	PHD finger protein 20	KGCEVPLKRPRLDKN	299	267–452	++
P21675	TAF1	transcription initiation factor TFIID subunit 1	SKKESLLKSRILLG	557	414–665	++
P51825	AFF1	AF4/FMR2 family member 1	KPAKPALKRSRREAD	883	668–916	++
Q5EE01	CENPW	centromere protein W	HVLAALKVILKKSRR	84	9–88	++
Q01518	CAP1	adenylyl cyclase-associated protein 1	THKNPALKAQSGPVR	286	1–320	+
Q9NS91	RAD18	E3 ubiquitin-protein ligase RAD18	ASRQSLKQGSRLMDN	127	64–232	+
Q96PU4	UHRF2	E3 ubiquitin-protein ligase UHRF2	SRGKTPLKNGSSCKR	166	81–350	+

PHF20, TAF1), and intracellular signaling via ubiquitylation (UHRF2, RAD18), the Ras pathway (FGD5, RAPH1) or other pathways (CAP1, AGAP2). Future studies will need to validate the SMYD2 dependent methylation of these proteins in cells and study its biological consequences.

Conclusions

It was the aim of this work to identify novel substrates of the SMYD2 PKMT. Using peptide arrays, we investigated the substrate sequence specificity profile of the enzyme, revealing a

rather low level of specificity, because only at the –1 site a real recognition of leucine (or at weaker level phenylalanine) was detected. At the +1 to +3 sites, some less specific additional effects were observed including the disfavor for acidic residues. Using the derived sequence specificity motif, novel SMYD2 peptide substrates were identified and, in a second step, methylation at the target lysine was validated. Strikingly, high fractions of predicted SMYD2 peptide substrates were indeed found to be methylated, 33% in the case of the full motif and still 12% in the case of the reduced [LF]-[K] motif. Our study has led to the discovery of 32 novel peptide sub-

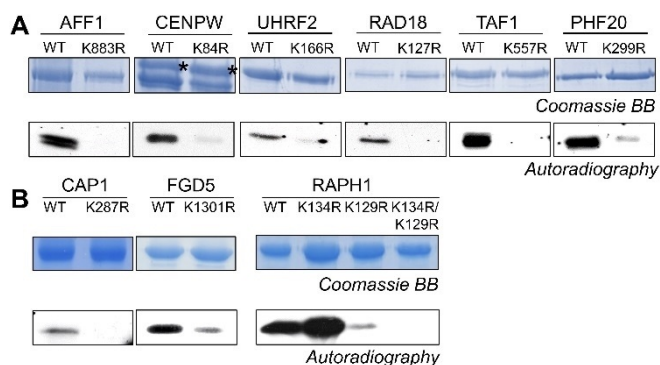


Figure 6. Validation of target lysine methylation of the novel SMYD2 protein substrates. Site directed mutagenesis was performed to create lysine to arginine mutants of the selected non-histone protein substrates. Similar protein amounts of the wild-type and K-to-R proteins were used for methylation experiments shown in the autoradiography images. For CENPW, black asterisks indicate the protein bands that correspond to methylation signal. A) Non-histone target protein domains and their K-to-R mutants identified in the Scansite search. B) Non-histone target protein domains and their K-to-R mutants identified in the PhosphoSite Plus database search of methylated proteins. Exposure of the autoradiography films was 8 h for all targets, except RAPH1 (1 day) and CAP1, FGD5 (3 days).

strates with validated target sites. Among them, 19 were already reported to be methylated at the target lysine in human cells, strongly suggesting that SMYD2 is the PKMT responsible for these methylation events. Next, methylation of the novel substrates was tested at the protein level leading to the identification of 14 novel protein substrates of SMYD2. However, as validation of protein methylation was very efficient for those proteins which could be successfully cloned, expressed and purified, it is very likely that many of the additional novel peptide substrates will also be methylated by SMYD2 at protein level. Six of the discovered protein substrates were methylated more strongly than p53, the best SMYD2 substrate described so far. The novel SMYD2 substrate proteins are involved in diverse biological processes like chromatin regulation, transcription and different intracellular signaling pathways. Overall, our study boosts the number of known SMYD2 substrates considerably providing a fundament for future studies to investigate the function of the methylation of these novel SMYD2 substrates in cells.

Experimental Section

Cloning, expression, and purification of the proteins: SMYD2 was cloned as described.^[14] The domains of the non-histone substrates and p53 were amplified by PCR using cDNA isolated from HEK293 cells as template. The amplified DNA was inserted into the pGEX-6p2 bacterial expression vector for protein expression. Protein domains of the non-histone substrates were predicted with the Scooby domain prediction tool (<http://www.ibi.vu.nl/programs/scoobywww/>).^[40] The different mutations were introduced by site-directed mutagenesis.^[41] All cloning steps were confirmed by sequencing. Protein overexpression was performed in BL21 DE3 Codon Plus cells (Novagen) in ampicillin containing lysogeny broth at 37 °C till the main culture reached an OD₆₀₀ of 0.6. Afterward, protein expression was induced by the addition of 1 mM isopropyl

β -D-thiogalactopyranoside (IPTG) at 22 °C for 12 h. Thereafter, the cells were centrifuged (4000g, 20 min, 4 °C) and stored at –20 °C. For protein purification, the pellet was resuspended in 25 mL sonication buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM DTT, 5% glycerol) and sonicated (13 rounds, 30% power, 40-second intervals). Following centrifugation (20000g, 80 min, 4 °C), the obtained supernatant was loaded onto a column containing pre-equilibrated glutathione-Sepharose 4B beads (GE Healthcare). The column was washed once with sonication buffer and twice with wash buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM DTT, 5% glycerol). Afterward, the protein was eluted from the column with wash buffer supplemented with 40 mM reduced glutathione (pH 8). The pooled fractions were dialyzed against dialysis buffer I (20 mM Tris-HCl pH 7.4, 100 mM KCl, 0.5 mM DTT, 10% glycerol) for 3 h and overnight in dialysis buffer II (20 mM Tris-HCl pH 7.4, 100 mM KCl, 0.5 mM DTT, 60% glycerol). The proteins were stored at –20 °C.

Peptide array synthesis and methylation: The peptide arrays were synthesized with an Autospot peptide array synthesizer (Intavis AG, Köln, Germany) using the SPOT synthesis method.^[36,37] The arrays contain 15 or 20 amino acid long peptides immobilized on a cellulose membrane. After synthesis, the membrane was pre-incubated in methylation buffer (50 mM Tris-HCl pH 9, 100 mM NaCl, 5 mM DTT) for 5 min. Thereafter, the membrane was incubated in methylation buffer supplemented with 0.76 μ M radioactively labeled AdoMet (PerkinElmer) and 0.6 μ M SMYD2 for 1 h at 25 °C on a shaker. Afterward, the membrane was washed five times for 5 min with wash buffer (100 mM NH₄HCO₃, 1% SDS), followed by the incubation in amplify NAMP100V (GE Healthcare) for 10 min. Then, the membrane was exposed to a Hyperfilm high-performance autoradiography film (GE Healthcare) at –80 °C in the dark for various periods of time and developed with a developing machine.

Protein methylation assay: Protein methylation was performed using 5–10 μ g of the substrate proteins in a total volume of 40 μ L methylation buffer (50 mM Tris-HCl pH 9, 100 mM NaCl, 5 mM DTT) supplemented with 0.76 μ M radioactive labeled AdoMet (PerkinElmer) and 0.6 μ M SMYD2 for 3–8 h at 25 °C. The reactions were stopped by the addition of SDS loading buffer and incubation at 95 °C for 5 min. Thereafter, the samples were separated by 12 or 16% SDS-PAGE. This step was followed by the incubation of the gel in amplify NAMP100V (GE Healthcare) for 45 min and drying of the gel in vacuum at 60 °C for 90 min. The dried SDS gel was then exposed to a Hyperfilm high performance autoradiography film as described above.

Circular dichroism analysis: Circular dichroism (CD) spectra of SMYD2 were measured using a J-815 circular dichroism spectrophotometer (Jasco) using 7 μ M protein diluted in a buffer containing 100 mM KCl. The spectra were measured in a wavelength range between 190 nm and 240 nm by using a 0.1 mm cuvette. The measurement was performed with an accumulation of 120 scans and a scan speed of 200 nm min^{–1}. The spectrum obtained with the dialysis buffer II was used as baseline. For the determination of the melting curve, 16 μ M protein was used. The measurement was performed at a wavelength of 210 nm in the temperature range of 20 and 80 °C using a heating rate of 1 °C min^{–1}. The obtained data was analyzed with Microsoft Excel as described.^[42]

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant JE 252/7-4.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: enzyme specificity · peptide array · protein lysine methyltransferase · SMYD2

- [1] S. Lanouette, V. Mongeon, D. Figeys, J. F. Couture, *Mol. Syst. Biol.* **2014**, *10*, 724.
- [2] S. G. Clarke, *Trends Biochem. Sci.* **2013**, *38*, 243–252.
- [3] X. Zhang, Y. Huang, X. Shi, *Cell. Mol. Life Sci.* **2015**, *72*, 4257–4272.
- [4] K. K. Biggar, S. S. Li, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 5–17.
- [5] P. A. Del Rizzo, R. C. Trievel, *Biochim. Biophys. Acta Gene Regul. Mech.* **2014**, *1839*, 1404–1415.
- [6] P. A. Boriack-Sjodin, K. K. Swinger, *Biochemistry* **2016**, *55*, 1557–1569.
- [7] N. Spellmon, J. Holcomb, L. Trescott, N. Sirinupong, Z. Yang, *Int. J. Mol. Sci.* **2015**, *16*, 1406–1428.
- [8] C. Tracy, J. S. Warren, M. Szulik, L. Wang, J. Garcia, A. Makaju, K. Russell, M. Miller, S. Franklin, *Curr. Opin. Physiol.* **2018**, *1*, 140–152.
- [9] M. A. Brown, R. J. Sims III, P. D. Gottlieb, P. W. Tucker, *Mol. Cancer* **2006**, *5*, 26.
- [10] M. Abu-Farha, J. P. Lambert, A. S. Al-Madhoun, F. Elisma, I. S. Skerjanc, D. Figeys, *Mol. Cell. Proteomics* **2008**, *7*, 560–572.
- [11] J. Huang, L. Perez-Burgos, B. J. Placek, R. Sengupta, M. Richter, J. A. Dorsey, S. Kubicek, S. Opravil, T. Jenuwein, S. L. Berger, *Nature* **2006**, *444*, 629–632.
- [12] L. A. Saddic, L. E. West, A. Aslanian, J. R. Yates, 3rd, S. M. Rubin, O. Gozani, J. Sage, *J. Biol. Chem.* **2010**, *285*, 37733–37740.
- [13] D. Marouco, A. V. Garabadgiu, G. Melino, N. A. Barlev, *Oncotarget* **2013**, *4*, 1556–1571.
- [14] A. D. Ferguson, N. A. Larsen, T. Howard, H. Pollard, I. Green, C. Grande, T. Cheung, R. Garcia-Arenas, S. Cowen, J. Wu, R. Godin, H. Chen, N. Keen, *Structure* **2011**, *19*, 1262–1273.
- [15] L. Wang, L. Li, H. Zhang, X. Luo, J. Dai, S. Zhou, J. Gu, J. Zhu, P. Atadja, C. Lu, E. Li, K. Zhao, *J. Biol. Chem.* **2011**, *286*, 38725–38737.
- [16] J. Wu, T. Cheung, C. Grande, A. D. Ferguson, X. Zhu, K. Theriault, E. Code, C. Birr, N. Keen, H. Chen, *Biochemistry* **2011**, *50*, 6488–6497.
- [17] R. Hamamoto, V. Saloura, Y. Nakamura, *Nat. Rev. Cancer* **2015**, *15*, 110–124.
- [18] X. Zhang, K. Tanaka, J. Yan, J. Li, D. Peng, Y. Jiang, Z. Yang, M. C. Barton, H. Wen, X. Shi, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17284–17289.
- [19] Y. Jiang, L. Trescott, J. Holcomb, X. Zhang, J. Brunzelle, N. Sirinupong, X. Shi, Z. Yang, *J. Mol. Biol.* **2014**, *426*, 3413–3425.
- [20] L. Piao, D. Kang, T. Suzuki, A. Masuda, N. Dohmae, Y. Nakamura, R. Hamamoto, *Neoplasia* **2014**, *16*, 257–264.
- [21] N. Reynoird, P. K. Mazur, T. Stellfeld, N. M. Flores, S. M. Lofgren, S. M. Carlson, E. Brambilla, P. Hainaut, E. B. Kaznowska, C. H. Arrowsmith, P. Khatri, C. Stresemann, O. Gozani, J. Sage, *Genes Dev.* **2016**, *30*, 772–785.
- [22] M. Nakakido, Z. Deng, T. Suzuki, N. Dohmae, Y. Nakamura, R. Hamamoto, *Neoplasia* **2015**, *17*, 367–373.
- [23] S. Weirich, A. Jeltsch in *Encyclopedia of Cancer*, 3rd ed. (Eds.: P. Boffetta, P. Hainaut), Elsevier, Amsterdam, **2019**, pp. 538–550.
- [24] X. Yi, X. J. Jiang, Z. M. Fang, *Clin. Epigenetics* **2019**, *11*, 112.
- [25] S. Weirich, S. Kudithipudi, A. Jeltsch, *J. Mol. Biol.* **2016**, *428*, 2344–2358.
- [26] J. B. Olsen, X. J. Cao, B. Han, L. H. Chen, A. Horvath, T. I. Richardson, R. M. Campbell, B. A. Garcia, H. Nguyen, *Mol. Cell. Proteomics* **2016**, *15*, 892–905.
- [27] S. Lanouette, J. A. Davey, F. Elisma, Z. Ning, D. Figeys, R. A. Chica, J. F. Couture, *Structure* **2015**, *23*, 206–215.
- [28] H. Ahmed, S. Duan, C. H. Arrowsmith, D. Barsyte-Lovejoy, M. Schapira, *J. Proteome Res.* **2016**, *15*, 2052–2059.
- [29] P. Rathert, A. Dhayalan, M. Murakami, X. Zhang, R. Tamas, R. Jurkowska, Y. Komatsu, Y. Shinkai, X. Cheng, A. Jeltsch, *Nat. Chem. Biol.* **2008**, *4*, 344–346.
- [30] A. Dhayalan, S. Kudithipudi, P. Rathert, A. Jeltsch, *Chem. Biol.* **2011**, *18*, 111–120.
- [31] S. Kudithipudi, A. Dhayalan, A. F. Kebede, A. Jeltsch, *Biochimie* **2012**, *94*, 2212–2218.
- [32] S. Kudithipudi, C. Lungu, P. Rathert, N. Happel, A. Jeltsch, *Chem. Biol.* **2014**, *21*, 226–237.
- [33] M. K. Schuhmacher, S. Kudithipudi, D. Kusevic, S. Weirich, A. Jeltsch, *Biochim. Biophys. Acta Gene Regul. Mech.* **2015**, *1849*, 55–63.
- [34] S. Kudithipudi, M. K. Schuhmacher, A. F. Kebede, A. Jeltsch, *ACS Chem. Biol.* **2017**, *12*, 958–968.
- [35] M. K. Schuhmacher, M. Rolando, A. Brohm, S. Weirich, S. Kudithipudi, C. Buchrieser, A. Jeltsch, *J. Mol. Biol.* **2018**, *430*, 1912–1925.
- [36] S. Kudithipudi, D. Kusevic, S. Weirich, A. Jeltsch, *J. Vis. Exp.* **2014**, e52203.
- [37] R. Frank, *J. Immunol. Methods* **2002**, *267*, 13–26.
- [38] J. C. Obenauer, L. C. Cantley, M. B. Yaffe, *Nucleic Acids Res.* **2003**, *31*, 3635–3641.
- [39] P. V. Hornbeck, J. M. Kornhauser, V. Latham, B. Murray, V. Nandhikonda, A. Nord, E. Skrzypek, T. Wheeler, B. Zhang, F. Gnad, *Nucleic Acids Res.* **2019**, *47*, D433–D441.
- [40] R. A. George, K. Lin, J. Heringa, *Nucleic Acids Res.* **2005**, *33*, W160–W163.
- [41] A. Jeltsch, T. Lanio, *Methods Mol. Biol.* **2002**, *182*, 85–94.
- [42] N. J. Greenfield, *Nat. Protoc.* **2006**, *1*, 2527–2535.

Manuscript received: September 18, 2019

Accepted manuscript online: October 15, 2019

Version of record online: October 29, 2019