The FEBS Journal

FEBS PRESS science publishing by scientists

Structural studies of protein arginine methyltransferase 2 reveal its interactions with potential substrates and inhibitors

Vincent Cura[1], Nils Marechal[1], Nathalie Troffer-Charlier[1], Jean-Marc Strub[2], Matthijs J. van Haren[3], Nathaniel I. Martin[3], Sarah Cianférani[2], Luc Bonnefond[1] and Jean Cavarelli[1]

1 Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U596, Université de Strasbourg, Illkirch, France

2 BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC, UMR7178, Université de Strasbourg, France

3 Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands

Keywords
CARM1; epigenetics; PRMT2; PRMT4; RSF1

Correspondence
J. Cavarelli, Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U596, Université de Strasbourg, 1 rue Laurent Fries, Illkirch F-67404, France
Fax: +33 (0)3 88 65 32 01
Tel: +33 (0)3 69 48 52 74
E-mail: cava@igbmc.fr

(Received 4 August 2016, revised 27 September 2016, accepted 2 November 2016)

doi:10.1111/febs.13953

PRMT2 is the less-characterized member of the protein arginine methyltransferase family in terms of structure, activity, and cellular functions. PRMT2 is a modular protein containing a catalytic Ado-Met-binding domain and unique Src homology 3 domain that binds proteins with proline-rich motifs. PRMT2 is involved in a variety of cellular processes and has diverse roles in transcriptional regulation through different mechanisms depending on its binding partners. PRMT2 has been demonstrated to have weak methyltransferase activity on a histone H4 substrate, but its optimal substrates have not yet been identified. To obtain insights into the function and activity of PRMT2, we solve several crystal structures of PRMT2 from two homologs (zebrafish and mouse) in complex with either the methylation product S-adenosyl-L-homocysteine or other compounds including the first synthetic PRMT2 inhibitor (Cp1) studied so far. We reveal that the N-terminal-containing SH3 module is disordered in the full-length crystal structures, and highlights idiosyncratic features of the PRMT2 active site. We identify a new nonhistone protein substrate belonging to the serine-/arginine-rich protein family which interacts with PRMT2 and we characterize six methylation sites by mass spectrometry. To better understand structural basis for Cp1 binding, we also solve the structure of the complex PRMT4:Cp1. We compare the inhibitor–protein interactions occurring in the PRMT2 and PRMT4 complex crystal structures and show that this compound inhibits efficiently PRMT2. These results are a first step toward a better understanding of PRMT2 substrate recognition and may accelerate the development of structure-based drug design of PRMT2 inhibitors.

Database
All coordinates and structure factors have been deposited in the Protein Data Bank: $zPRMT2_{1-408}$-SFG = 5g02; $zPRMT2_{73-408}$-SAH = 5fub; $mPRMT2_{1-445}$-SAH = 5ful; $mPRMT2_{1-445}$-Cp1 = 5fwa, $mCARM1_{130-487}$-Cp1 = 5k8v.

Abbreviations
mPRMT2, mouse (*Mus musculus*) PRMT2; PDB, Protein Data Bank; PRMT, protein arginine methyltransferase; RMSD, root-square-mean deviation; SAH, S-adenosyl-L-homocysteine; SAM, S-adenosyl-L-methionine; SAXS, small-angle X-ray scattering; SFG, sinefungin; zPRMT2, zebrafish (*Danio rerio*) PRMT2.

Introduction

Protein arginine N-methyltransferases (PRMTs) are eukaryotic enzymes that catalyze the transfer of methyl groups from the methyl donor S-adenosyl-L-methionine (SAM) to guanidine nitrogens in arginine residues of target proteins. The PRMT family has been the focus of many biological researches and at least nine members of the PRMT family (PRMT1 to PRMT9) have been identified in mammalian cells [1–3]. PRMTs are divided into four major classes (type I to type IV) depending on the chemistry of methyl-arginine they generate. Type I and type II PRMT, the two most populated classes have been extensively studied. Type I PRMTs (PRMT1, PRMT2, PRMT3, PRMT4/CARM1, PRMT6, and PRMT8) produce asymmetric dimethylarginine, whereas type II (PRMT5, PRMT9) form symmetric dimethylarginine. PRMT7 produces monomethylarginine residues only, which defines it as the first type III PRMT. A fourth type which results in the monomethylation of the internal δ guanidino nitrogen has only been described in yeast [4]. Thanks to extensive research by many teams, several dozen crystal structures from seven different PRMTs have been determined either free or in complex with cofactor analogs [5–19]. Only PRMT2 and PRMT9 remain to be characterized at the structural level.

Human PRMT2 (or HRMT1L1) was first identified by homology to the yeast methyltransferase *HMT1* gene in 1998 [20]. PRMT2 contains a highly conserved catalytic SAM-binding domain and is characterized by an N-terminal extension partially folded into a Src homology 3 (SH3) domain. PRMT2 harbors a weak methyltransferase activity on histone H4 tail *in vitro*, but no optimal substrate has been found yet [21].

It has been shown that PRMT2 interacts with nuclear hormone receptors, including estrogen receptor alpha, retinoic acid receptors, and androgen receptors, and enhances their transcriptional activity in a ligand-dependent manner [22,23]. In addition to its role as a transcriptional coactivator, PRMT2 has also been found to be involved in diverse cellular processes, through its interaction with partners such as the adenoviral early region 1B-associated protein (E1B-AP5) [24], the retinoblastoma protein [25], β-catenin [26], or the transcription factor STAT3 [27].

Implication of PRMT2 in breast carcinogenesis in correlation with nuclear receptors expression has been reported [28,29]. Loss of PRMT2 nuclear expression in breast cancer cells is linked to increased cyclin D1 expression, thus promoting breast tumor cell proliferation. In addition to the original full-length PRMT2, four alternatively spliced PRMT2 isoforms (PRMT2L2, PRMT2α, β, and γ) have been identified. Together, all these isoforms showed increased expression in breast tumor samples [30,31].

So far no inhibitor of PRMT2 has been studied. Because of the lack of structural data and the weak enzyme activity, very few information is available to understand how this protein recognizes its substrates.

In the present study, we report the crystal structures of the PRMT2 catalytic module from zebrafish (*Danio rerio*) and from mouse (*Mus musculus*) in complex with either the methylation product S-adenosyl-L-homocysteine (SAH) or other compounds including a potent CARM1 inhibitor. We identify a new nonhistone protein substrate belonging to the serine-/arginine-rich protein family which interacts with PRMT2. We compare the inhibitor–protein interactions occurring in the PRMT2 and CARM1 complex structures and show that this compound inhibits efficiently PRMT2.

Results

PRMT2 from zebrafish

PRMT2 from zebrafish (*D. rerio*) was expressed in insect cells using the baculovirus system, purified and crystallized. The first crystals were obtained from a full-length construct ($zPRMT2_{1\text{-}408}$) in the presence of sinefungin, a natural inhibitor. The structure was solved and refined at 2.45 Å resolution (Table 1). The electron density map revealed the absence of residues 1–71 corresponding to the N-terminal part containing the SH3 domain. A SDS/PAGE analysis performed on dissolved crystals showed that the crystallized product was shorter than the purified protein, indicating that proteolysis occurred in the crystallization drop (Fig. 1A). Mass spectrometry LC-MS/MS analysis was performed on a trypsin digest of the gel-extracted protein band to localize the cleavage site. Sequence resulting from tryptic peptide identification revealed that the crystallized molecule started at residue 68, four residues before the catalytic module visible in the zPRMT2 crystal structure. zPRMT2 N-terminal degradation was potentially caused by a bacterial contamination of the sinefungin solution used for crystallization. Subsequent attempts to crystallize the full-length zPRMT2 with a fresh sinefungin solution were successful but the resulting crystals diffracted only at low resolution (5 Å). To improve the initial crystals obtained from a degraded protein, a new construct $zPRMT2_{73\text{-}408}$ restricted to the catalytic domain was cloned. The protein was expressed, purified and crystallized in the presence of SAH. The structure was

Table 1. X-ray data collection and refinement statistics. Statistics for the highest resolution shell are shown in parentheses.

	$zPRMT2_{1-408}$	$zPRMT2_{73-408}$	$mPRMT2_{1-445}$	$mPRMT2_{1-445}$	$mCARM1_{130-487}$
Ligand	Sinefungin	SAH	SAH	Compound 1	Compound 1
PDB ID	5g02	5fub	5ful	5fwa	5k8v
Data processing					
Wavelength (Å)	0.979	0.980	0.873	0.979	0.980
Resolution range (Å)	45.3–2.45 (2.55–2.45)	57.0–2.0 (2.05–2.00)	43.3–1.9 (1.93–1.89)	44.2–1.8 (1.84–1.8)	47.7–2.25 (2.30–2.25)
Space group	H32	H32	$C222_1$	$C222_1$	$P2_12_12$
Unit cell (Å,°)	148.0 148.0 127.7 90 90 120	147.1 147.1 127.4 90 90 120	65.5 115.4 133.0 90 90 90	66.2 114.7 132.7 90 90 90	75.2 98.1 206.3 90 90 90
Total reflections	128 263 (13 045)	191 435 (14 027)	371 882 (21 162)	308 208 (18 911)	474 150 (28 897)
Unique reflections	19 847 (2193)	35 929 (2639)	40 533 (3933)	47 105 (2780)	73 068 (4372)
Multiplicity	6.5 (5.9)	5.3 (5.3)	9.2 (8.5)	6.5 (6.8)	6.5 (6.6)
Completeness (%)	99.8 (98.7)	99.9 (100.0)	99.8 (97.3)	100 (100)	99.8 (97.6)
$I/\sigma(I)$	11.8 (1.1)	9.2 (1.0)	15.4 (1.2)	11.7 (1.0)	5.7 (1.1)
Resolution limit for $I/\sigma(I) > 2$	2.63	2.14	1.99	1.90	2.52
Wilson B-factor (Å²)	56.9	35.4	30.1	29.9	29.6
R_{meas}[a]	0.087 (1.315)	0.114 (1.904)	0.096 (2.025)	0.097 (1.967)	0.307 (2.314)
$CC_{1/2}$[b]	0.999 (0.698)	0.998 (0.308)	0.999 (0.448)	0.999 (0.470)	0.986 (0.349)
Refinement					
Resolution range (Å)	45.3–2.45 (2.54–2.45)	38.4–2.0 (2.07–2.0)	35.1–1.9 (1.96–1.89)	35.0–1.8 (1.86–1.80)	47.7–2.2 (2.28–2.25)
% *R*-work	19.96 (32.73)	18.17 (30.59)	16.50 (27.80)	17.17 (30.52)	19.24 (30.02)
% *R*-free	24.20 (39.38)	21.75 (31.29)	19.04 (28.43)	19.07 (34.97)	23.48 (32.30)
Number of non-H atoms	2778	2934	3088	3078	11 755
Protein	2728	2722	2749	2723	10 989
Ligands	37	31	24	51	170
Water	13	181	315	304	596
Validation					
RMS (bonds)	0.002	0.004	0.018	0.004	0.002
RMS (angles)	0.68	0.61	1.50	0.86	0.477
Ramachandran favored (%)	96	98	97	98	96.6
Ramachandran outliers (%)	0	0	0	0.3	0
Average B-factor (Å²)	76.92	45.60	36.28	35.55	41.0
Rotamer outliers (%)	1.0	0	0.33	0	0.5

[a] R_{meas} is the redundancy-independent merging.
[b] See [73] for definitions. The resolution limits for $I/\sigma(I) > 2$ are reported.

refined to 2.0 Å resolution and was almost identical to the one obtained after proteolysis (RMSD = 0.27 Å). A small difference occurs at the C-terminal extremity which is rotated by 180° after M406, changing the orientation of the two last residues.

PRMT2 from mouse

Full-length PRMT2 from mouse (*M. musculus*), $mPRMT2_{1-445}$, was expressed in insect cells using the baculovirus system, purified and crystallized. The structure was solved and refined at 2.0 Å resolution (Table 1). The visible part of the structure is limited to the catalytic module (residues 107–445) as for zPRMT2. The molecular mass of the purified mPRMT2 measured by mass spectrometry under denaturing conditions is 50 728 Da for a calculated mass of 50 726 Da for the complete protein. Furthermore, a SDS/PAGE performed on a dissolved mPRMT2 crystal revealed no detectable size difference between the dissolved crystal and the purified protein (Fig. 1B). Therefore, unlike zPRMT2, the absence of residues 1–106 in the electron density map was not the result of proteolysis before or during the crystallization process, but could be explained by a high flexibility of the N-terminal region within the crystals.

Matthews coefficient and solvent content calculated for one full-length monomer per asymmetric unit

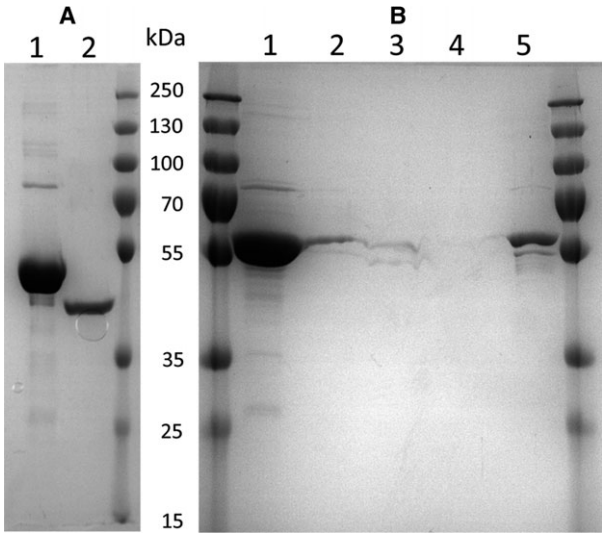


Fig. 1. PAGE analysis of purified PRMT2 and crystals. Proteins were visualized by Coomassie blue staining. For crystals analysis, each crystal was washed three times in 5 μL of reservoir solution and dissolved in SDS/PAGE loading buffer. (A) Lane 1: Purified zPRMT2 (calculated MW 46 651 Da), lane 2: dissolved zPRMT2 crystal. (B) Lane 1: Purified mPRMT2 (calculated MW 50 726 Da); lanes 2, 3, and 4: crystal wash solutions. Lane 5: dissolved mPRMT2 crystal.

(V_m = 2.48 Å^3·Da^{-1}, ~ 50% solvent content) indicates that the unit cell could easily accommodate SH3 domain. In the crystal packing, no steric hindrance could prevent the positioning of SH3 domain in the vicinity of residue D107.

Overall structure

Phylogenetic analysis [32] and sequence comparisons have established that PRMT2 methylation module is closely related to all type I PRMTs (35% to 39% sequence identity between PRMT8, PRMT1, PRMT3, PRMT6, and CARM1 from mouse). As expected, the monomeric structure of PRMT2 catalytic module is very similar to that of all the PRMT known structures especially type I PRMTs. It consists in a SAM-binding domain (residues 107–254, mPRMT2 numbering) adopting a Rossmann fold, a helical dimerization arm (residues 266–298), interrupting the β-barrel domain (residues 255–265 and residues 299–445) (Fig. 2). The two domains are connected by the strictly conserved cis-proline 254.

zPRMT2 and mPRMT2 share 55% sequence identity and up to 59% in their catalytic modules. The structure of the catalytic modules is highly similar and can be well superimposed (RMSD = 0.68 Å) (Fig. 3). The major difference is the orientation of the loop β14-β15 (426–437, 392–400 in zPRMT2). This loop is involved in the crystal packing in mPRMT2 (hydrogen bond between main-chain O L429 and N E205 from a neighbor monomer), whereas in zPRMT2, it is turned toward solvent and opens slightly the β barrel.

SAH-binding pocket

The active site pocket is strictly conserved between the mouse and zebrafish enzymes. The SAH and the sinefungin molecules are well defined in the electronic density. Residues involved in the SAH recognition are conserved among all PRMT2s (Fig. 2). Bifurcated hydrogen bonds involving residue E209 carboxylate and S237 Oγ maintain the adenine amino group. N1 interacts with the V208 main-chain carbonyl and the E180 carboxylate forms two hydrogen bonds with the ribose hydroxyl oxygens (Fig. 4). For the homocystein moiety, the carboxylate group binds to R133, the amino group interacts with the C158 carbonyl and M127 makes van der Waals contacts with the S atom. Helix αX harboring the conserved YFxxY motif closes the SAM-binding pocket as observed in the structures of PRMT3 [6], CARM1 [8], and PRMT6 [18]. Helix αX interacts via Y118 and Y114 with catalytic E232 allowing the proper formation of the substrate arginine pocket required for catalysis (Fig. 3).

Sinefungin binds slightly differently from SAH because of a different orientation of the homocystein moiety. The homocystein amino group is too far to form a hydrogen bond but the sinefungin-specific NH2ε is involved in hydrogen bonds with E189 carboxylate Oε1 and carbonyl oxygens (Fig. 4A).

Substrate-binding pocket

The highly conserved active site glutamate residues E223 and E232 of the so-called double-E loop [7] form a pair of salt bridges with a calcium ion in mPRMT2 (Fig. 4C). Calcium was added to the crystallization drop in order to obtain diffracting crystals. In mPRMT2:SAH structure, this ion occupies the site where the guanidino moiety of the arginine substrate would bind. In zPRMT2, calcium is replaced by a water molecule. E198 of the double-E loop interacts with the ammonium of a MES molecule which fills the arginine substrate channel (Fig. 4A,B).

Dimerization

Homodimerization is a feature conserved in the PRMTs and essential for the catalytic activity (for a discussion and review see [16,18,33]). Exception is the

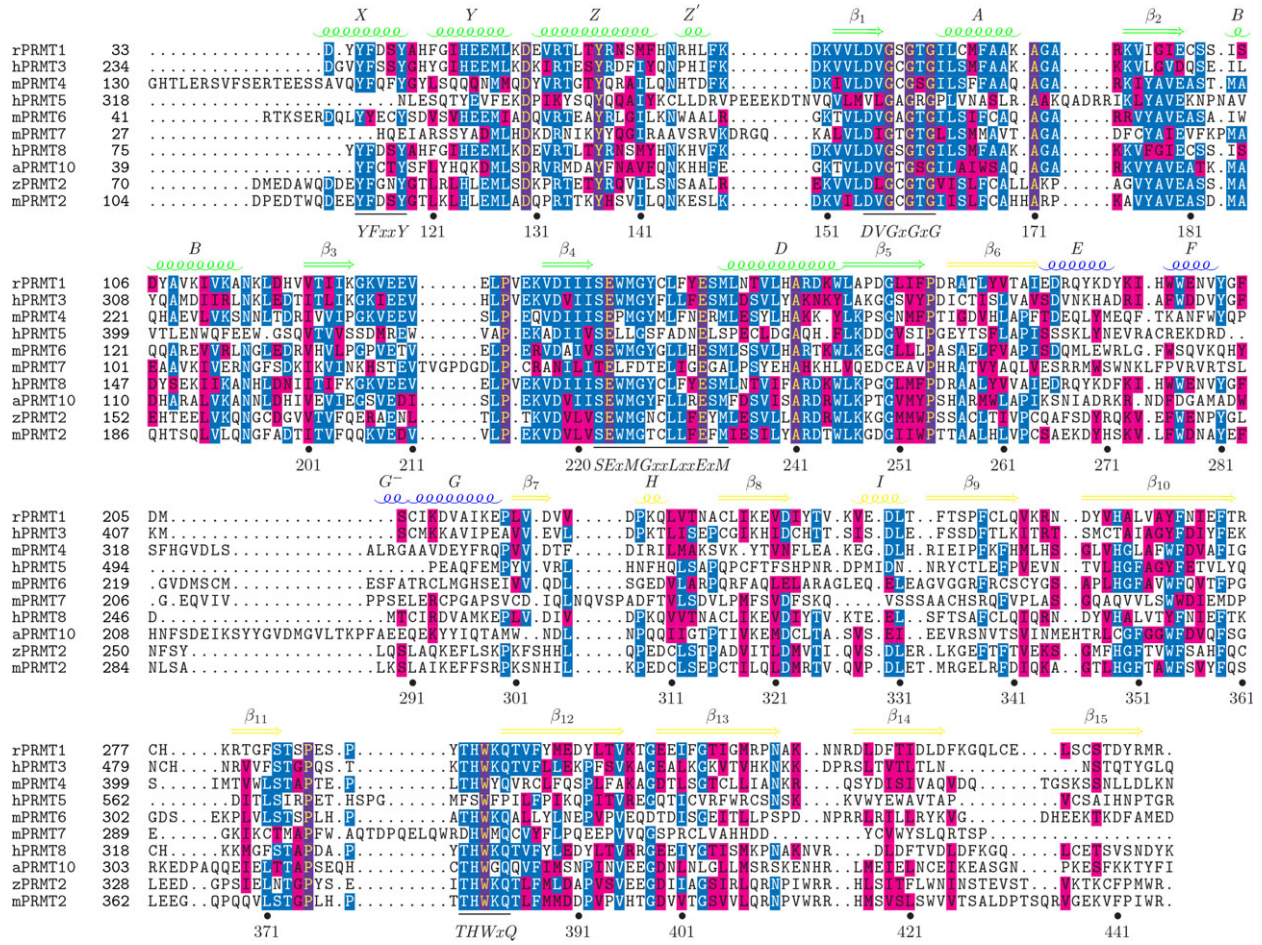


Fig. 2. Structure-based sequence alignment of selected PRMTs. Ten PRMTs sequences have been aligned based on their crystal structures. The alignment is restricted to the catalytic core. The secondary structure of mPRMT2 is drawn above the alignment with color. The SAM-binding domain, the β-barrel domains, and the dimerization arm are colored green, yellow, and blue, respectively. The mPRMT2 residue numbering is shown below the sequences. The four signature sequences are localized and their consensus written below. Amino acids are shaded according to similarity to the consensus sequence. Amino acids highlighted are either invariant (violet) or similar (blue) as defined by the following grouping: F, Y, and W; I, L, M, and V; R and K; D and E; G and A; S, T, N, and Q. Abbreviations are as follow: m/*Mus musculus*, h/*Homo sapiens*, r/*Rattus norvegicus*, a/*Arabidopsis thaliana*, and z/zebrafish (*D. rerio*). This figure was produced with the program TEXSHADE [70].

monomeric PRMT7 which consist in two PRMT modules in tandem forming a pseudo dimer [18,34,35]. Higher order oligomers of some PRMTs, resulting of association of dimers, have been also found in solution both *in vivo* and *in vitro* [5,7,8,11,19,36–40]. zPRMT2 quaternary structure was studied using small-angle X-ray scattering. To ensure that zPRMT2$_{1\text{-}408}$ was not degraded, SDS/PAGE analysis was performed before SAXS experiments. Furthermore, the molecular mass of zPRMT2 monomer was determined from the scattering intensity values $I(0)$ using bovine serum albumin (BSA) as a reference. The resulting value of 47 630 Da corresponds to the size of zPRMT2$_{1\text{-}408}$ (theoretical mass of 46 845 Da), confirming the presence of the SH3 domain in the protein sample. The scattering curves show that the full-length protein behaves as a dimer in solution according to the radius of gyration (R_g = 3.3 nm) (Table 2). The dimeric zPRMT2$_{73\text{-}408}$ exhibits a R_g of 2.9 nm identical to the R_g calculated from the X-ray structure (Fig. 5). However, the comparison of *ab initio* low-resolution models generated from zPRMT2 full-length and zPRMT2$_{73\text{-}408}$ scattering data did not reveal significant differences (data not shown). Therefore, it was not possible to localize SH3 domain on the whole structure using SAXS, probably because of the lack of rigidity of this N-terminal region. In both mPRMT2 and zPRMT2 crystal structures, the two monomers are related by a

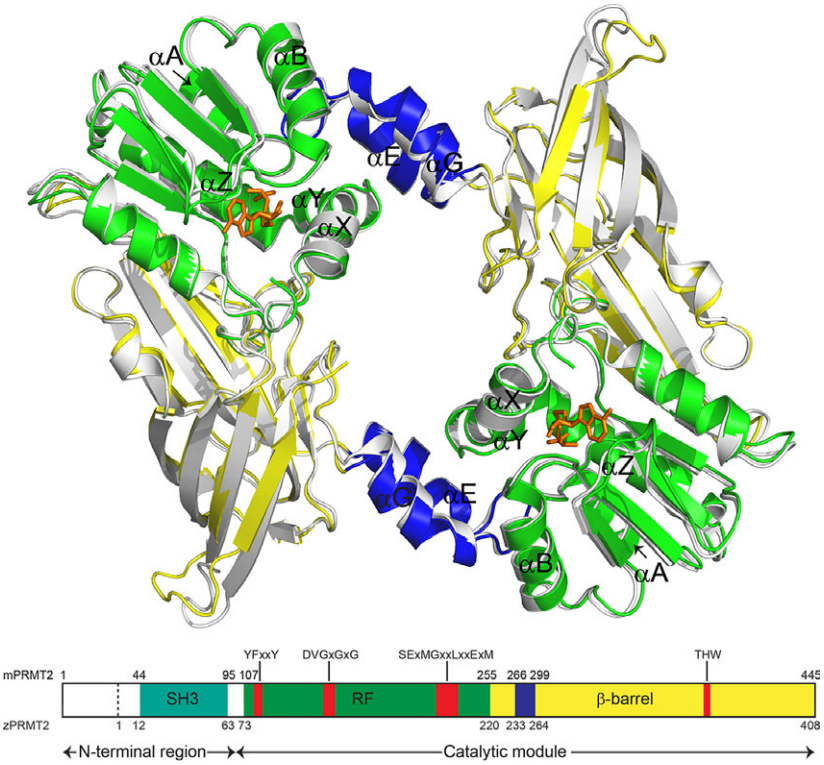


Fig. 3. Superimposed methylation modules of PRMT2 dimers from mouse (in gray) and from zebrafish (in color) and schematic representation illustrating the modular organization of full-length PRMT2. Rossmann fold is shown in green, the beta barrel in yellow, and the dimerization arm in blue. SAH is displayed in orange, as a stick model in the structure. On the scheme, SH3 domain is indicated in cyan and the four motifs in red.

crystallographic twofold axis and hence are structurally identical. The dimeric interface is stabilized by hydrophobic and hydrogen bonds and buries a surface area of 1602 and 1574 $Å^2$ for zPRMT2 and mPRMT2, respectively (calculated using PISA) [41]. The dimer formation involves the dimerization arm from one monomer, namely helices αE and αG and two short 3_{10}-helices F and G′, and helices αY, αZ, αA, and αB from the other monomer and delineates a central hole of ~ 22 × 12 $Å^2$. The orientation of the monomer in the functional dimer observed in zPRMT2 is perfectly conserved in mPRMT2 (Fig. 3). Furthermore, sequence conservation in the dimerization arm suggests that this feature could be common to all PRMT2 orthologs (Fig. 2).

Activity

The *in vitro* methyltransferase activity of recombinant PRMT2 on histones is described as being weak compared to PRMT1, CARM1, and PRMT6, suggesting that the optimal substrate for PRMT2 is probably unknown. A small, but significant, activity has been previously described using the histone H4 as the substrate [21]. *In vivo* evidence has shown that PRMT2 associated with β-catenin deposits an asymmetric dimethyl mark on H3R8 of target gene promoters [23].

Synthetic peptides corresponding to the N-terminal tails of histone H3 (residues 1–34) and H4 (residues 2–22) were used as substrates to test the activity of the purified recombinant mPRMT2. The resulting gels showed incorporation of radioactivity in four different species: weak signal was detected for either $H4_{2–22}$ or $H3_{1–34}$ as expected; a much stronger signal was attributed to PRMT2 likely as a result of automethylation, whereas an additional strong signal was revealed for a minor contaminant with an estimated mass of 16 kDa (Fig. 6). The radioactive signals were all linearly dependent to the concentration in enzyme, confirming that the activity observed is due to PRMT2.

The 16 kDa contaminant copurifying with mPRMT2 was identified by LC-MS/MS mass spectrometry as the repressor splicing factor 1 (RSF1) from the insect expression host *Spodoptera frugiperda*. This splicing repressor antagonizes serine- and arginine-rich (SR) protein function [42] or coregulates alternative splicing with the other SR proteins in drosophila [43]. It has a modular organization with an N-terminal RRM-SRSF3-like RNA-binding domain and a long arginine-/glycine-rich C-terminal part. RSF1 is related to the serine-/arginine-rich (SR) family of splicing regulators, especially with the serine-/arginine-rich splicing factor 7 (SRSF7) involved in pre-mRNA splicing and mRNA export (51% sequence identity in the RRM domain). RSF1 from *S. frugiperda* was expressed

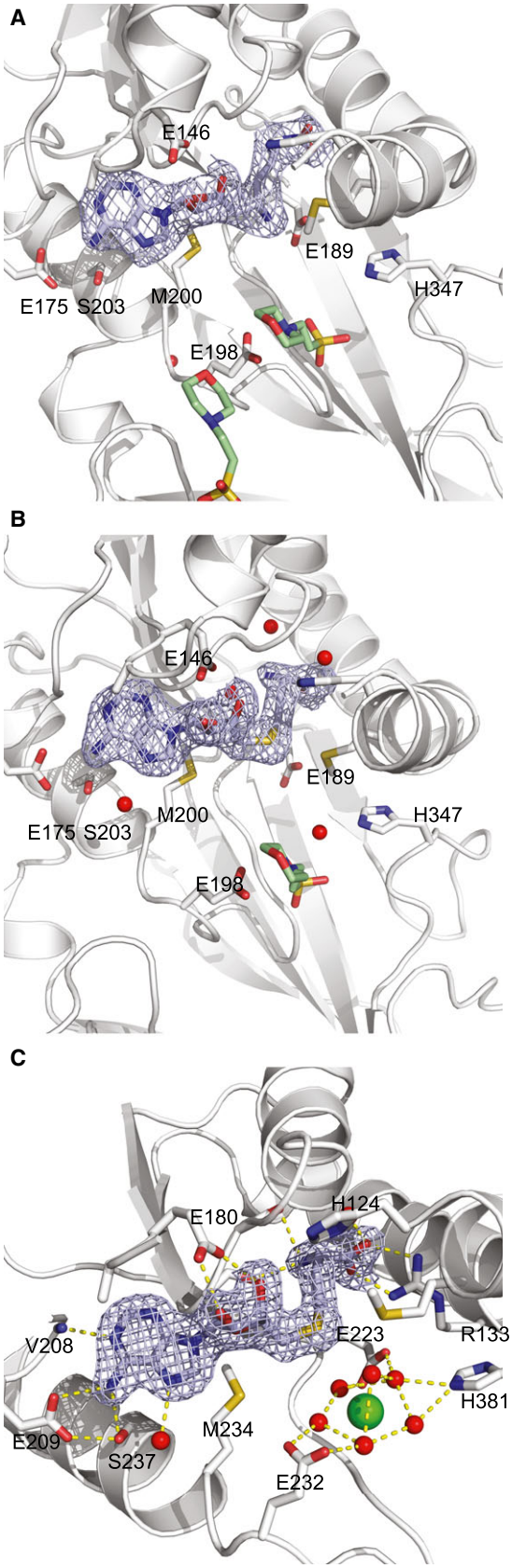


Fig. 4. A detailed view of the amino acid interactions in the active site of zPRMT2 in complex with sinefungin (A), with SAH (B), and mPRMT2 with SAH (C). For clarity, helix X has been removed. SAH or sinefungin are drawn as ball and stick models in light blue. Electron density (2Fobs-Fcalc)-weighted maps around sinefungin and SAH are represented as a mesh, contouring level was contoured at 1σ. MES molecules and calcium ion are shown in green. Water molecules are drawn as red spheres. Hydrogen bonds are shown as yellow dotted lines.

in *Escherichia coli* and purified. To localize the methylation sites, recombinant RSF1 was incubated with SAM and mPRMT2 and analyzed by mass spectrometry (Fig. 7). Six methylation sites were identified on RSF1, four arginine residues are monomethylated (position 24, 100, 111, and 139), and two are dimethylated (position 7 and 120). Interestingly, the two dimethylated arginines are both located in short motifs displaying a unique consensus sequence SxGGxRxY. R7 and R24 belong to the RRM domain of RSF1 (Fig. 8). According to a model based on the structure of SRSF7 (PDB ID 2hvz), these residues would be located in an α helix and in a β strand, respectively, with their side chains exposed at the surface of the domain and therefore accessible to the methyltransferase. The four other methylation sites would be located in the unfolded arginine-rich C-terminal part of RSF1.

Effect of calcium on the activity

A calcium ion was found bound in the bottom of active site pocket of mPRMT2:SAH structure, at a position corresponding to the binding site of the charged guanidino group of the target arginine residue. The hydrated calcium atom is surrounded by the two highly conserved glutamate residues of the 'double-E loop' and other conserved residues of 'THW' loop. This binding site is therefore conserved in all arginine methyltransferases. A calcium atom has also been observed at the same position in PRMT6 [18] and could explain the reduced activity of PRMT7 in the presence of magnesium or calcium [44]. To investigate the effect of calcium on PRMT2, the enzyme was incubated with increasing concentrations of calcium chloride and its activity was quantified (Fig. 9A). These experiments revealed that rather high calcium concentrations are necessary to inhibit the enzyme, as 25% of the initial activity remains in the presence of 10 mM $CaCl_2$. At higher concentrations, the solubility of the enzyme was affected, causing consequently an artifactual drop of activity. These results suggest that calcium is a relatively weak inhibitor of PRMT2. The metal chelator EDTA increased slightly enzyme activity with respect to the positive control, indicating a possible contamination of purified mPRMT2 by divalent cations.

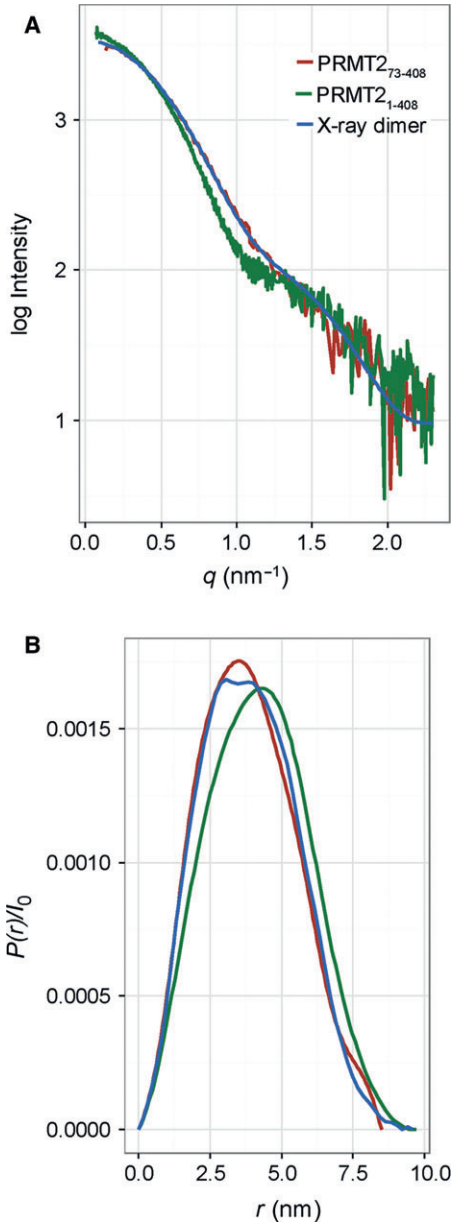


Fig. 5. SAXS data. (A) Comparison of scattering curves and (B) distance distributions $P(r)/I_0$ for zPRMT2$_{1-408}$, zPRMT2$_{73-408}$, and crystallographic structure of zPRMT2$_{73-408}$ (X-ray dimer) in the presence of SAH.

SH3 domain truncation

PRMT2 is characterized by its N-terminal extension partially folded into a SH3 domain. To check whether this module is involved in PRMT2 methyltransferase function, enzymatic activities of zPRMT2$_{73-408}$ and zPRMT2$_{1-408}$ were compared using RSF1 as substrate (Fig. 9B). A sevenfold decrease in activity was detected for the truncated construct with respect to the full-length enzyme. These results reveal that PRMT2 lacking SH3 domain exhibits activity *in vitro* but less than the full-length enzyme, showing the importance of the N-terminal region for the enzymatic function.

Table 2. Small-angle X-ray scattering data. Radius of gyration (R_g) and maximum diameter of the particle (D_{max}) values are indicated for zPRMT2 full-length and catalytic module in the presence of SAH.

	zPRMT2$_{1-408}$ + SAH	zPRMT2$_{73-408}$ + SAH
Protein concentration	2.2 mg·mL^{-1}	2.4 mg·mL^{-1}
R_g (nm) from $P(r)$	3.22 ± 0.005	2.97 ± 0.02
R_g (nm) from Guinier	3.25 ± 0.01	2.90 ± 0.04
D_{max} (nm)	9.00 ± 0.05	8.50 ± 0.05

Inhibitor binding

A set of SAM-based compounds have been recently described as potent inhibitors of PRMTs [45]. These molecules were designed to simultaneously occupy both the adenosine and the peptide substrate-binding sites. Among them, compound 1 (Cp1) has been reported as an inhibitor for PRMT1, PRMT6, and PRMT4/CARM1 with the highest selectivity and potency for CARM1 vs. PRMT1 or PRMT6. It was therefore interesting to evaluate the inhibitor potency of Cp1 on PRMT2 and to compare it to the effects observed on related PRMTs.

Cp1 binding on mPRMT2 was first compared to SAH and sinefungin by thermal shift assay experiments (Fig. 10A). Binding of SAH or sinefungin increased the melting temperature by 5.3 °C ± 1.2 °C and 5.0 °C ± 1.4 °C with respect to the apo protein. Surprisingly, when Cp1 was added to mPRMT2, the melting temperature difference reached 15 °C ± 1.9 °C, suggesting a better affinity of PRMT2 for Cp1 than for SAH or sinefungin.

To further investigate the affinity of Cp1 toward PRMT2, we determined IC_{50} values using RSF1 as a substrate (Fig. 10B). We obtained similar IC_{50} for Cp1 (16.3 ± 3.8 μM) and for SAH (18.3 ± 2.0 μM). These values are much higher than the ones that we obtained for mouse CARM1 (mCARM1) (460 ± 134 nM for Cp1 and 820 ± 35 nM for SAH) which are in good agreement with the published data [45]. The IC_{50} values obtained for PRMT2 with Cp1 are in the order of magnitude of those published for PRMT6 (20.23 ± 8.67 μM) and PRMT1 (11.09 ± 2.77 μM). However, the high potency and the selectivity of Cp1 toward CARM1 were not observed for PRMT2. To better understand structural basis for Cp1 binding, we crystallized and solved the structure of the complexes

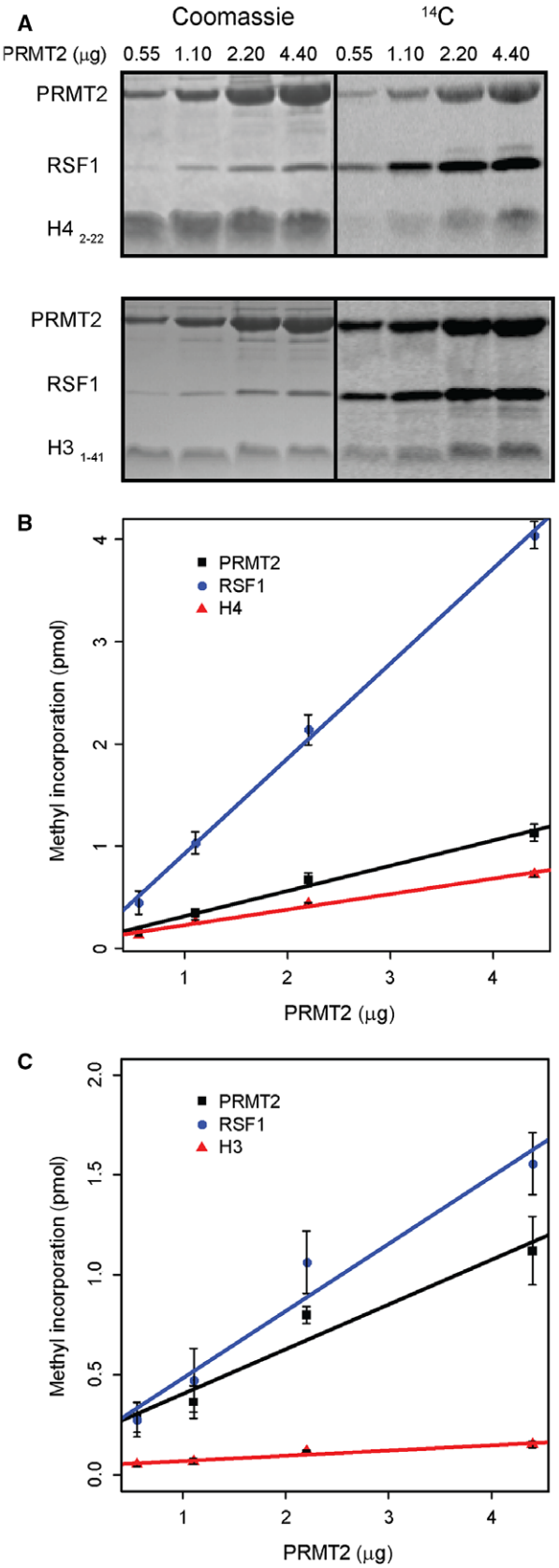


Fig. 6. Methylated product analyses by gel-based assays: (A) Increasing concentrations of mPRMT2 were incubated with $H4_{2–22}$ peptide (upper panel) and $H3_{1–41}$ peptide (lower panel). RSF1 band corresponds to the 16 kDa mPRMT2 contaminant. (B) Curves showing the methylation of the different species with respect to PRMT2 concentration in the presence of $H4_{2–22}$ or (C) $H3_{1–41}$ peptide. The error bars represent SD from experiments run in triplicate.

mPRMT2:Cp1 at 1.8 Å resolution and also mCARM1:Cp1 at 2.25 Å resolution (Table 1).

Structure of mPRMT2:Cp1

Cp1 lacks the methionine moiety typical of all the SAM analogs but harbors a guanidine connected with a spacer to the adenosine moiety (Fig. 11). Structure comparison revealed that the hydrogen bond network observed between the key residues of PRMT2 and the adenosine part of SAH is also well conserved for Cp1 despite a slight conformational change in the ribose cycle. The guanidinium group of Cp1 enters into the arginine pocket where it is maintained by H-bonds and salt bridges between the carboxylates of the motif III residues E223 and E232. The ligand-free methionine pocket is filled by several water molecules. Additional H-bonds occur with the M125 main-chain oxygen and with a water molecule mediating an indirect interaction with the H381 side chain. Thermal shift assays revealed that Cp1 binds more strongly to mPRMT2 than SAH. This result can be explained by the different sets of interactions occurring in the two distinct pockets of the active site, i.e., the methionine pocket for SAH or the arginine pocket for Cp1.

CARM1–PRMT2 Cp1 comparison

Comparing mPRMT2 and mCARM1 structures solved with Cp1 reveals a highly similar positioning of the inhibitor in both the active site pockets. Almost all the key residues involved in Cp1 binding are conserved and their side chains are interacting identically with the inhibitor in both structures (Fig. 11C). However, two differences appear. First, one additional helix W, specific to CARM1 (residues 136–141), packs against the adenine group of Cp1 shielding it from the solvent. This helix, which does not exist in PRMT2, enables an indirect interaction between the guanidino group of residue R141 from CARM1 and N7 from Cp1 via one water molecule. Additionally, F138 forms a hydrophobic interaction with F151, closing the pocket on the adenine ring. Second, in mPRMT2, the histidine residue H124 can form a hydrogen bond with the O3′

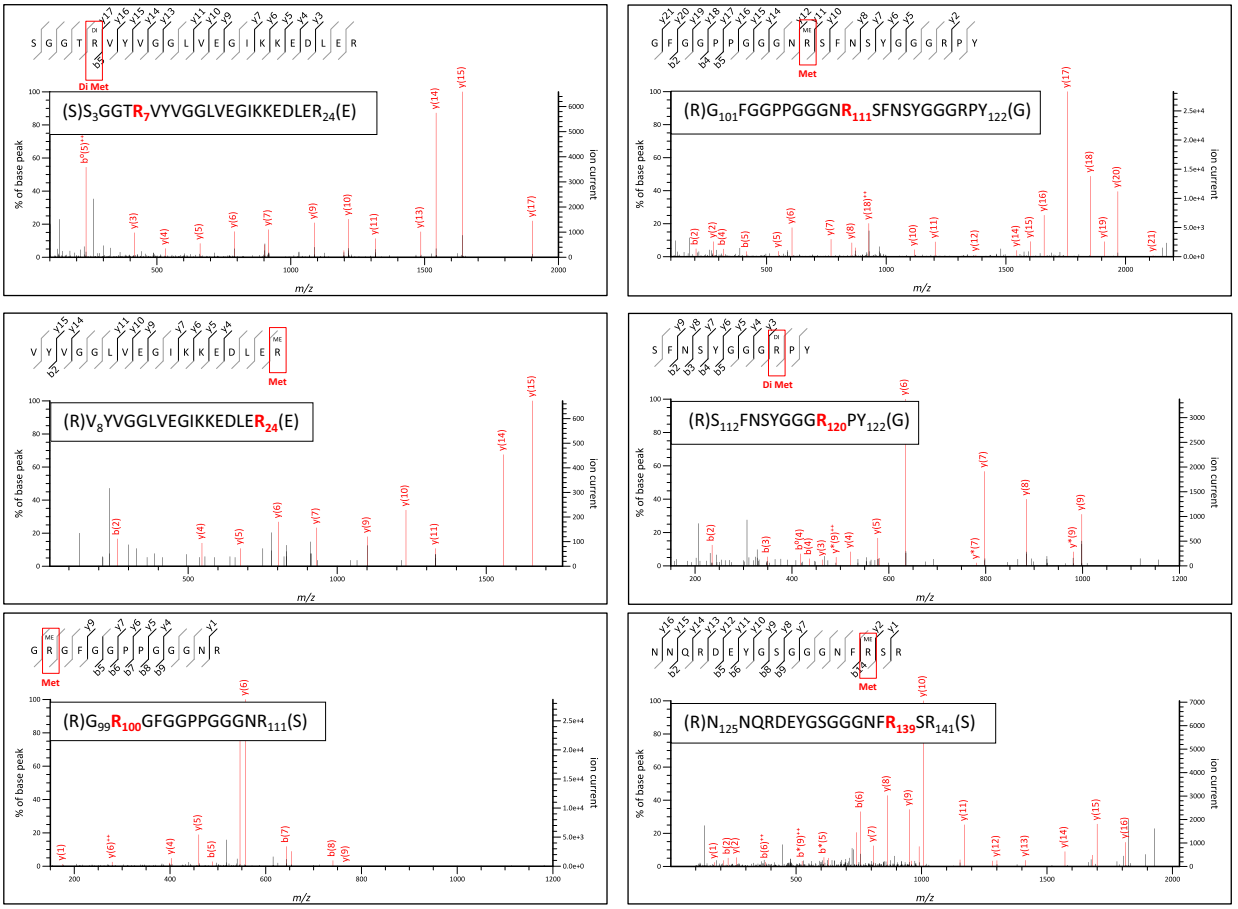


Fig. 7. Analysis of RSF1 methylation by mass spectrometry. See main text for details.

ribose hydroxyl group of Cp1. This residue is replaced by a glutamine (E160) in mCARM1 and the distance to Cp1 is thus too far to be compatible with a hydrogen bond. It is interesting to note that this histidine residue is conserved among almost all PRMT2 sequences known with a few exceptions in which cases it is changed into a glutamine. This histidine residue is also conserved in other PRMTs including PRMT1 and PRMT6. It is probable that these discrepancies impact the effect of Cp1 on the different PRMTs. The high potency and selectivity of Cp1 for CARM1 could be the result of its specific helix W which reinforces the interactions with the inhibitor.

Discussion

The crystal structures of PRMT2 from zebrafish and from mouse revealed a highly conserved dimeric architecture formed by two typical type I PRMT catalytic modules. The global shape of the functional dimer is a consequence of variations in relative orientations of the two monomers. It varies significantly in the structures of different PRMTs and is linked to active site accessibility and therefore substrate recognition [10,16]. The N-terminal module of mPRMT2 (residues 1–106) partially folded in a SH3 domain is missing in the electron density although present in the crystal. A similar case has already been mentioned in mCARM1 [8], where the PH domain assumes different positions in the crystal structure, behaving as a wobbly domain. This behavior is the consequence of the lack of direct interaction between the N-terminal domain and the catalytic module as well as of the flexibility of the connection between functional modules. SH3 motifs are known to interact to proline-rich sequences in their binding partner, mediating the assembly of large multiprotein complexes. The wobbly SH3 domain would probably be stabilized in its functional position in the presence of partners or PRMT2 substrates.

It is well established that in PRMTs, the methyltransferase function is exclusively borne by the catalytic domain, containing both SAM- and arginine-binding sites. However, the presence of additional domains can strongly affect enzymatic activity, as

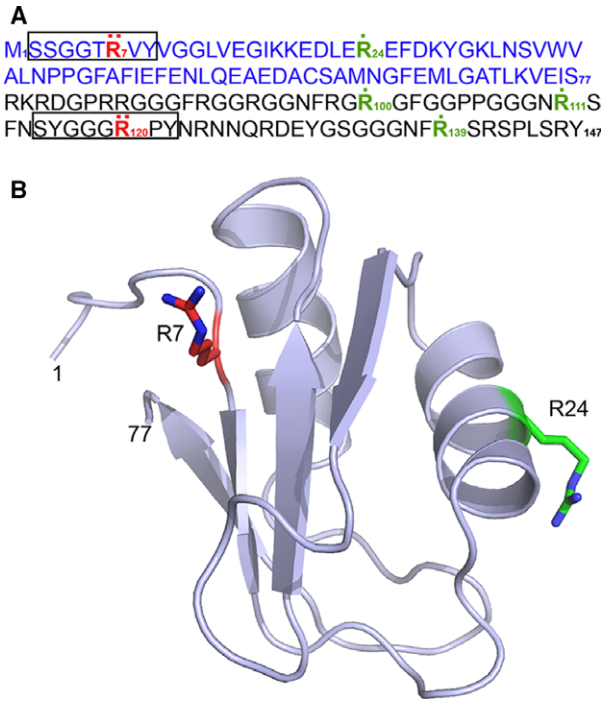


Fig. 8. Methylation sites on RSF1 sequence and model of the RRM domain. (A) Peptidic sequence of *S. frugiperda* RSF1. The folded part corresponding to the RRM domain is shown in blue. Mono- (˙) and dimethylated (¨) arginines are in bold. Consensus sequence SxGGxRxY is boxed. (B) 3D-Model of RSF1 RRM domain generated by Swiss-Model [71]. The side chains of the methylated arginine residues are shown in stick.

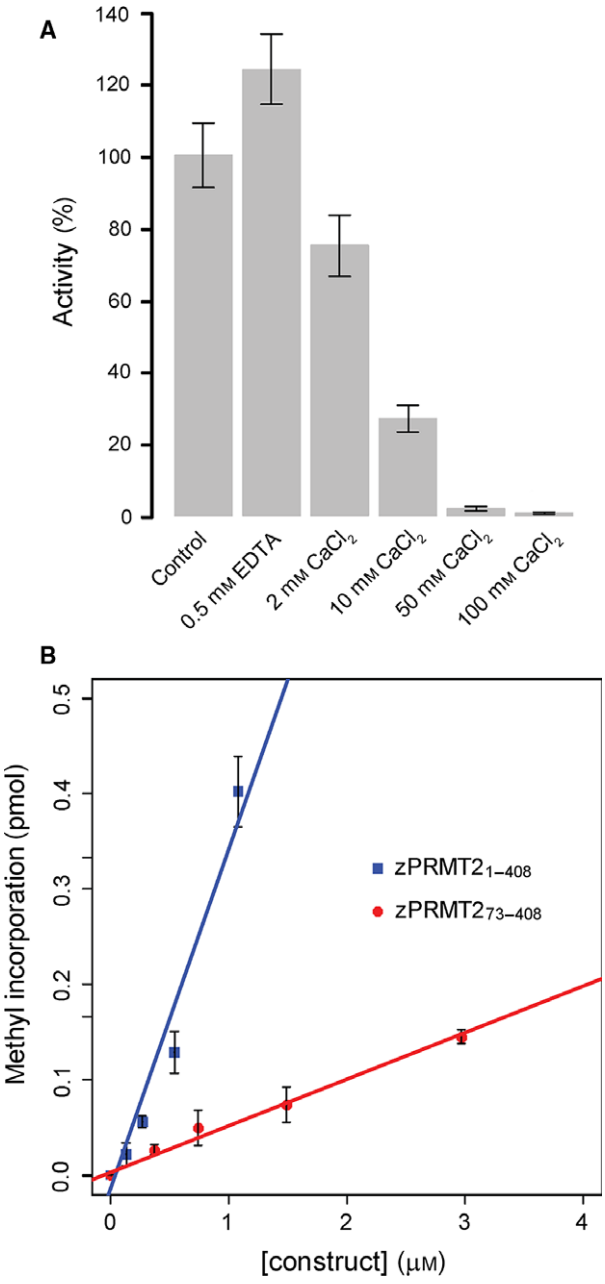


Fig. 9. Methyltransferase activity quantified by gel-based assays: (A) Effect of calcium on mPRMT2 activity. Relative activity was quantified using recombinant RSF1 as a substrate and 1 μM $mPRMT2_{1\text{-}445}$. The positive control was performed with no additive and used as a reference. (B) Effect of SH3 domain deletion on zPRMT2 activity. Curves comparing the methyltransferase activity of $zPRMT2_{1\text{-}408}$ (blue) and $zPRMT2_{73\text{-}408}$ (red). The error bars represent the SD of three sets of experiments.

shown for PRMT5 [46] or PRMT8 [47]. We showed that it is also the case for PRMT2 as N-terminal-truncated enzyme displays reduced methyltransferase activity. Our structural results indicate that this truncation does not impair enzyme stability or dimerization. Therefore, the SH3 domain and/or adjacent sequences belonging to the N-terminal extension of the catalytic module may be also involved in substrate binding or recognition through long distance interactions.

We could identify RSF1, a host-cell contaminant coeluting with PRMT2 and furthermore methylated by the enzyme. Although RSF1 is not present in mouse and PRMT2 has no ortholog in insects, the interaction observed between an arginine methyl transferase and a protein involved in RNA splicing is not surprising. Indeed, arginine methylation is a post-translational modification that occurs in RNA-binding proteins including the spliceosome and SR proteins [48]. It has been shown that PRMT5 catalyzes sDMA formation on the Sm proteins [49] and CARM1 methylates three splicing factors: SmB, U1-C, and SAP49 [50]. Furthermore, methylation of SR proteins by PRMTs has been observed: SF2/ASF and its close paralog SRp30c both belonging to the SR family are methylated by PRMT1 [51,52]. More recently, it has been shown that human PRMT9 symmetrically dimethylates arginine residues on splicing factor SF3B2 [35]. Taken together, these

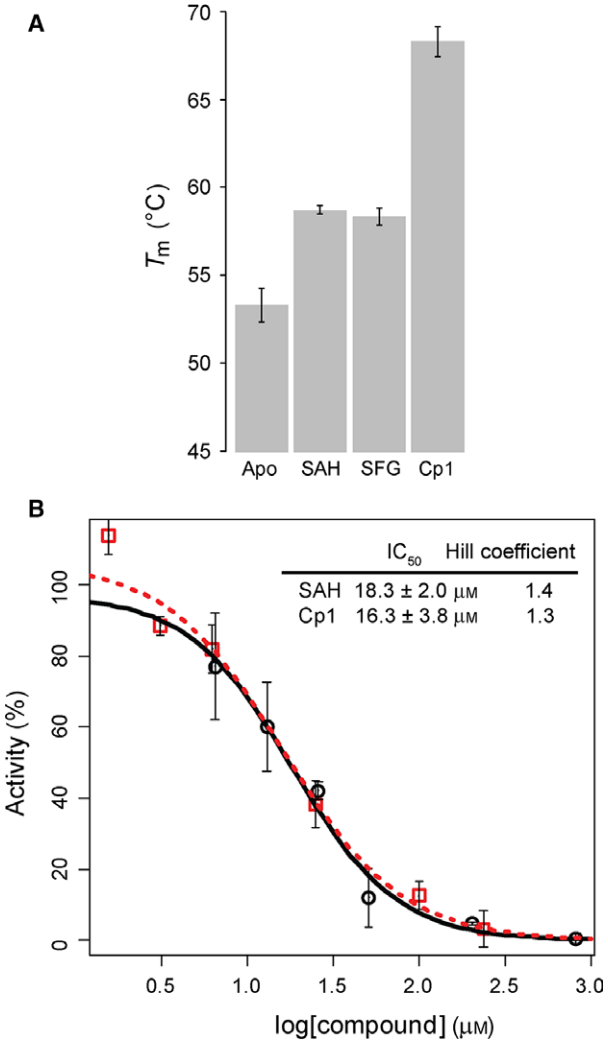


Fig. 10. Effect of inhibitor binding. (A) Melting temperatures (T_m) determined by thermal shift assay for mPRMT2 alone (apo) or incubated with 1 mM SAH, sinefungin (SFG), and Cp1. (B) Dose–response curves for SAH (circles, plain line) and Cp1 (square, red dotted line) at 1.2 µM of mPRMT2. The error bars represent the SD of three sets of experiments.

results and our initial findings suggest that substrates for PRMT2 could be found in the SR family. We tried to methylate *in vitro* mouse SRSF7 expressed and purified from insect cell culture (data not shown). This preliminary attempt was unsuccessful. Nevertheless, it is not excluded that other members of SR protein family are PRMT2 methylation targets. So far, PRMT2 has not been mentioned to interact with splicing regulators or RNA-binding proteins but this possibility needs to be further investigated.

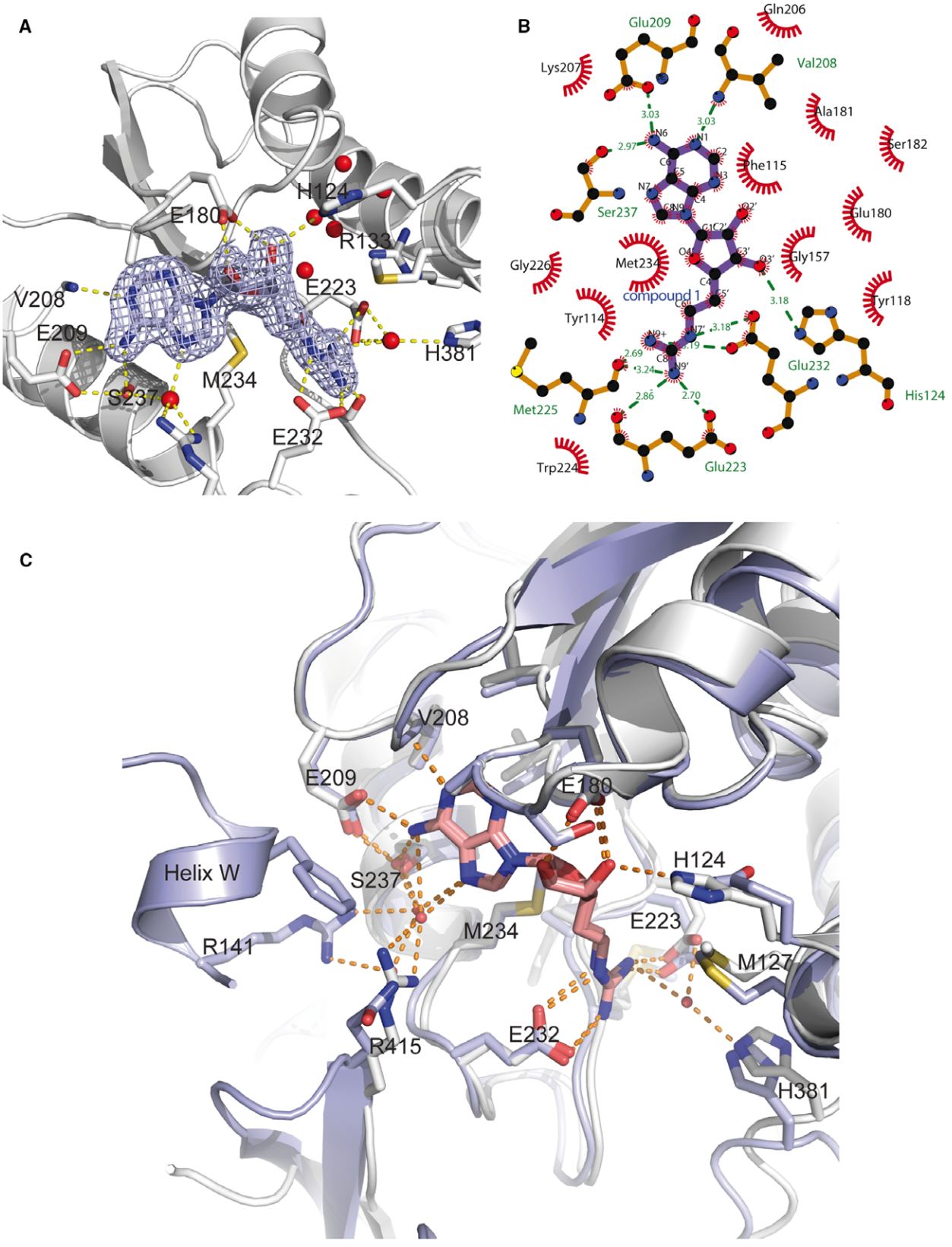
The crystal structure of mPRMT2 in complex with Cp1, a specific CARM1 inhibitor, has been solved and analyzed. The IC_{50} values obtained for Cp1 with mPRMT2 were higher than with CARM1, revealing a weaker affinity for mPRMT2. Nevertheless, Cp1 inhibits mPRMT2 slightly more efficiently than SAH and can be then considered as the first synthetic PRMT2 inhibitor studied so far. Altogether, these results are a first step toward a better understanding of PRMT2 substrate recognition and may accelerate the development of structure-based drug design of PRMT2 inhibitors.

Material and methods

PRMT2 cloning, expression, and purification

The *D. rerio prmt2* gene sequence (Genbank BC129171.1) corresponding to the full-length protein and to the PRMT core (residues 73–408) were amplified by PCR from a cDNA clone (Source BioScience, Nottingham, UK; IRAK293-C16). The *M. musculus prmt2* full-length gene sequence (Genbank BC122563.1) was amplified by PCR from a cDNA clone (Source BioScience IRCL32A04). The sequences were cloned in the pDONR207™ (Fisher Scientific, Illkirch, France) vector using a BP reaction (Gateway® Cloning; Thermo Fisher Scientific, Waltham, MA, USA). The positive clones were confirmed by sequencing (GATC BIOTECH, Konstance, Germany). The sequences were subcloned in a pDEST20™ vector (for *D. rerio prmt2*) or a pBacGGWA vector [53] (for *M. musculus prmt2*) using a LR reaction. DH10Bac competent cells containing the baculovirus genome were transformed with the pDEST20™-PRMT2 plasmids and plated onto LB agar media containing 15 $mg{\cdot}mL^{-1}$ tetracycline, 7 $mg{\cdot}mL^{-1}$ gentamicin, 50 $mg{\cdot}mL^{-1}$ kanamycin, 25 $mg{\cdot}mL^{-1}$ X-Gal, and 40 $mg{\cdot}mL^{-1}$ isopropyl thio-β-d-galactoside (IPTG). Bacmid DNA purified from recombination-positive white colonies was transfected into Sf21 cells using the Lipofectin reagent (Fisher Scientific). Viruses were harvested 10 days after transfections. The resulting recombinant proteins harbor an amino-terminal glutathione S-transferase (GST) tag followed by a Tobacco etch virus (TEV) protease cleavage site.

Fig. 11. Overview of Cp1-binding site on mPRMT2 and mCARM1. (A) View of Cp1 interactions in mPRMT2 SAM- and arginine-binding pockets. Electron density (2Fobs-Fcalc)-weighted maps around Cp1, represented as a mesh, was contoured at 1σ. (B) Flat representation done with LigPlot [72] showing contacts between Cp1 (compound 1) and active site residues in mPRMT2. The ligand bonds are colored in purple. Hydrogen bonds are shown as dotted lines. The spoked arcs represent protein residues making nonbonded contacts with Cp1. (C) Superimposition of the mPRMT2 (gray) and mCARM1 (purple) active sites. Cp1 is drawn in pink in both structures. Numbered residues belong to mPRMT2 except R141 belonging to mCARM1. Helix W specific to mCARM1 is also shown.


A
H124
E180
R133
E223
V208
E209
H381
M234
S237
E232
B
Gln206
Glu209
Lys207
Val208
Ala181
Ser182
Phe115
Ser237
Glu180
Gly157
Met234
Gly226
compound 1
Tyr118
Tyr114
Glu232
His124
Met225
Trp224
Glu223
C
V208
E209
E180
Helix W
S237
H124
E223
M234
R141
M127
E232
R415
H381

mPRMT2 also harbors a R445W point mutation due to a mutation in the PCR product. Sf21 cells were grown at 27 °C in suspension culture in SF900II (Thermo Fisher Scientific) using Erlenmeyer flasks. Two liter of Sf21 cell culture (at 0.8×10^6 cells·mL^{-1}) was infected with recombinant GST-zPRMT2 viruses with an infection multiplicity of 1. Cells were harvested 48 h postinfection. Cells lysis was performed by sonication in 100 mL buffer *A* [50 mM Tris/HCl pH 8.0, 250 mM NaCl, 5% glycerol, 10 mM DTT, 0.01% NP40, and anti-proteases (Roche, Basel, Switzerland, Complete™, EDTA-free)] and cellular debris were removed after centrifugation of the lysate at 40 000 ***g*** for 30 min. The supernatant was incubated at 4 °C for 4 h with 3 mL glutathione Sepharose 4B resin (GE Healthcare, Velizy, France). After a short centrifugation, the supernatants were discarded and the beads were poured in an Econo-column (Bio-Rad, Hercules, CA, USA). After two wash steps with 10 mL buffer *A*, 3 mL buffer *A* supplemented with in-house produced TEV protease were applied to the columns and digestion was performed overnight at 22 °C with gentle mixing. The digests were concentrated with an Amicon Ultra 10K (Merck Millipore, Darmstadt, Germany), loaded on a gel-filtration column (HiLoad 16/60 Superdex S200, GE Healthcare) and eluted at 1 mL·min^{-1} with buffer *B* [20 mM Tris/HCl pH 8.0, 75 mM NaCl, 5 mM TCEP and 1 mM EDTA] for zPRMT2, using an ÄKTA Explorer device (GE Healthcare). Fractions containing zPRMT2 were pooled and concentrated to 17 mg·mL^{-1}. mPRMT2 and zPRMT2$_{73\text{-}408}$ were purified from baculovirus-infected cells using the same protocol with the following modifications: the crude extract was incubated with glutathione sepharose overnight at 4 °C, TEV digestion was performed during 4 h at 30 °C and buffer C was used for gel-filtration (20 mM Tris/HCl pH 8.0, 100 mM NaCl , and 5 mM TCEP). Fractions containing mPRMT2 were pooled and concentrated to 3.5 mg·mL^{-1}. Fractions containing zPRMT2$_{73\text{-}408}$ were pooled and concentrated to 14 mg·mL^{-1}.

PRMT2 crystallization

Crystallization conditions were screened using commercially or home-made available kits by the sitting-drop vapor-diffusion method in 96-well MRC2 plates (Swissci, Zug, Switzerland), and employing a Mosquito robot (TTP Labtech, Melbourn, UK) dispensing droplets made of 100 nL protein solution and 100 nL reservoir solution. The plates were placed in a Rock Imager (Formulatrix, Bedford, MA, USA) at 20 °C and monitored periodically.

The zPRMT2 crystal used for X-ray diffraction was obtained in a 96-well MRC2 plates in sitting drops made with a zPRMT2 sample at 17 mg·mL^{-1}, mixed with 1 mM sinefungin, and a reservoir solution containing 100 mM MES pH 6.0, 9% (w/v) PEG 20 000, and 100 mM LiCl. Single diffracting crystals grew to $120 \times 100 \times 20$ μm within 4 weeks.

The zPRMT2$_{73\text{-}408}$ crystal used for X-ray diffraction was obtained in a 96-well MRC2 plates in sitting drops made with a zPRMT2 sample at 14 mg·mL^{-1}, mixed with 1 mM SAH, and a reservoir solution containing 100 mM MES pH 6.0, 9% (w/v) PEG 20 000, and 300 mM NaCl. Single diffracting crystals grew to $50 \times 40 \times 20$ μm within 3 days.

First hits for mPRMT2 were obtained in several conditions of the PACT (Molecular Dimensions, Newmarket, UK) and The Peg Suite (Qiagen, Hilden, Germany) crystallization screens. Crystallization conditions were further optimized in 24-well VDX plates (Hampton Research, Aliso Viejo, CA, USA). Crystals used for X-ray diffraction were obtained within days at 20 °C in a hanging drop made with 1 μL mPRMT2 at 3.5 mg·mL^{-1}, mixed with 1 mM inhibitor, and 1 μL reservoir solution containing 16% (w/v) PEG 6000, 100 mM Hepes-NaOH pH 7.0 and 100 mM $CaCl_2$.

Data collection and structure solution

zPRMT2 crystals were soaked in the reservoir solution supplemented with 15–20% (v/v) PEG 400 or 20% (v/v) ethylene glycol and flash-cooled in liquid nitrogen.

The diffraction datasets were collected on the SOLEIL PROXIMA1 and PROXIMA2 beamlines, using a PILATUS 6M (DECTRIS, Baden-Dättwil, Switzerland) or a Quantum 315 (ADSC, Poway, CA, USA) detector and processed with XDS [54], Imosflm [55], and HKL-2000 [56]. The crystals belonged to the *H32* space group with one zPRMT2 molecule in the asymmetric unit. The dataset was submitted to the BALBES server for automated molecular replacement at York Structural Biology Laboratory [57] after unsuccessful manual attempts with known PRMT structures or models. An initial solution with a BALBES template based on CARM1 from *Homo sapiens* (PDB ID 2y1w) and the sequence of *z*PRMT2 was obtained with one monomer in the asymmetric unit.

mPRMT2 crystals were cryoprotected by a short soak in 20% (w/v) PEG 6000, 100 mM Hepes pH 7.0, 100 mM $CaCl_2$, 15% (v/v) PEG 400, and flash-cooled in liquid nitrogen for data collection. Diffraction datasets were collected on the ID23-2 beamline at the ESRF using a PILATUS 2M detector (DECTRIS) and processed with XDS. The crystal belonged to the $C222_1$ space group with one mPRMT2 molecule in the asymmetric unit. The structure was solved by molecular replacement using PHASER and zPRMT2 structure as a starting model.

Iterative cycles of model building and refinement were carried out using COOT [58], PHENIX [59], and BUSTER [60]. TLS refinement with four groups per polypeptide chain was used. All other crystallographic calculations were carried out with the CCP4 package [61]. A summary of the data statistics for the refined structure is provided in Table 1. The atomic coordinates and experimental data (PDB ID 5g02, 5fub, 5ful, 5fwa) have been deposited in the Protein Data Bank.

CARM1 purification, crystallization, and structure determination

A construct of mouse CARM1 from residues 130 to 487 ($mCARM1_{130-487}$) was expressed as a GST fusion protein using the baculovirus system, purified as previously described [62] and crystallized in 15% (w/v) PEG 2000MME, 100 mM Tris/HCl pH 8.0, 100 mM NaCl in the presence of 1 mM Cp1 at 20 °C. mCARM1 crystals were cryoprotected by a short soak in 22% (w/v) PEG 2000MME, 100 mM Tris/HCl pH 8.0, 100 mM NaCl, 15% (v/v) PEG 400, and flash-cooled in liquid nitrogen. Diffraction datasets were collected on SOLEIL PROXIMA2 beamline, using a Quantum 315 (ADSC) detector and processed with XDS. The structure was solved and refined using PHENIX. Model building and refinement were carried out using COOT [58], PHENIX [59], and BUSTER [60]. TLS refinement with three groups per polypeptide chain was used. All other crystallographic calculations were carried out with the CCP4 package [61]. A summary of the data statistics for the refined structure is provided in Table 1. The atomic coordinates and experimental data (PDB ID 5k8v) have been deposited in the Protein Data Bank.

RSF1 expression and purification

A synthetic gene encoding RSF1 from *S. frugiperda* and codon-optimized for the expression in *E. coli* (Eurofins Genomics, Ebersberg, Germany) was inserted into the *Nde*I and *Bam*HI sites of a pET-15b vector (Novagen, Madison, WI, USA). RSF1 was overexpressed in *E. coli* BL21 (DE3) as a N-terminal His-tagged fusion protein in 1 L LB/ampicilin medium. Protein expression was induced for 3 h at 37 °C by the addition of 0.4 mM IPTG when A600 nm reached 0.6. Harvested cells were suspended in 20 mM Tris/HCl pH 8.0, 250 mM NaCl and lysed by sonication. The lysate was centrifuged for 30 min at 40 000 ***g*** and the crude extract was applied to a Ni-NTA Superflow Agarose (Qiagen) column equilibrated in the same buffer. His6-RSF1 was eluted with 250 mM imidazole and concentrated at 0.9 $mg{\cdot}mL^{-1}$ in 20 mM Tris/HCl pH 8.0, 100 mM NaCl, and 0.5 mM DTT.

Gel-based activity assays

Methylation assays were performed at 23 °C in 20 mM Tris/HCl pH 8, 100 mM NaCl, 0.5 mM DTT. PRMT2 at varying concentrations was incubated for 3–5 h with a substrate (500 μM histone H3 or H4 tail peptides, or 5 μM RSF1) and 0.02 mM S-adenosyl-L-[Me-^{14}C]methionine (^{14}C-SAM) at 50 $mCi{\cdot}mmol^{-1}$ (PerkinElmer, Richmond, VA, USA). The reaction was stopped by addition of SDS loading buffer 4× and analyzed on 16% Tricine/SDS/PAGE gel. After electrophoresis, the gel was stained with Coomassie brilliant blue, dried on a Whatman 3MM paper sheet and exposed for 48 h against a BAS-MS 2025 imaging plate (GE Healthcare). The amount of radioactivity incorporated by the substrate was quantified by phosphorimager analysis (Typhoon; GE Healthcare). The amount of arginine methylation was deduced from the radioactive signal using spots of ^{14}C-SAM as a reference.

IC_{50} determination

All the SAM analogs were dissolved in water and their concentration was determined by UV absorption (ε_{260} = 15 400 $M^{-1}{\cdot}cm^{-1}$). Inhibition assays were adapted from the gel-based activity assay described above: Solutions containing either 10 μM RSF1 and 1.2 μM mPRMT2 in 20 mM Tris/HCl pH 8.0, 100 mM NaCl, 0.5 mM DTT or 0.8 μM mCARM1, 10 μM H3 peptide in 50 mM Tris/HCl pH 8, 200 mM NaCl, 0.5 mM DTT were incubated at 23 °C with increasing concentrations of SAM analogs (0.1–500 μM). The reaction was initiated by adding 20 μM ^{14}C-labeled SAM and stopped after 4 h for PRMT2 or 30 min in the case of CARM1 assays. IC_{50} values were determined by fitting the data for sigmoidal dose–response curves using the R version 2.13.0 software package [63].

Thermal shift assay

Thermal shift assay experiments were performed on a MiniOpticon Real-Time PCR System MJ Mini (Bio-Rad). The protein sample (4 μM) was incubated in 20 mM Tris/HCl pH 8.0, 100 mM NaCl with 1 mM of potential ligands in a total volume of 23 μL for 30 min at room temperature. The samples were heated from 20 °C to 94 °C at 0.1 $°C{\cdot}s^{-1}$ after addition of 7 μL Sypro Orange 25× (Thermo Fisher Scientific). The fluorescence intensity was plotted as a function of temperature using the MJ OPTICON MONITOR ANALYSIS software. Melting temperature (T_m) was given by the inflection point of the fluorescence curve.

SAXS experiments and data processing

$zPRMT2_{73-408}$ SAXS data were collected on a BioSAXS-1000 camera (Rigaku, Ettlingen, Germany) mounted on a MicroMax 007 HF Rigaku rotating anode at a sample-detector distance of 500 mm, covering the range of momentum transfer $0.008 < q < 0.65$ $Å^{-1}$ ($q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 0.154$ nm is the X-ray wavelength) in four frames (30 min each) to check for possible radiation damage. The q scale was calibrated by silver behenate powder diffraction and all data were collected up to a maximum q of 0.35 $Å^{-1}$. Scattering measurements were carried out at 20 °C in a thermostated quartz capillary with a diameter of 1.5 mm. zPRMT2 full-length SAXS data were collected at the SAXS beamline P12 at the PETRA III storage ring (Deutsches Elektronen-Synchrotron, Hamburg, Germany). The exposure time was 0.045 s.

All zPRMT2 samples were solubilized in 20 mM Tris/HCl pH 8.0, 100 mM NaCl, 5 mM TCEP, 1 mM SAH and measured at several concentrations. The SAXS data were processed by standard procedures using the ATSAS software package (version 2.7.2). The scattering intensity $I(0)$ and the radius of gyration R_g were evaluated using the Guinier approximation with PRIMUS [64]. These parameters were also determined from the entire scattering pattern using GNOM which computes the distance distribution function $P(r)$ and the maximum particle dimension D_{max} [65]. The program CRYSOL was used to fit the theoretical scattering pattern to the experimental data [66]. For *ab initio* model construction, multiple runs were performed using the program DAMMIF [67] to verify the stability of the solution. Ten models were averaged using DAMAVER [68] to produce a starting model for a second modeling step using DAMMIN. The final model was superimposed to the X-ray structure using SUPCOMB [69].

Mass spectrometry analysis

Protein preparation for liquid digestion

The Rapid Gest was reconstituted with 1 mL of 25 mM NH_4HCO_3. Eighty-five microliter of the Rapid Gest solution was added to 15 μL of sample. The reduction was done with 5 mM DTT (final concentration) in 25 mM NH_4HCO_3 (1 h at 57 °C) and alkylation was performed with 15 mM iodoacetamide (final concentration) in 25 mM NH_4HCO_3 (30 min in the dark). The trypsin (Promega, Madison, WI, USA; V511A) was suspended in 100 μL of 25 mM NH_4HCO_3. Five microliter was added to the sample and incubated overnight at 37 °C. The digestion was stopped by the addition of 1 μL of TFA.

A second digestion was performed with the pepsin. The pepsin was suspended at 1 mg·mL^{-1} in water at pH 1. Five microliter was added to the sample and incubated 3 h at 37 °C. The digestion was stopped by heating the sample at 90 °C during 5 min. The sample was centrifuged 5 min at 10 000 ***g*** and the supernatant analyses by nanoLC-MS/MS.

Chromatography conditions on NanoAcquity

The analysis was performed on a nanoACQUITY Ultra-Performance-LC (UPLC; Waters, Milford, MA, USA). The samples were trapped on a 20 × 0.18 mm, 5 μm Symmetry C18 precolumn (Waters Corp.), and the peptides were separated on a ACQUITY UPLC® BEH130 C18 column (Waters Corp.), 75 μm × 250 mm, 1.7 μm particle size. The solvent system consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Trapping was performed during 3 min at 5 μL·min^{-1} with 99% of solvent A and 1% of solvent B. Elution was performed at a flow rate of 400 nL·min^{-1}, using 1–40% gradient (solvent B) over 35 min at 45 °C followed by 65% (solvent B) over 5 min.

MS and MS/MS conditions on SYNAPT mass spectrometer

The MS and MS/MS analyses were performed on the SYNAPT™, a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Waters) equipped with a *Z*-spray ion source and a lock-mass system. The capillary voltage was set at 3.5 kV and the cone voltage at 35 V. Mass calibration of the TOF was achieved using Glu-fibrino-peptide B fragments obtained at 27 V on trap cell on the (50–2000) m/z range in positive mode. Online correction of this calibration was performed with Glu-fibrino-peptide B as the lock-mass. The ion $(M + 2H)^{2+}$ at m/z 785.8426 is used to calibrate MS data and the fragment ion $(M + H)^{+}$ at m/z 684.3469 is used to calibrate MS/MS data during the analysis.

For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes (MS 0.5 s/scan on m/z range [250–1500] and MS/MS 0.7 s/scan on m/z range [50–2000]). The three most abundant peptides (intensity threshold 60 counts·s^{-1}), preferably doubly and triply charged ions, were selected on each MS spectrum for further isolation and CID fragmentation with two energies set using collision energy profile. Fragmentation was performed using argon as the collision gas. The complete system was fully controlled by MassLynx 4.1 (SCN 566; Waters). Raw data collected during nanoLC-MS/MS analysis were processed and converted with ProteinLynx Browser 2.5 (Waters) into. pkl peak list format. Normal background subtraction type was used for both MS and MS/MS with 5% threshold and polynomial correction of order 5, and deisotoping was performed.

MS and MS/MS conditions on TripleTOF 5600+ mass spectrometer

The MS and MS/MS analyses were performed on the TripleTOF 5600, a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (AB Sciex, Villebon, France). The mass spectrometer was operated in positive mode, with the following settings: ion spray voltage floating (ISVF) 2300 V, curtain gas (CUR) 25 psi, interface heater temperature (IHT) 75 °C, ion source gas 1 (GS1) 2 psi, declustering potential (DP) 100 V. Information-dependent acquisition (IDA) mode was used with top-five MS/MS scans. The MS scan had an accumulation time of 250 ms on m/z (400–1250) range and the MS/MS scans had 100 ms m/z (150–1600) range in high sensitivity mode. Switching criteria were set to ions with charge state of 2–4 and an abundance threshold of more than 150 counts, exclusion time was set at 12 s. IDA rolling collision energy script was used for automatically adapting the CE. Mass calibration of the analyzer was achieved using peptides from digested BSA. The complete system was fully controlled by AnalystTF 1.6 (AB Sciex).

MS/MS data interpretation

The peak list has been searched against a Swiss Prot-derived combined target-decoy database using MASCOT (version 2.4.1; Matrix Science, London, UK). The database contained sequences of human proteins including common contaminants (human keratins) and porcine trypsin and was created using an in-house database generation toolbox (http://msda.u-strasbg.fr). During database search, up to one missed cleavage by trypsin and four variable modifications [oxidation of methionine (+16 Da), carbamidomethylation of cysteine (+57 Da), and methylation or dimethylation on arginine] were considered. The search window was set to 15 ppm for precursor ions and 0.05 Da for fragment ions. MASCOT result files (.dat) were imported into SCAFFOLD 3 software (version 3.6.5; Proteome Software Inc., Portland, OR, USA) and filtering criteria based on probability-based scoring of the identified peptides were taken into account: peptides having an (Mascot ion score minus identity score) above 0 and a Mascot Ion score above 30 were validated. The false discovery rate (FDR) was calculated to be < 1% based on the number of decoy hits.

Acknowledgements

This work was supported by grants from CNRS, Université de Strasbourg, INSERM, Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) supported by national member subscriptions as well as the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01, grant ANR-10-LABX-0030-INRT), a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir labeled ANR-10-IDEX-0002-02 to JC and IGBMC, grants from the French Proteomic Infrastructure (ANR-10-INBS-08; ProFI project) to SC and grants from Association pour la Recherche contre le Cancer (ARC) (No. A09/4/5005, No. SFI20121205902). We thank the members of the IGBMC common services and the members of the structural biology platform of IGBMC for technical assistance. We thank Alastair McEwen for help during data collection at synchrotron sites. We thank Johann Stojko for help during MS analysis. We thank Natacha Rochel for help during SAXS data collection and PETRA III P12 beamline staff for technical assistance during data collection at the Deutsches Elektronen-Synchrotron. We thank members of SOLEIL Proxima1/Proxima2 beamlines and the European Synchrotron Radiation Facility–European Molecular Biology Laboratory joint Structural Biology groups for the use of beamlines facilities and for help during X-ray data collection.

Author contributions

VC, NM, NT-C performed the biochemical and biophysical experiments. VC, NM, LB, and JC performed the structure determination by X-ray crystallography. JMS and SC performed and analyzed the mass spectrometry experiments. MJH and NIM provided the cp1 compounds. VC, LB, SC, and JC wrote the article. JC directed the study.

References

1 Bedford M & Clarke S (2009) Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33**, 1–13.
2 Lee Y-H & Stallcup MR (2009) Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. *Mol Endocrinol* **23**, 425–433.
3 Yang Y & Bedford MT (2013) Protein arginine methyltransferases and cancer. *Nat Rev Cancer* **13**, 37–50.
4 Niewmierzycka A & Clarke S (1999) S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J Biol Chem* **274**, 814–824.
5 Weiss VH, McBride AE, Soriano MA, Filman DJ, Silver PA & Hogle JM (2000) The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nat Struct Biol* **7**, 1165–1171.
6 Zhang X, Zhou L & Cheng X (2000) Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J* **19**, 3509–3519.
7 Zhang X & Cheng X (2003) Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure* **11**, 509–520.
8 Troffer-Charlier N, Cura V, Hassenboehler P, Moras D & Cavarelli J (2007) Functional insights from structures of coactivator-associated arginine methyltransferase 1 domains. *EMBO J* **26**, 4391–4401.
9 Yue WW, Hassler M, Roe SM, Thompson-Vale V & Pearl LH (2007) Insights into histone code syntax from structural and biochemical studies of CARM1 methyltransferase. *EMBO J* **26**, 4402–4412.
10 Cheng Y, Frazier M, Lu F, Cao X & Redinbo MR (2011) Crystal structure of the plant epigenetic protein arginine methyltransferase 10. *J Mol Biol* **414**, 106–122.
11 Antonysamy S, Bonday Z, Campbell RM, Doyle B, Druzina Z, Gheyi T, Han B, Jungheim LN, Qian Y, Rauch C *et al.* (2012) Crystal structure of the human PRMT5:MEP50 complex. *Proc Natl Acad Sci USA* **109**, 17960–17965.
12 Ho MC, Wilczek C, Bonanno JB, Xing L, Seznec J, Matsui T, Carter LG, Onikubo T, Kumar PR, Chan MK *et al.* (2013) Structure of the arginine

methyltransferase PRMT5-MEP50 reveals a mechanism for substrate specificity. *PLoS One* **8**, e57008.
13 Wang C, Zhu Y, Chen J, Li X, Peng J, Zou Y, Zhang Z, Jin H, Yang P, Wu J *et al.* (2014) Crystal structure of arginine methyltransferase 6 from *Trypanosoma brucei*. *PLoS One* **9**, e87267.
14 Wang C, Zhu Y, Caceres Tamar B, Liu L, Peng J, Wang J, Chen J, Chen X, Zhang Z, Zuo X *et al.* (2014) Structural determinants for the strict monomethylation activity by *Trypanosoma brucei* protein arginine methyltransferase 7. *Structure* **22**, 756–768.
15 Hasegawa M, Toma-Fukai S, Kim J-D, Fukamizu A & Shimizu T (2014) Protein arginine methyltransferase 7 has a novel homodimer-like structure formed by tandem repeats. *FEBS Lett* **588**, 1942–1948.
16 Cura V, Troffer-Charlier N, Wurtz JM, Bonnefond L & Cavarelli J (2014) Structural insight into arginine methylation by the mouse protein arginine methyltransferase 7: a zinc finger freezes the mimic of the dimeric state into a single active site. *Acta Crystallogr D Biol Crystallogr* **70**, 2401–2412.
17 Lee WC, Lin WL, Matsui T, Chen ES, Wei TY, Lin WH, Hu H, Zheng YG, Tsai MD & Ho MC (2015) Protein arginine methyltransferase 8: tetrameric structure and protein substrate specificity. *Biochemistry* **54**, 7514–7523.
18 Bonnefond L, Stojko J, Mailliot J, Troffer-Charlier N, Cura V, Wurtz JM, Cianferani S & Cavarelli J (2015) Functional insights from high resolution structures of mouse protein arginine methyltransferase 6. *J Struct Biol* **191**, 175–183.
19 Toma-Fukai S, Kim JD, Park KE, Kuwabara N, Shimizu N, Krayukhina E, Uchiyama S, Fukamizu A & Shimizu T (2016) Novel helical assembly in arginine methyltransferase 8. *J Mol Biol* **428**, 1197–1208.
20 Scott HS, Antonarakis SE, Lalioti MD, Rossier C, Silver PA & Henry MF (1998) Identification and characterization of two putative human arginine methyltransferases (HRMT1L1 and HRMT1L2). *Genomics* **48**, 330–340.
21 Lakowski TM & Frankel A (2009) Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4. *Biochem J* **421**, 253–261.
22 Qi C, Chang J, Zhu Y, Yeldandi AV, Rao SM & Zhu Y-J (2002) Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor alpha. *J Biol Chem* **277**, 28624–28630.
23 Meyer R, Wolf SS & Obendorf M (2007) PRMT2, a member of the protein arginine methyltransferase family, is a coactivator of the androgen receptor. *J Steroid Biochem Mol Biol* **107**, 1–14.
24 Kzhyshkowska J, Kremmer E, Hofmann M, Wolf H & Dobner T (2004) Protein arginine methylation during lytic adenovirus infection. *Biochem J* **383**, 259–265.
25 Yoshimoto T, Boehm M, Olive M, Crook MF, San H, Langenickel T & Nabel EG (2006) The arginine methyltransferase PRMT2 binds RB and regulates E2F function. *Exp Cell Res* **312**, 2040–2053.
26 Blythe SA, Cha SW, Tadjuidje E, Heasman J & Klein PS (2010) beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Dev Cell* **19**, 220–231.
27 Iwasaki H, Kovacic JC, Olive M, Beers JK, Yoshimoto T, Crook MF, Tonelli LH & Nabel EG (2010) Disruption of protein arginine N-methyltransferase 2 regulates leptin signaling and produces leanness *in vivo* through loss of STAT3 methylation. *Circ Res* **107**, 992–1001.
28 Zhong J, Cao RX, Liu JH, Liu YB, Wang J, Liu LP, Chen YJ, Yang J, Zhang QH, Wu Y *et al.* (2014) Nuclear loss of protein arginine N-methyltransferase 2 in breast carcinoma is associated with tumor grade and overexpression of cyclin D1 protein. *Oncogene* **33**, 5546–5558.
29 Oh TG, Bailey P, Dray E, Smith AG, Goode J, Eriksson N, Funder JW, Fuller PJ, Simpson ER, Tilley WD *et al.* (2014) PRMT2 and RORgamma expression are associated with breast cancer survival outcomes. *Mol Endocrinol* **28**, 1166–1185.
30 Zhong J, Cao RX, Hong T, Yang J, Zu XY, Xiao XH, Liu JH & Wen GB (2011) Identification and expression analysis of a novel transcript of the human PRMT2 gene resulted from alternative polyadenylation in breast cancer. *Gene* **487**, 1–9.
31 Zhong J, Cao RX, Zu XY, Hong T, Yang J, Liu L, Xiao XH, Ding WJ, Zhao Q, Liu JH *et al.* (2012) Identification and characterization of novel spliced variants of PRMT2 in breast carcinoma. *FEBS J* **279**, 316–335.
32 Wang YC, Wang JD, Chen CH, Chen YW & Li C (2015) A novel BLAST-based relative distance (BBRD) method can effectively group members of protein arginine methyltransferases and suggest their evolutionary relationship. *Mol Phylogenet Evol* **84**, 101–111.
33 Morales Y, Caceres T, May K & Hevel JM (2016) Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Arch Biochem Biophys* **590**, 138–152.
34 Miranda TB, Miranda M, Frankel A & Clarke S (2004) PRMT7 is a member of the protein arginine methyltransferase family with a distinct substrate specificity. *J Biol Chem* **279**, 22902–22907.
35 Hadjikyriacou A, Yang Y, Espejo A, Bedford MT & Clarke SG (2015) Unique features of human protein arginine methyltransferase 9 (PRMT9) and its substrate RNA splicing factor SF3B2. *J Biol Chem* **290**, 16723–16743.
36 Lin WJ, Gary JD, Yang MC, Clarke S & Herschman HR (1996) The mammalian immediate-early TIS21

protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J Biol Chem* **271**, 15034–15044.
37 Rho J, Choi S, Seong YR, Cho WK, Kim SH & Im DS (2001) Prmt5, which forms distinct homo-oligomers, is a member of the protein-arginine methyltransferase family. *J Biol Chem* **276**, 11393–11401.
38 Lim Y, Kwon YH, Won NH, Min BH, Park IS, Paik WK & Kim S (2005) Multimerization of expressed protein-arginine methyltransferases during the growth and differentiation of rat liver. *Biochim Biophys Acta* **1723**, 240–247.
39 Herrmann F, Pably P, Eckerich C, Bedford MT & Fackelmayer FO (2009) Human protein arginine methyltransferases *in vivo*–distinct properties of eight canonical members of the PRMT family. *J Cell Sci* **122**, 667–677.
40 Feng Y, Xie N, Jin M, Stahley MR, Stivers JT & Zheng YG (2011) A transient kinetic analysis of PRMT1 catalysis. *Biochemistry* **50**, 7033–7044.
41 Krissinel EB, Winn MD, Ballard CC, Ashton AW, Patel P, Potterton EA, McNicholas SJ, Cowtan KD & Emsley P (2004) The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **60**, 2250–2255.
42 Labourier E, Bourbon H-M, Gallouzi I-E, Fostier M, Allemand E & Tazi J (1999) Antagonism between RSF1 and SR proteins for both splice-site recognition *in vitro* and *Drosophila* development. *Genes Dev* **13**, 740–753.
43 Bradley T, Cook ME & Blanchette M (2015) SR proteins control a complex network of RNA-processing events. *RNA* **21**, 75–92.
44 Feng Y, Hadjikyriacou A & Clarke SG (2014) Substrate specificity of human protein arginine methyltransferase 7 (PRMT7): the importance of acidic residues in the double E loop. *J Biol Chem* **289**, 32604–32616.
45 van Haren M, van Ufford LQ, Moret EE & Martin NI (2015) Synthesis and evaluation of protein arginine N-methyltransferase inhibitors designed to simultaneously occupy both substrate binding sites. *Org Biomol Chem* **13**, 549–560.
46 Pesiridis GS, Diamond E & Van Duyne GD (2009) Role of pICLn in methylation of Sm proteins by PRMT5. *J Biol Chem* **284**, 21347–21359.
47 Sayegh J, Webb K, Cheng D, Bedford M & Clarke S (2007) Regulation of protein arginine methyltransferase 8 (PRMT8) activity by its N-terminal domain. *J Biol Chem* **282**, 36444–36453.
48 Yu MC (2011) The role of protein arginine methylation in mRNP dynamics. *Mol Biol Int* **2011**, 163827.
49 Friesen WJ, Paushkin S, Wyce A, Massenet S, Pesiridis GS, Duyne GV, Rappsilber J, Mann M & Dreyfuss G (2001) The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol Cell Biol* **21**, 8289–8300.
50 Cheng D, Cote J, Shaaban S & Bedford MT (2007) The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell* **25**, 71–83.
51 Sinha R, Allemand E, Zhang Z, Karni R, Myers MP & Krainer AR (2010) Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol Cell Biol* **30**, 2762–2774.
52 Bressan GC, Moraes EC, Manfiolli AO, Kuniyoshi TM, Passos DO, Gomes MD & Kobarg J (2009) Arginine methylation analysis of the splicing-associated SR protein SFRS9/SRP30C. *Cell Mol Biol Lett* **14**, 657–669.
53 Abdulrahman W, Uhring M, Kolb-Cheynel I, Garnier JM, Moras D, Rochel N, Busso D & Poterszman A (2009) A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal Biochem* **385**, 383–385.
54 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
55 Battye TGG, Kontogiannis L, Johnson O, Powell HR & Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271–281.
56 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In Macromolecular Crystallography (Carter CW & Sweet RM, eds), pp. 307–326. Academic Press, New York.
57 Long F, Vagin AA, Young P & Murshudov GN (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**, 125–132.
58 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
59 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
60 Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C *et al.* (2011) BUSTER version 2.10.0. Global Phasing Ltd., Cambridge, UK.
61 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
62 Troffer-Charlier N, Cura V, Hassenboehler P, Moras D & Cavarelli J (2007) Expression, purification,

crystallization and preliminary crystallographic study of isolated modules of the mouse coactivator-associated arginine methyltransferase 1. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 330–333.
63 R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
64 Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282.
65 Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**, 495–503.
66 Svergun D, Barberato C & Koch MHJ (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768–773.
67 Franke D & Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342–346.
68 Volkov VV, Lapuk VA, Kayushina RL, Shtykova EV, Varlamova EY, Malfois M & Svergun DI (2003) Low-resolution structure of immunoglobulins IgG1, IgM and rheumatoid factor IgM-RF from solution X-ray scattering data. *J Appl Crystallogr* **36**, 503–508.
69 Kozin MB & Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33–41.
70 Beitz E (2000) TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **16**, 135–139.
71 Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252–W258.
72 Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778–2786.
73 Karplus PA & Diederichs K (2015) Assessing and maximizing data quality in macromolecular crystallography. *Curr Opin Struct Biol* **34**, 60–68.