



Characterisation and Structural Biology of Protein Arginine Methyltransferases

Alexandra Knopp

Thesis submitted for the Degree of Doctor of Philosophy

20th June 2020

Newcastle University

Faculty of Medical Sciences

Translational and Clinical Research Institute

Abstract

Characterisation and Structural Biology of Protein Arginine Methyltransferases

Post-translational and epigenetic modifications of proteins and nucleic acids are known to play major roles in influencing cell fate. Enzymes that catalyse modifications such as phosphorylation, acetylation and methylation have been identified as promising drug targets. Protein methyltransferase 2 (PRMT2) and Coactivator-associated arginine methyltransferase 1 (CARM1) belong to the class of Type 1 PRMTs which catalyse the asymmetric dimethylation of substrate arginine residues. CARM1 has been shown to be overexpressed in different cancer types including breast and prostate cancer. PRMT2 has been identified as a potential target for oncology with reported links to androgen receptor signalling, NF- κ B signalling and induction of apoptosis. However, selective chemical probes that could be used as tools for target validation and which could potentially be a starting point for drug discovery are still missing.

The work presented here aims to identify selective CARM1 and PRMT2 inhibitors that target the cofactor- and substrate-binding sites. Crystal structures of mouse PRMT2 in the apo-state and in complex with Sinefungin are presented. Crystal structures of the catalytic domain of CARM1 in complex with the cofactor S-adenosyl L-homocysteine (SAH) and different small molecule inhibitors were also determined. Surface plasmon resonance was used to characterise inhibitor binding to CARM1 and identify structure-activity relationships. To further map the CARM1 active site, ligand soaks of CARM1 with a library of small fragments called FragLites were performed. These small fragments can more readily find potential binding pockets than larger more drug-like inhibitors. A direct and label-free mass spectrometry-based assay was developed to measure CARM1 activity and its inhibition. Together these findings can be used to further develop inhibitors that target the PRMT family. These inhibitors will be useful tools to investigate the biology of PRMT2 and CARM1 and to understand their biological role in cancer.

Acknowledgements

Throughout this time of conducting my PhD I have met so many people to whom I owe my gratitude. First and foremost, I want to thank my parents Monika and Wolfgang and my fiancé Josh for guiding me through every stage of the PhD journey and supporting me whenever necessary.

A big thank you to my primary supervisor Prof Jane Endicott and Prof Mike Waring for providing feedback and guiding me through the years of my PhD. Equally, I want to thank Arnaud Baslé and Rick Lewis for their continuous, valuable, and encouraging support in the development of my thesis despite various challenges and setbacks. Thanks to my first supervisor Julie Tucker and my colleagues Claire Jennings and Rachel Heap for their guidance, support, and help. Thanks to my sister Steffi for cheering me up and giving me support. Also thank you to Ruth for supporting me throughout the whole journey as a PhD (and before), and for providing help whenever it was needed. Also, special thanks to those amazing people that I am happy to call my friends, Maurizio, Marcus, Andreas, Janina, Zandile, Elena, Hugo for your encouragement, your support, your opinions, and help.

Thanks to Rick Davies and Karolina Peciak for your support during my time at AstraZeneca and to my colleagues Martyna, Max, Mat, Richard, Marco, Lan, Svitlana, and Natalie for supporting me with encouragement, advice, idea generation and feedback. I also want to thank so many people that I cannot all name here individually.

Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Tables	viii
Table of Figures	ix
Abbreviations	xiii
Chapter 1 Introduction	1
1.1 Role of Histone Modification in Epigenetic Regulation.....	1
1.2 The Writers of Methyl-Lysine and Methylarginine Marks	6
1.3 Domain Architecture of PRMTs.....	11
1.4 Structure of PRMTs.....	13
1.5 Active Site and Substrate Specificity of Type 1 PRMTs.....	19
1.6 Catalytic Mechanism of Type 1 Protein Arginine Methyltransferases	21
1.7 Consequences of Arginine Methylation.....	22
1.8 Arginine Methyltransferase Substrate Recognition	23
1.9 Methyl-arginine Binding Proteins- The Readers of Arginine Methylation	25
1.10 Arginine Demethylases, Erasers of Methyl-arginine Marks?	29
1.11 Regulation of PRMT Activity	32
1.11.1 Post-translational Modifications	32
1.11.2 PRMT-Binding Partners	33
1.11.3 Alternative Splicing of PRMTs	34
1.11.4 Modification of Substrate Arginines	35
1.12 Development of PRMT Inhibitors.....	36
1.12.1 Non-selective PRMT Inhibitors	36
1.12.2 PRMT6 and CARM1 Selective Inhibitors.....	37
1.12.3 Inhibitors of PRMT1 and PRMT5.....	40
1.12.4 Selective Inhibitors of PRMT5 and PRMT7	40
1.13 Research Aim	43
Chapter 2 Materials and Methods	45
2.1 Chemicals and Reagents	45
2.2 Expression Constructs	45
2.3 <i>E. coli</i> Strains	46
2.4 Nucleic Acid Techniques.....	47
2.4.1 DNA Concentration Measurement.....	47
2.4.2 Primer Design	47
2.4.3 PCR Amplification of DNA Inserts	49
2.4.4 Preparative Restriction Digest.....	49
2.4.5 InFusion Cloning.....	50

2.4.6	Ligation of PCR Amplicons into Expression Vectors	51
2.4.7	Site-directed Mutagenesis by PCR	51
2.5	Heat-Shock Transformation of Competent <i>E. coli</i> Cells.....	52
2.6	DNA Sequencing	52
2.7	Agarose Gel Electrophoresis.....	52
2.8	Recombinant Protein Expression in <i>E. coli</i> Cells.....	53
2.9	Recombinant Protein Expression in Insect Cells.....	53
2.9.1	Maintenance of Sf9 and Sf21 Cell Lines	54
2.9.2	Transformation and Blue/White Screening of pFastBac1 DNA into DH10EMBacYFP Cells	54
2.9.3	Preparation of X-gal/IPTG Agar Plates for Blue/White Screening	55
2.9.4	Isolation of Bacmid DNA	55
2.9.5	Transfection of Bacmid DNA into Sf9/Sf21 Cells	55
2.10	Small- Scale Expression Test in Sf9 Insect Cells	56
2.11	Small- Scale Expression Test in Sf21 Insect Cells	56
2.12	Small- Scale Expression Test in <i>E. coli</i> Cells	57
2.13	24-well Expression Screen Analysis using Spin Columns	57
2.14	PhyNexus 24-well Expression Screen Purification	58
2.15	Large-Scale Expression in Insect Cells	59
2.16	Protein Purification.....	59
2.16.1	Cell Lysis.....	60
2.16.2	Expression and Purification of His ₆ -RSF1	60
2.16.3	Expression and Purification of His ₁₀ -MBP-CARM1	61
2.16.4	Expression and Purification of GST-MmPRMT2-10.....	61
2.16.5	Expression and Purification of His ₆ -RnPRMT1-1	62
2.16.6	Expression and Purification of HsPRMT2_E223Q.....	63
2.16.7	Expression and Purification of His ₁₀ -MBP-HsPRMT2-6.....	63
2.16.8	Expression and Purification of His ₆ -HsPRMT2-1	64
2.17	Protein Analysis	64
2.17.1	SDS- Polyacrylamide Gel Electrophoresis (PAGE)	64
2.17.2	Western Blotting Analysis	65
2.17.3	Protein Quantification.....	66
2.17.4	Intact Protein Analysis using Mass Spectrometry	66
2.17.5	Limited Proteolysis	66
2.17.6	Thermal Stability Assay using the ThermoFluor-iCycler	67
2.17.7	Analytical SEC using the Ettan System	67
2.18	<i>In vitro</i> Methyltransferase Glo Assay	68
2.18.1	Determination of the Optimal Enzyme Concentration.....	68
2.18.2	Determination of K _m Values for PRMT Substrates.....	70
2.19	MALDI-TOF MS-Based Assay	70
2.20	Surface Plasmon Resonance Assay.....	71

2.20.1	Calculation of the Maximal Response.....	71
2.20.2	Steady-State Affinity Determinations	71
2.20.3	CARM1 Fragment Screening.....	72
2.21	Isothermal Scanning Calorimetry.....	74
2.21.1	CARM1 Preparation.....	74
2.21.2	Analysis of Ligand Binding.....	74
2.21.3	Data Analysis.....	75
2.22	X-Ray Crystallography.....	75
2.22.1	CARM1 Crystallisation.....	75
2.22.2	Crystal Harvest	76
2.22.3	DMSO Tolerance Testing of CARM1 Crystals.....	76
2.22.4	Ligand Soaking into CARM1 Crystals.....	76
2.22.5	Co-crystallization of SNF and CARM1	76
2.22.6	Co-crystallization of Ligands and CARM1	77
2.22.7	Data Collection and Analysis	77
Chapter 3	Recombinant Protein Expression and Characterisation of	
Different Type 1 PRMTs	79
3.1	Introduction.....	79
3.2	Design of Protein Constructs for Recombinant Expression	80
3.3	Small Scale Expression Tests of PRMT1, PRMT2 and CARM1 in <i>E. coli</i> Cells	84
3.3.1	Small-Scale Expression Test in <i>E. coli</i> BL21(DE3)Gold	84
3.3.2	Small Scale Expression Test in <i>E. coli</i> BL21(DE3)pLysS and DH5 α	88
3.3.3	Small Scale Expression Test in <i>E. coli</i> Rosetta(DE3)pLysS and ArcticExpress... ..	89
3.4	Large Scale Expression Tests of Selected Constructs in <i>E. coli</i>	93
3.4.1	Large Scale Purification of His ₆ -RnPRMT1-1	95
3.4.2	Large Scale Purification of GST-RnPRMT1-1	98
3.4.3	Analysing the Oligomeric State of PRMT1 Using Analytical SEC.....	99
3.4.4	Large Scale Purification of GST-HsPRMT2-9.....	100
3.5	Changing the Expression System from <i>E. coli</i> cells to the MultiBac™ System.....	103
3.5.1	Small Scale Expression Tests in Sf9 Cells	104
3.5.2	Small Scale Expression Tests in Sf21 Cells	107
3.5.3	Large Scale Purification of His ₁₀ -MBP-HsCARM1	109
3.5.4	Large Scale Purification of His ₆ -HsPRMT2-1	111
3.5.5	Large Scale Purification of His ₁₀ -MBP-HsPRMT2-6	113
3.6	Limited Proteolysis to optimise HsPRMT2 Construct Design.....	115
3.7	Buffer Optimization for HsPRMT2 and RnPRMT1 using Thermal Melt Assays	116
3.8	Purification of GST-MmPRMT2-10.....	121
3.9	Summary	125
Chapter 4	Enzyme Activity Assay Development and Evaluation.....	127
4.1	The MTase Glo Assay.....	130
4.2	Substrates of CARM1 and PRMT2	131

4.3	RSF1- a New PRMT2 Substrate?	139
4.4	Expression Test of His ₆ -RSF1	140
4.5	PRMT Substrate GST-GAR.....	143
4.6	MALDI-ToF MS-based Methylation Assay.....	145
4.7	Purity Assessment of H3 and H4 Peptide	146
4.8	Production of a Catalytical Inactive HsPRMT2 Mutant.....	153
4.9	Summary	156
Chapter 5	Structural Characterisation of PRMT2	159
5.1	Overall Structure of PRMT2	162
5.1.1	Overview of SH3 Domains	163
5.1.2	The N-terminal SH3 Domain of HsPRMT2	167
5.2	Aims.....	171
5.3	Crystallisation Trials of HsPRMT2 and MmPRMT2.....	172
5.4	Overall Crystal Structure of MmPRMT2 in Complex with Sinefungin	174
5.5	The Sinefungin Binding Site of MmPRMT2	177
5.6	Comparison of the Mouse MmPRMT2 Apo and SNF-bound Structures	182
5.7	Summary	189
Chapter 6	Development of Small-Molecule Inhibitors of CARM1.....	191
6.1	PRMT Inhibitor Design	195
6.2	Targeting the CARM1- Cofactor Binding Site.....	198
6.3	Targeting the CARM1-Substrate Binding Site	201
6.4	Crystallisation Trials of the Catalytic Domain of HsCARM1	202
6.5	SPR Analysis.....	204
6.6	Small-Molecules Targeting the Cofactor Binding Pocket.....	208
6.7	Fragments Targeting the Substrate Binding Pocket.....	213
6.7.1	Stereochemical Study of the Secondary Amine of the Ethylenediamine	229
6.8	Structural Relationship Study of the Phenyl Moiety.....	231
6.8.1	Characterisation of the Pyridine Nitrogen Substituted Series	233
6.8.2	Substitution of the 2-Pyridine or Phenyl Ring.....	238
6.9	Analysis of Bisubstrate Fragment Binding to CARM1	240
6.10	Analysis of FragLite Binding to CARM1	246
6.11	Summary.....	247
Chapter 7	Conclusion and Future Direction	249
References	251
	Appendix A.....	279
	Appendix B.....	280
	Appendix C.	281
	Appendix D.	283
	Appendix E.....	287
	Appendix F.....	288
	Appendix G.	289

Appendix H.....	291
Appendix I.....	292
Appendix J.....	293
Appendix K.....	295
Appendix L.....	296
Appendix M.....	297
Appendix N.....	301
Appendix O.....	302

List of Tables

Table 1-1 Examples of Histone Modifications and their Functions.	5
Table 2-1 Summary of Protein Constructs for Insect Cell Expression.	45
Table 2-2 Summary of Protein Constructs for E. coli Expression.	46
Table 2-3 E. coli Growth Media.	47
Table 2-4 Antibiotics and their used Working Concentration for Bacterial Selection.	47
Table 2-5 Summary of the Designed Primers.	48
Table 2-6 PCR Reaction Cycles for DNA-Insert Amplification.	49
Table 2-7 Preparative Restriction Digest Components.	49
Table 2-8 InFusion Reactions for Purified PCR Fragments.	50
Table 2-9 Primers for Site-Directed Mutagenesis of HsPRMT2.	51
Table 2-10 Reaction Cycles for SDM by PCR.	52
Table 2-11 Buffer Composition for Multiple Parallel Expression Screening.	58
Table 2-12 Buffer Composition for PhyNexus 24-well Purification.	59
Table 2-13 Gel Filtration Standard Components and their Molecular Weight.	68
Table 2-14 Components of the 2× Substrate Reaction Mix.	69
Table 3-1 Summary of PRMT Constructs Tested for Expression in E. coli BL21(DE3)Gold.	85
Table 3-2 Summary of Small-Scale Expression Tests for Rat PRMT1 and Human CARM1 Constructs in the Whole Cell Fraction or the Soluble Fraction.	92
Table 3-3 Summary of Buffers that Increased RnPRMT1 Stability as Determined Using a Thermal Denaturation Assay in Comparison with its Original Buffer.	119
Table 3-4 Summary of Buffers that increased HsPRMT2-1 Stability in Thermal Denaturation Assay.	120
Table 3-5 New Mouse PRMT2 Expression Constructs for Insect Cell Expression.	122
Table 5-1 Growth and Harvest Conditions for MmPRMT2-10 Crystals used for Structure Determination.	173
Table 5-2 Crystallographic Data and Refinement Statistics of MmPRMT2-10 in Complex with Sinefungin.	176
Table 5-3 Crystallographic Data and Refinement Statistic of Apo MmPRMT2-10.	185
Table 6-1 Growth and Harvest Conditions for HsCARM1^{CAT} Crystals used for Structure Determination	203
Table 6-2 Crystallographic Data and Refinement Statistics of Human CARM1 in Complex with Sinefungin.	212
Table 6-3 Literature Fragments Targeting PRMT6 and CARM1.	214
Table 6-4 Newly Designed Alanine- and Ethylenediamine-based Fragments.	225
Table 6-5 Chemical Structure of the Stereoisomers.	229
Table 6-6 Chemical Structure and Affinity Data of the Stereoisomers.	231
Table 6-7 2-Pyridine or Phenyl Ring Substitutions.	238
Table 6-8 Chemical Structure of Designed Bisubstrate Ligands.	244

Table of Figures

Figure 1-1 Histone Structure.	2
Figure 1-2 Post-transcriptional Histone Modifications.	4
Figure 1-3 The Histone Code and its Interpretation.	6
Figure 1-4: Representative Structures of Protein Methyltransferases.	8
Figure 1-5 Phylogenetic Tree of Human PKMTs and PRMTs.	9
Figure 1-6 Arginine Methylation by PRMTs.	10
Figure 1-7 The PRMT Family.	12
Figure 1-8 Overview of PRMT Crystal Structures.	15
Figure 1-9 Structural Overview of Type 1 PRMTs.	16
Figure 1-10 Regulatory Elements of the Rossmann Fold.	18
Figure 1-11 Active Site of the Crystal Structure of CARM1.	20
Figure 1-12 Catalytic Mechanism for Type 1 PRMTs, Exemplified by PRMT1.	22
Figure 1-13 Hydrogen Bond Donor Sites and Electrostatic Surface Potential of Unmethylated and Methylated Arginine Side Chains.	23
Figure 1-14 Arginine Methylation Sites on N-terminal Histone Tails.	24
Figure 1-15 Royal Superfamily Domains- The Reader of Methylation Marks.	26
Figure 1-16 Tudor Domain Structure of SMN and its Recognition of SDMA and ADMA.	28
Figure 1-17 Structure and Function of PADs.	30
Figure 1-18 Demethylation Reaction by JMJD6.	31
Figure 1-19 Overview of PRMT1 and PRMT7 Isoforms.	35
Figure 1-20 Early non-selective direct and indirect PRMT Inhibitors.	37
Figure 1-21 Early Inhibitors of CARM1 and PRMT6.	38
Figure 1-22 Selective CARM1 Inhibitors.	39
Figure 1-23 Bisubstrate Inhibitors of PRMT1 and PRMT5.	40
Figure 1-24 PRMT Type 1 and Type 2 Inhibitors in Clinical Trials.	41
Figure 1-25 Selective PRMT Inhibitors that target the SAM-binding Site or allosteric Sites.	42
Figure 2-1 Ligand Immobilisation via Amine Coupling.	73
Figure 3-1 Schematic Illustration of Rat PRMT1 and Human CARM1 and Expression Construct Selection.	80
Figure 3-2 Expression Constructs Designed for Human PRMT2.	81
Figure 3-3 Designed Expression Constructs for Mouse PRMT2.	82
Figure 3-4 Secondary Structure Prediction for Human PRMT2 with the PSIPRED Server.	82
Figure 3-5 In silico Prediction of Disordered Regions of Human PRMT2 using RONN.	83
Figure 3-6 Small Scale Expression Test of Different GST-tagged Constructs in E. coli BL21(DE3)Gold.	87
Figure 3-7 Small Scale Expression Test of different GST-tagged Constructs in E. coli BL21(DE3)Gold.	88
Figure 3-8 SDS-PAGE Analysis of Small Scale Expression Test of different GST-tagged Constructs in E. coli DH5α and BL21(DE3)pLysS.	89

Figure 3-9 SDS-PAGE Analysis of Small-Scale Expression Test of different GST-tagged Constructs.....	90
Figure 3-10 SDS-PAGE Analysis of Large-Scale Expression Test of different GST- and His-tagged Constructs in different E. coli strains.....	94
Figure 3-11 Western Blot with anti-GST-Antibody to Detect PRMTs.....	95
Figure 3-12 RnPRMT1-1 Purification.....	96
Figure 3-13 Mass Determination of Purified RnPRMT1-1 Sample by Q-TOF LC/MS.....	98
Figure 3-14 SDS-PAGE Analysis of the GST-RnPRMT1-1 Purification.....	99
Figure 3-15 Oligomerisation State Analysis of RnPRMT1-1 upon Addition of Cofactor.....	100
Figure 3-16 SDS-PAGE Analysis of GST-HsPRMT2-9 Purification.....	102
Figure 3-17 Principle of the Bac-to-Bac Baculovirus Expression System.....	104
Figure 3-18 Small-Scale Expression Test Analysis in Sf9 Cells.....	105
Figure 3-19 Large Scale Expression Test in Sf9 Cells.....	106
Figure 3-20 SDS-PAGE Analysis of Protein Expression Tests in Sf21 Cells.....	108
Figure 3-21 Small Scale Expression Test of Large Scale MBP-CARM1-1.....	109
Figure 3-22 Example Purification of His₁₀-MBP-tagged Human CARM1 from Sf21 Cells.....	110
Figure 3-23 Example Purification of His₆-tagged Human HsPRMT2-1 from Sf9 Cells.....	112
Figure 3-24 Example Purification of His₁₀-MBP-tagged HsPRMT2-6 from Sf9 Cells.....	114
Figure 3-25 Limited Proteolysis and Resulting Proteolytic Fragments of HsPRMT2-1.....	115
Figure 3-26 Principle of Thermal Denaturation Assay.....	117
Figure 3-27 Exemplary Thermal Denaturation Assay of RnPRMT1-1.....	118
Figure 3-28 Exemplary Thermal Denaturation Assay of HsPRMT2-1.....	120
Figure 3-29 MmPRMT2 Protein Sequence Alignment.....	121
Figure 3-30 SDS-PAGE Analysis of GST-MmPRMT2-10 Expressed in Sf9 cells and Affinity Purified.....	123
Figure 3-31 SEC and SDS-PAGE Analysis of MmPRMT2-10.....	124
Figure 3-32 IEX Chromatography of MmPRMT2-10.....	125
Figure 4-1 Overview of PRMT Activity Assays.....	128
Figure 4-2 Principle of the Bioluminescence-based MTase-Glo Assay.....	131
Figure 4-3 PRMT Arginine Methylation Sites on Histone H3 and H4.....	133
Figure 4-4: Measurement of PRMT Activity using the MTase Glo Assay.....	134
Figure 4-5 Determining the K_m value of H4 for PRMT1.....	135
Figure 4-6 Histone H4 Titration for PRMT2.....	136
Figure 4-7 PRMT2 Auto-Methylation.....	138
Figure 4-8 S. frugiperda RSF1.....	140
Figure 4-9 SDS-PAGE Analysis of His₆-RSF1 Purification.....	141
Figure 4-10 Purification of RSF1 by SEC.....	142
Figure 4-11 RSF1 Activity Assessment using the MTase Glo Assay.....	143
Figure 4-12 GST-GAR Expression Test in BL21(DE3)pLysS.....	144
Figure 4-13 Principle of the Direct Label-free PRMT MALDI-ToF Mass Spectrometry Assay.....	145
Figure 4-14 Purity Analysis of H3 (22-45) Peptide.....	147
Figure 4-15 Detection of H3(22-45) Peptide by Mass Spectrometry.....	148

Figure 4-16 MALDI-TOF MS Spectra of CARM1 Activity Assay with H3(22-45).	149
Figure 4-17 Effect of Sodium Chloride on CARM1 Activity.	150
Figure 4-18 Purity Analysis of H3(22-45) Peptide.	151
Figure 4-19 Effects of Sodium Chloride on H3 Peptide Ionisation.	152
Figure 4-20 Measurement of CARM1 Activity using the MTase Glo Assay.	153
Figure 4-21 SDS-PAGE Analysis of the Purification of HsPRMT2_E223Q.	155
Figure 4-22 Analysis of HsPRMT2_E223Q Sample by Mass Spectrometry.	156
Figure 5-1 Overview of PRMT2 Isoforms.	160
Figure 5-2 Structure-based Sequence Alignment of the Catalytic Core of Different PRMTs.	163
Figure 5-3 Structural Architecture of SH3 Domains.	164
Figure 5-4 SH3 Domain Binding to a Class I or Class II Peptide.	165
Figure 5-5 NMR Structure of HsPRMT2 SH3 Domain.	168
Figure 5-6 Structural Analysis of the HsPRMT2 SH3 Domain and its Binding Surface.	170
Figure 5-7 HsPRMT2 and MmPRMT2 Crystallisation Trials. Protein used for Crystallisation Experiments.	172
Figure 5-8. MmPRMT2-10 Crystallisation Trials with/without Sinefungin.	173
Figure 5-9 Structural Overview of MmPRMT2 in Complex with Sinefungin.	175
Figure 5-10 Chemical Structures of the PRMT cofactor SAM and the fungal-derived pan-PRMT Inhibitor Sinefungin.	177
Figure 5-11 Overview of the SNF-binding Domain of MmPRMT2.	178
Figure 5-12 The Sinefungin Binding Pocket of Type 1 PRMTs.	179
Figure 5-13 Structural Comparison of MmPRMT2 in Complex with SNF and SAH.	180
Figure 5-14 Cofactor and Substrate Binding Pockets of Different PRMTs.	181
Figure 5-15 Structural Overview of Apo-MmPRMT2.	182
Figure 5-16 Structural Comparison of Apo-MmPRMT2 and MmPRMT2 in Complex with SNF.	183
Figure 5-17 Observed Electron Density at the N-terminus of MmPRMT2.	184
Figure 5-18 Surface Representation of the PRMT2 Structures.	186
Figure 5-19 Structural Dynamics of the α-helix in PRMTs.	187
Figure 5-20 Structural Dynamics of the α-helix of CARM1 and PRMT2.	188
Figure 6-1 Unique Features of the Catalytical Domain of CARM1.	192
Figure 6-2 N-terminal PH-Domain of CARM1.	194
Figure 6-3 Mechanism of Action of PRMT Inhibitors.	196
Figure 6-4 Targeting CARM1 Binding Pockets.	197
Figure 6-5 Structural Diversity of the Cofactor Binding Pocket.	199
Figure 6-6 Structural Diversity of the Substrate-Binding Pocket.	201
Figure 6-7 HsCARM1-1 Crystallisation Trials. Protein used for Crystallisation Experiments.	202
Figure 6-8 HsCARM1^{CAT} Crystallisation Trials with and without SAH.	203
Figure 6-9 Principle of Surface Plasmon Resonance Analysis.	205
Figure 6-10 SPR Sensorgram.	206
Figure 6-11 HsCARM1^{CAT} Amine-Coupling to CM5 Sensor Chip.	207
Figure 6-12 Fragments Designed to Target the Cofactor Binding Pocket.	209
Figure 6-13 SPR Analysis of NCL-00024109 and SNF.	210

Figure 6-14 Characterisation of the CARM1-Sinefungin Complex.	211
Figure 6-15 PRMT6 and CARM1 Inhibitors Used for Original Fragmentation Study.	213
Figure 6-16 Overview of Fragment Design.	215
Figure 6-17 Characterisation of NCL-00024058 Binding to CARM1.	217
Figure 6-18 Characterisation of the Binding of NCL-00025070 to CARM1.	218
Figure 6-19 Comparison of the NCL-00025069 Binding Site of CARM1 and PRMT6.	219
Figure 6-20 PRMT6 and CARM1 Superposition with NCL-00025069 and CMPD-1.	220
Figure 6-21 Binding Analysis of NCL-00024056 to CARM1.	222
Figure 6-22 ITC Analysis of NCL-00024056 Binding to HsCARM1^{CAT}.	223
Figure 6-23 SPR and ITC Analysis of NCL-00024057 Binding to CARM1.	224
Figure 6-24 SPR Analysis of NCL-00024860 Binding to CARM1.	225
Figure 6-25 Analysis of Fragment Binding to CARM1 by SPR.	226
Figure 6-26 Characterisation of the Binding of Ethylenediamine-Based Fragments to CARM1 by X-Ray Crystallography.	227
Figure 6-27 Characterisation of the Alanine-Based Ligand NCL-00024055.	228
Figure 6-28 Comparison of the Binding of NCL-00024053 and NCL-00024858 to CARM1.	230
Figure 6-29 Characterisation of the Binding NCL-00024852 and NCL-00024854 to CARM1.	232
Figure 6-30 Structures of the Designed Fragments.	233
Figure 6-31 SPR Binding Analysis of Pyridine Nitrogen Substituted Series.	234
Figure 6-32 Characterisation of the Binding of NCL-00024856 and NCL-00024855 to CARM1.	235
Figure 6-33 Characterisation of NCL-00024861 and NCL-00024862 Binding to CARM1.	237
Figure 6-34 Structures of CARM1 in Complex with NCL-00025544 and NCL-00025547 Determined by Co-Crystallisation.	239
Figure 6-35 Approaches Used for Fragment Optimisation.	241
Figure 6-36 Location of the CARM1-SAH and NCL-00024062 Binding Pockets.	241
Figure 6-37 Bisubstrate PRMT Inhibitors.	242
Figure 6-38 Newly Identified Bisubstrate Inhibitors of CARM1.	243
Figure 6-39 SPR Analysis of Bisubstrate Ligand Binding to CARM1.	245
Figure 6-40 ITC Analysis of NCL-0002471 and NCL-00025094 and CARM1.	246

Abbreviations

°C	Degree Celsius
ADMA	Asymmetrical Di-Methylated Arginines
AIM	Auto Induction Medium
AR	Androgen Receptor
ATP	Adenosine triphosphate
BAF155	BRG1-associated Factor 155
BSA	Bovine Serum Albumin
CAF1	CCR4-Associated Factor 1
cAMP	Cyclic Adenosine Monophosphate
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CARM1	Coactivator Associated Arginine Methyltransferase 1
CHCA	Cyano-4-Hydroxycinnamic Acid
C-terminal	Carboxy-terminal
COMT	Catechol-O-Methyltransferase
CV	Column Volume
Da	Dalton
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNTM	DNA Methyltransferase
dNTP	Deoxyribonucleotide Triphosphate
DOT1L	DOT1 Like Histone H3K79 Methyltransferase
<i>Dr</i>	<i>Danio rerio</i> (Zebrafish)
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylene Diamine Tetraacetic Acid
EG	Ethylene Glycol
EHV1	<i>Drosophila</i> -enabled/vasodilator-stimulated Phosphoprotein Homology 1

ER α	Estrogen Receptor α
FBXO	F-box only
G/ xg	g-force
GAR	Glycine/Arginine-Rich
GST	Glutathione S-transferase
h	Hour
H	Enthalpy
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HnRNP	Heterogeneous Nuclear Ribonucleoprotein
<i>Hs</i>	<i>Homo sapiens</i> (Human)
Hsp	Heat shock protein
IC50	Half Maximal Inhibitory Concentration
IMAC	Immobilized-metal-ion Affinity Chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRF4	Interferon Regulatory Factor 4
ITC	Isothermal Titration Calorimetry
JMJD	Jumonji C Domain-Containing
K	Kelvin
K _d	Equilibrium Dissociation Constant
KDM	Histone Lysine Demethylase
K _m	Michaelis Constant
LB	Lysogeny Broth
Log P	Partition Coefficient octanol:water
MBP	Maltose Binding Protein
MMA	Mono-Methylated Arginines
MEP50	Methylosome Protein 50
MES	2-(N-Morpholino) Ethane Sulfonic Acid
METTL23	Methyltransferase Like 23
<i>Mm</i>	<i>Mus musculus</i> (Mouse)
MMA	Mono Methylated Arginine
MAPK	Mitogen-Activated Protein Kinase

MTase	Methyltransferase
NDUFAF7	NADH:ubiquinone Oxidoreductase Complex Assembly factor 7
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic acid
N-terminal	Amino-terminal
NUMAC	Nucleosomal Methylation Activator Complex
GlcNAc	N-acetylglucosamine
PAD	Protein Arginine Deiminases
PAGE	Polyacrylamide Gel Electrophoresis
PABP1	Poly(A)-binding Protein 1
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PGMs	Proline, Glycine, Methionine, and Arginine-rich Sites
PHD	Plant Homeodomain
PH	Pleckstrin Homology
PKA	Protein Kinase A
PMT	Protein Methyltransferase
PKMTs	Protein Lysine Methyltransferases
PPII	Polyproline Type II
PPAR γ	Peroxisome Proliferator-Activated Receptor γ
PR	Proline-Rich
PRMTs	Protein Arginine Methyltransferases
PSA	Polar Surface Area
R	Response
RAR α	Retinoic Acid Receptor α
<i>Rn</i>	<i>Rattus norvegicus</i> (Rat)
RNA	Ribonucleic Acid
RNMT	RNA Methyltransferase
rpm	Revolutions per Minute

RT	Room Temperature
S	Entropy
SDMA	Symmetrical Di-Methylated Arginines
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SCF	SKP1-Cullin1-F-box Protein
SEC	Size Exclusion Chromatography
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
<i>Sf</i>	<i>Spodoptera frugiperda</i>
SH3	Src Homology 3
SDS	Sodium Dodecyl Sulfate
SMN	Survival Motor Neuron
S _N 2	Bimolecular Nucleophilic Substitution
SNF	Sinefungin
SPF30	Splicing Factor 30
SKP2	S-Phase Kinase-associated Protein 2
SPR	Surface Plasmon Resonance
<i>Tb</i>	<i>Trypanosoma brucei</i>
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TDRD	Tudor Domain-containing Proteins
TFA	Trifluoroacetic Acid
TIM	Triosephosphate Isomerase
T _m	Melting Temperature
Tris	Tris(hydroxymethyl)aminomethane
TPR	Tetratricopeptide Repeat
U2AF65	U2 Small Nuclear Ribonucleoprotein Auxiliary Factor 65-kD Subunit
UV	Ultraviolet
V	Volume
W	Weight
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside

YFP

Yellow Fluorescent Protein

ZnF

Zinc Finger

Amino Acid Name	3 Letter Code	1 Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Hydroxyproline	Hyp	O
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Pyroglutamic	Glp	U
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter 1 Introduction

1.1 Role of Histone Modification in Epigenetic Regulation

The field of epigenetics comprises all molecular mechanisms which cause an increase or decrease of gene transcription without changing the genetic information. Today the term epigenetics is defined as “the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence” (Morris, 2001). Research has shown that epigenetic regulation of gene expression is very important for cell differentiation and that epigenetic dysregulation is often the cause of human diseases and has been observed in most human cancer types, reviewed by Pfister and Ashworth (2017). Different molecular mechanisms control the epigenetic regulation and include DNA-methylation which occurs at cytosine residues and leads to gene silencing and post-translational modifications (PTMs) of histones, which can be either gene activating or repressing depending on the modified amino acid residue (Mohammad *et al.*, 2019).

Histones are small basic structural proteins of chromosomes around which the DNA is wrapped. Four major histone proteins exist named H2A, H2B, H3 and H4. The DNA is organised in regular, repeating units which are called nucleosomes, the core consists of an octameric histone complex. It consists of 2 molecules of each histone H2A, H2B, H3 and H4, which are positively charged, and around which 145-147 negatively charged DNA nucleotides are wound (Luger *et al.*, 1997, Lawrence *et al.*, 2016). The crystal structure of the histone core and DNA is shown in Figure 1-1.

The linker histone H1 is located outside of the histone core and together with the linker DNA, they form a higher order complex named the nucleosome, as illustrated in Figure 1-1 B (Allan *et al.*, 1980, Zhou *et al.*, 2019). The linker DNA has varying length of 7-101 base pairs and, together with the nucleosome core particle, the DNA-protein complex is named chromatin (van Holde and Zlatanova, 1995). Supercoil DNA is a form of DNA in which the DNA strand is either over- or under-wound compared to the relaxed state. In eukaryotes, each nucleosome constrains one under-wound DNA supercoil, whereas the linker DNA between the nucleosome core particles is unconstrained (Corless and Gilbert, 2016). The nucleosome arrays, the well-known beads-on-a-string structure (Thoma *et al.*, 1979), can condense into higher-order fibres and

large-scale chromatin structures via inter-nucleosomal, intra-fibre, and inter-fibre interactions (Horn and Peterson, 2002) influencing gene sequence accessibility and thus gene regulation (Corless and Gilbert, 2016).

A



B

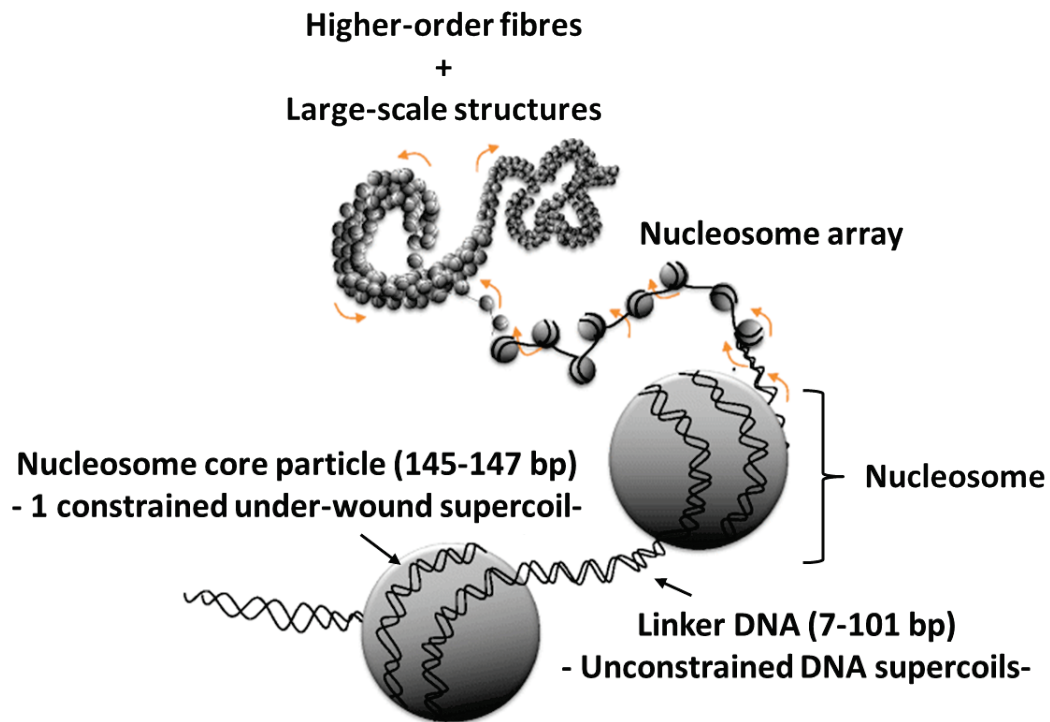


Figure 1-1 Histone Structure.

A) Crystal structure of the histone core and DNA (PDB: 1KX5 (histone octamer); PDB: 1ZBB (DNA)). Histone H3 and H4 are shown in purple and green, histone H2A and H2B in yellow and red, respectively. **B)** Chromatin Structure. The negatively charged chromosomal DNA is wrapped around positively charged histone proteins, together they form nucleosomes. Each consists of ~145 DNA base pairs wound around an octameric histone core, with a size of 10 nm. Together with the linker histone H1 they form 30 nm chromatin fibres, that fold into higher-order structures. DNA supercoiling can influence chromatin structure and genome regulation. Adapted, with permission, from Zhou et al. (2019) and Corless and Gilbert (2016).

Moreover, it was suggested that DNA supercoiling can influence gene regulation through chromatin re-arrangements and shaping across long distances as reviewed by Björkegren and Baranello (2018).

Most of the nucleosomal DNA is bound to structurally conserved histone domains named the histone fold. However, flexible N-terminal histone tails flank both ends of the histone fold, as shown for H3 in Figure 1-1 (Zhou *et al.*, 2019). These tails not only interact with the nucleosomal DNA but also with linker DNA and adjacent nucleosomes and are very important for the structural stability of chromatin (Iwasaki *et al.*, 2013). Moreover, nucleosomes are not static structures but highly dynamic and are influenced by many factors including not only post-translational modifications (PTMs), but also histone variants and interactions with chromatin-binding proteins. Chromatin can exist in different conformational states. For clarity, a condensed state (heterochromatin) in which the DNA is tightly packed and a relaxed state (euchromatin) is differentiated, even if more than 15 different chromatin states do exist (Ernst *et al.*, 2011).

The extreme tight packaging of the DNA in the heterochromatin state inhibits the interaction of the DNA with proteins, such as transcription factors, resulting in a reduced gene transcription rate compared to the euchromatin form (Luger *et al.*, 1997). Moreover, gene transcription can be influenced by direct covalent modifications of DNA molecules, post-translational modifications of histones, or interaction with other non-histone proteins. They can influence gene transcription by controlling the transition process between the chromatin states, influence chromatin compaction and accessibility (Bannister and Kouzarides, 2011). Different types of post-translational modifications (PTMs) exist, which mainly occur on the N-terminal tails of the core histones which are rich in lysine (Lys) and arginine (Arg) residues (Lawrence *et al.*, 2016). An overview of different histone tail modifications is shown in Figure 1-2.

However, PTMs not only occur on the histone tails but also on the globular domain (Tropberger and Schneider, 2013). A variety of histone modifications exist such as phosphorylation, acetylation, lysine and arginine methylation, ubiquitination, sumoylation, and deimination (Kouzarides, 2007, Kebede *et al.*, 2015). An example for PTMs in core histone regions is the methylation of Arg42 of histone H3 by coactivator-associated arginine methyltransferase 1 (CARM1) and protein arginine methyltransferase (PRMT) 6, which activates transcription most likely by loosening histone-DNA interactions (Casadio *et al.*, 2013).

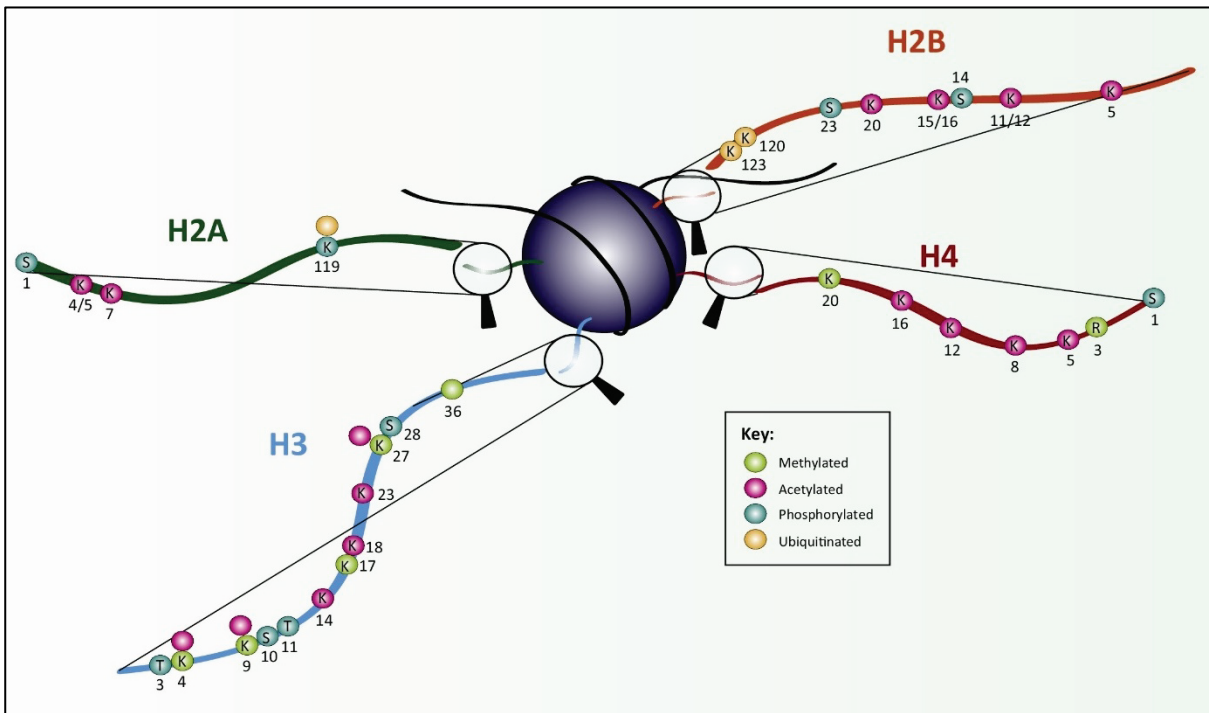


Figure 1-2 Post-transcriptional Histone Modifications.

Histone modifications mainly occur on the N-terminal tails, which are reaching out of the histone core and are accessible. The amino acid name and position of the most common PTMs on each tail are labelled and the type of modification is shown (green= methylation, pink= acetylation, blue= phosphorylation, yellow= ubiquitination). Figure reproduced, with permission, from Lawrence *et al.* (2016).

An overview of all currently known histone modifications and their function is shown in Table 1.1 and was recently reviewed by Zhao and Shilatifard (2019). PTMs can influence gene transcription by alternating the histone overall charge or inter-nucleosomal interactions but also by recruiting specific binding proteins for example Chromo- or Plant Homeodomain (PHD) containing proteins (Zhao and Shilatifard, 2019).

Histone modifications can cause changes in the chromatin structure and thus can regulate gene expression. PTMs can cause transcription activation but also decrease transcription. In most cases, they affect chromatin structure and transcription indirectly by recruitment of other proteins such as transcription factors or chromatin remodelling proteins (Clements *et al.*, 2003, Wysocka *et al.*, 2006). However, some direct cases are known (Lawrence *et al.*, 2016). For example, acetylation of Lys16 of the N-terminal tail of histone H4 activates transcription by reducing chromatin condensation (Shogren-Knaak *et al.*, 2006).

Table 1-1 Examples of Histone Modifications and their Functions.

Table adapted, with permission from Kouzarides (2007).

Histone PTMs	Modified Residues	Regulated Functions
Lysine Acetylation	Lys-ac	Transcription, Repair, Replication, Condensation
Lysine Methylation	Lys-me1, Lys-me2, Lys-me3	Transcription, Repair
Arginine Methylation	Arg-me1, Arg-me2a, Arg-me2s	Transcription
Phosphorylation	Ser-ph, Thr-ph	Transcription, Repair, Condensation
Ubiquitylation	Lys-ub	Transcription, Repair
Neddylation	Lys-N8	
Sumoylation	Lys-su	Transcription
ADP Ribosylation	Glu-ar	Transcription
Deimination	Arg > Cit	Transcription
Proline Isomerization	Pro-cis > Pro-trans	Transcription
Lysine Crotonylation	Lys-Cr	Transcription
Butyrylation	Lys-Bu	Transcription
Propionylation	Lys-Pr	Transcription
Tyrosine Hydroxylation	Tyr-OH	
Biotinylation		-
O-GlcNAc	Ser/Thr-OGA	Transcription, Repair, Replication
N-formylation	Lys-fo	Repair

The enzymes that perform these histone modifications are of great interest and many of them have been identified and characterised. The enzymes that add these so-called histone “marks” are described as “writers” (Figure 1-3). They can be grouped according to the chemical modification they catalyse and their target substrate. Histone-modifying enzymes include acetyltransferases, deacetylases, methyltransferases, demethylases, kinases, ubiquitin ligases, and proline isomerases (Bannister and Kouzarides, 2011).

Enzymes that remove the histone marks are named “erasers” and proteins that recognize the specific chemical groups on the histone tails and interact with them are called “readers” (Tarakhovsky, 2010). For example, chromodomains interact with methylated lysine residues of histone H3 and H4 tails (Blus *et al.*, 2011, Hard *et al.*, 2018). The correct interactions between those enzymes are important to ensure precise gene regulation.

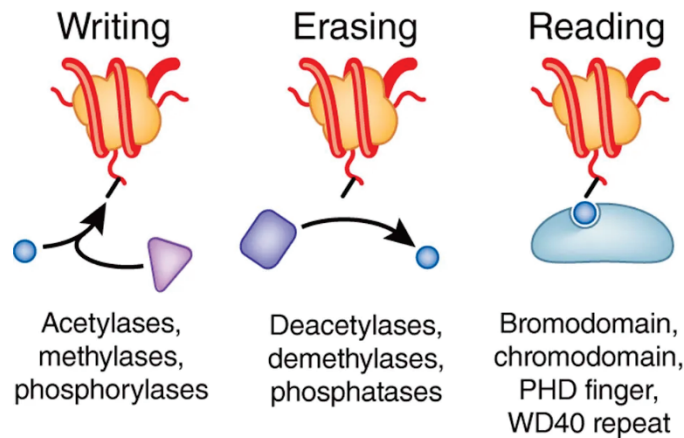


Figure 1-3 The Histone Code and its Interpretation.

More than 100 enzymes are able to add (writer), remove (eraser) or interact (reader) with the chemical modifications on the tail of histone proteins. Figure reproduced, with permission from Tarakhovsky (2010).

Histone methylation can take place on the basic amino acid residues lysine, arginine, and histidine. Protein methyltransferases (PMTs) are the writers of this mark and can methylate lysine or arginine residues of histones but also non-histone proteins (Greer and Shi, 2012). This thesis will focus on arginine methyltransferases which will be further explored in the next chapters.

1.2 The Writers of Methyl-Lysine and Methylarginine Marks

Protein methyltransferases (PMTs) are enzymes that catalyse the transfer of a methyl group from the cofactor S-adenosyl-methionine (SAM) to histones and other proteins and can be divided into two families according to the residue they methylate (Di Lorenzo and Bedford, 2011). Protein lysine methyltransferases (PKMTs) catalyse the methylation of lysine residues, whereas protein arginine methyl transferases (PRMTs) methylate arginine residues (Copeland *et al.*, 2009).

More than 50 human PKMTs exist that can mono-, di-, and/or tri-methylate lysine residues (Martin and Zhang, 2005). Based on their catalytic domain sequence, PKMTs can be divided into two groups, disruptor of telomeric silencing-1 (Dot1) domain-containing proteins and Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain-containing proteins. The latter contain a 130 amino acid long catalytic SET-domain defined by specific amino acid motifs and were first identified in different *Drosophila* proteins namely Su(var.) 3-9 (suppressor of position-effect variegation 3-9),

En(zeste) (enhancer of the eye colour mutant zeste), and Trithorax (homeotic gene regulator) (Dillon *et al.*, 2005, Herz *et al.*, 2013). SET-domain containing PKMTs also contain an I-SET (immunoglobulin-SET) and post-SET domain, which are involved in formation of the substrate binding and in some cases also cofactor binding pocket (Schapira, 2011). As an example, the crystal structure of the SET-domain containing PKMT SETD7 (PDB: 1MT6116) is shown in Figure 1.4 A. Structural features include the I-SET domain (yellow), the SET domain (green) that adopts a pseudoknot structure, and the post-SET domain (cyan).

SET-domain containing PKMTs can methylate lysine residues of histones but also non-histone proteins (Herz *et al.*, 2013) and their SET domains adopt a unique β -fold which is not found in other SAM-dependent methyltransferases (Schubert *et al.*, 2003). Their active site is highly conserved across the family and consists of a knot substructure containing two conserved motifs which is surrounded by small β -sheets and is located next to the cofactor binding pocket (Jacobs *et al.*, 2002). The SET domain containing proteins can be further divided into 6 groups according to sequence similarities: the suppressor of variegation 3–9 (SUV39) family; the SET1 or MLL family; the SET2 family; the SET and MYND (named after myeloid translocation protein 8, Nery, and DEAF-1) domain-containing (SMYD) family; the enhancer of zeste (EZ) family; and the SUV4–20 family (Dillon *et al.*, 2005). Other SET domain containing proteins such as SETD7 and SETD8 have not been assigned to any group yet.

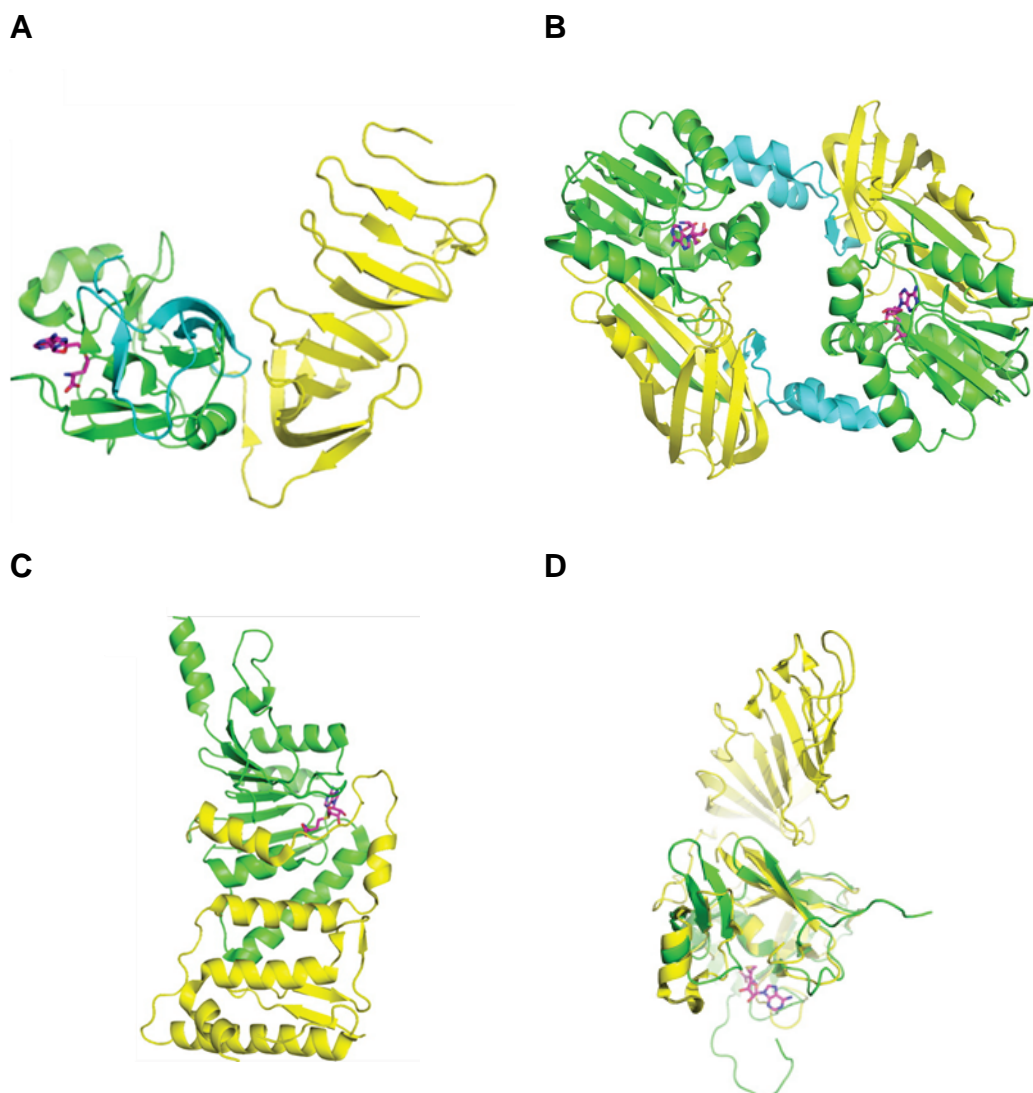


Figure 1-4: Representative Structures of Protein Methyltransferases.

The cofactor SAM/SAH is shown as magenta sticks. **A)** PKMT SETD7 bound to SAH (PDB: 1MT6116). Structural features including the I-SET domain (yellow), SET domain (green), and post-SET domain (cyan) are shown. **B)** PRMT3 in complex with SAH (PDB: 1F3L85). The methyltransferase domain (green), β -barrel domain (yellow), and dimerization domain (cyan) are shown. **C)** DOT1L (PDB: 3QOW5). Structural features include the N-terminal domain (yellow) and the C-terminal α/β domain that binds SAM (green). **D)** PRDMs are structurally related to SET domains. PRMD2 (green, PDB: 2QPW4) is overlaid onto SETD7 (yellow, PDB: 1MT6116). Figure reproduced, with permission from Boriack-Sjodin and Swinger (2016).

The phylogenetic tree of human PKMTs (Figure 1-5 A) also incorporates 16 members of the PRDM (PRDI-BF1 and RIZ homology domain containing) family, that all contain a unique N-terminal domain, known as PRDI-BF1-RIZ1 homologous (PR) domain. The PR domain was firstly discovered and named after the two proteins, PRDI-BF1 (positive regulatory domain I-binding factor 1) and RIZ1 (retinoblastoma protein-interacting zinc finger gene 1) (Buyse *et al.*, 1995). The PR domain is closely

related to the SET domain, but they miss the catalytically important NHXC motif which is highly conserved in PKMTs (NHxC). However, some family members are able to methylate lysine residues (Eram *et al.*, 2014). A structural overlay of PRMD2 (green, PDB: 2QPW4) and SETD7 (yellow, PDB: 1MT6116) is shown in Figure 1-4 D.

In humans, only one Dot1-like protein is known, named HsDOT1L. Its crystal structure is shown in Figure 1.4 C. It exists as a monomer and does not contain a SET domain but a C-terminal α/β domain that binds the cofactor SAM, shown in green (Min *et al.*, 2003). Sequence analysis revealed that it possesses some SAM binding motifs that are similar to those of the PRMTs (Dlagic, 2001), which is why it is included in the phylogenetic tree of PRMTs (Figure 1-5 B). However, DOT1L does not methylate arginine residues but lysine residues on nucleosomal histones and is thus a lysine methyltransferase (Scheer *et al.*, 2019).

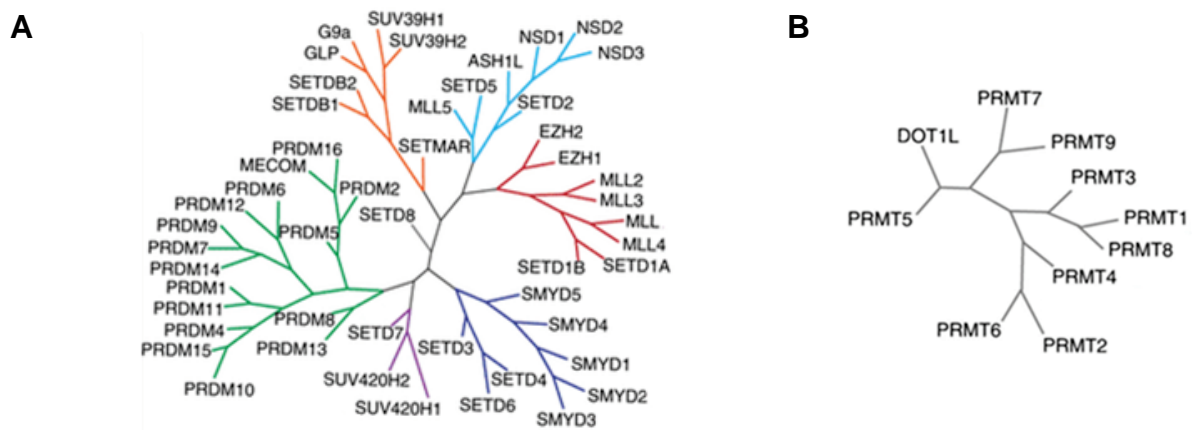


Figure 1-5 Phylogenetic Tree of Human PKMTs and PRMTs.

Phylogenetic Tree of **A)** PR and SET-domain containing PKMTs and **B)** β -barrel domain containing PRMTs and DOT1L proteins. Figure adapted, with permission, from Scheer *et al.* (2019).

PRMTs catalyse the transfer of the methyl group from the cofactor SAM to the guanidinium group of protein arginine side chains generating the by-product S-adenosyl-L-homocysteine (SAH) (Bedford and Clarke, 2009). In humans, nine classic PRMT family members are distinguished which can be divided into three groups according to their methylation position (Figure 1-6). The methylation reaction catalysed by the PRMTs results in either mono-methylated arginine (MMA) residues, asymmetric di-methylated (ADMA) or symmetric di-methylated (SDMA) arginine residues (Bedford, 2007).

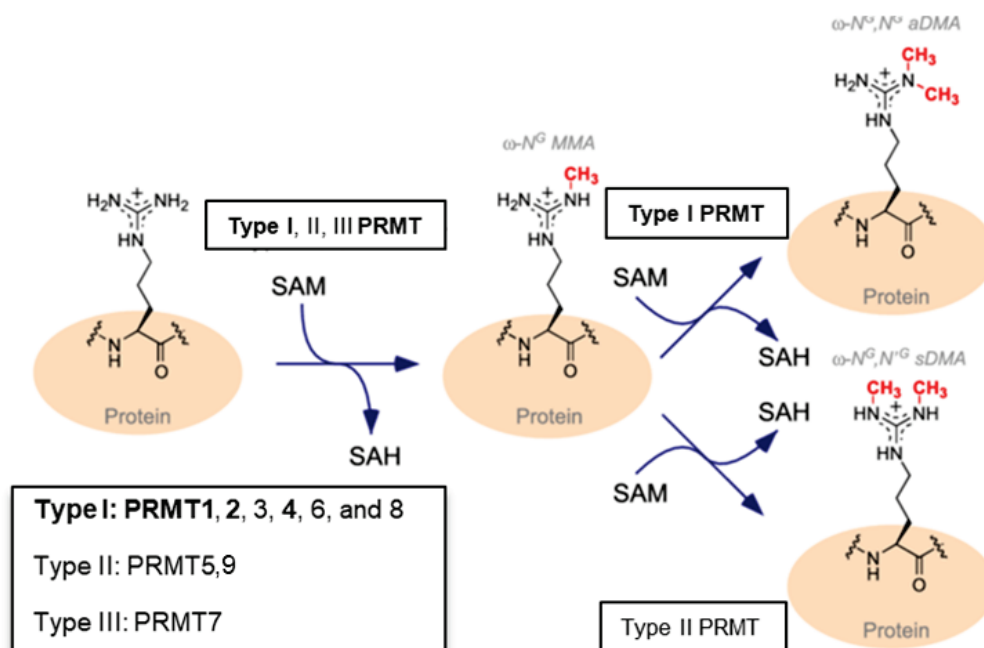


Figure 1-6 Arginine Methylation by PRMTs.

PRMTs catalyse the methylation of arginines and can be divided into three groups. PRMT Type 1, 2, and 3 can mono-methylate (ω -MMA) the terminal guanidinium nitrogen atom of arginine residues. Type 1 PRMTs, which include PRMT2, PRMT1 and PRMT4/CARM1, can in a second reaction asymmetrically di-methylate arginines (ω -ADMA). Type 2 PRMTs are capable of symmetric di-methylation (ω -SDMA) of arginines. Figure adapted, with permission from Stopa et al. (2015).

In human cells, PRMT1 (UniProtKB: Q99873) is the predominant PRMT family member that is found throughout the cell and conducts over 85 % of arginine methylation (Bedford, 2007). Type 1 PRMTs are capable of ω -NG-mono- (Figure 1-6, ω -MMA) and ω -NG,NG-asymmetrical dimethylation (Figure 1-6, ω -ADMA) of the guanidinium group, whereas Type 2 PRMTs produce mono- and ω -NG,N'G-symmetric di-methylation (Figure 1-6, ω -SDMA). In contrast, Type 3 PRMTs can only carry out mono-methylation. Six PRMTs are classified as Type 1 enzymes; PRMT1, PRMT2 (UniProtKB: P55345), PRMT3 (UniProtKB: O60678), coactivator-associated arginine methyltransferase 1 (CARM1) also known as PRMT4 (UniProtKB: Q86X55), PRMT6 (UniProtKB: Q96LA8), and PRMT8 (UniProtKB: Q9NR22). In contrast, PRMT5 (UniProtKB: O14744) and PRMT9 (UniProtKB: Q6P2P2) are Type 2 PRMTs (Yang et al., 2015). PRMT7 (UniProtKB: Q9NVM4) was assigned as a Type 3 enzyme that is only able to generate mono-methyl-arginines (Feng et al., 2013), but different groups reported that it can also produce symmetrical di-methylation marks (Lee et al., 2005). However, this is not the case as shown by Feng and colleagues (Feng et al., 2014). Substrates of

the histone methyltransferase family included free histones, histones that complex DNA, and non-histone proteins (Bedford and Clarke, 2009).

1.3 Domain Architecture of PRMTs

The nine human PRMTs are variable in size (353-845 residues) and possess different domains but share a highly conserved catalytic core of approximately 310 amino acids as shown in Figure 1-7 (Cheng *et al.*, 2005). The catalytic core of all human PRMTs consist of a Rossmann fold which includes the cofactor binding pocket, and a β -barrel domain that is important for substrate binding (Cheng *et al.*, 2005). Most PRMTs have only one catalytic core region, but mouse PRMT7 and human PRMT9 have two copies of the catalytic domain, most likely caused by gene duplication (Miranda *et al.*, 2004, Yang *et al.*, 2015). Even if the catalytic core of PRMTs is very similar, PRMTs differ from one another in their N-terminal region. For example, PRMT3 has a unique zinc-finger domain (Cura *et al.*, 2014), CARM1 a pleckstrin homology (PH) domain (Troffer-Charlier *et al.*, 2007), PRMT2 contains a Src homology 3 (SH3) domain (Cura *et al.*, 2017) and PRMT9 has three tetratricopeptide repeats (TPR) (Yang *et al.*, 2015).

The N-terminal domains are involved in many different protein-protein interactions (Swiercz *et al.*, 2005) and influence enzymatic activity and biology (Sayegh *et al.*, 2007). They can also lead to the formation of higher oligomers and have been proposed to contribute to substrate recognition (Bedford and Clarke, 2009). For example, PRMT5 exists as a multimeric protein complex that influences its activity and specificity (Antonyamy *et al.*, 2012).

Over the last couple of years, other arginine methyltransferases have been identified including the Type 1 PRMT NADH:ubiquinone oxidoreductase complex assembly factor 7 NDUFAF7 (UniProtKB: Q7L592) and the Methyltransferase Like 23 protein METTL23 (UniProtKB: Q86XA0) (Fuhrmann *et al.*, 2015). The mitochondrial arginine methyltransferase NDUFAF7 symmetrically di-methylates the NDUF52 subunit of the human mitochondrial complex I (Rhein *et al.*, 2013). The Type 2 PRMT METLL23 was shown to catalyse asymmetric di-methylation of Histone H3 in mouse oocytes (Hatanaka *et al.*, 2017). Both share only some sequence identity with the classical PRMTs, with less than 20 % overall sequence identity.

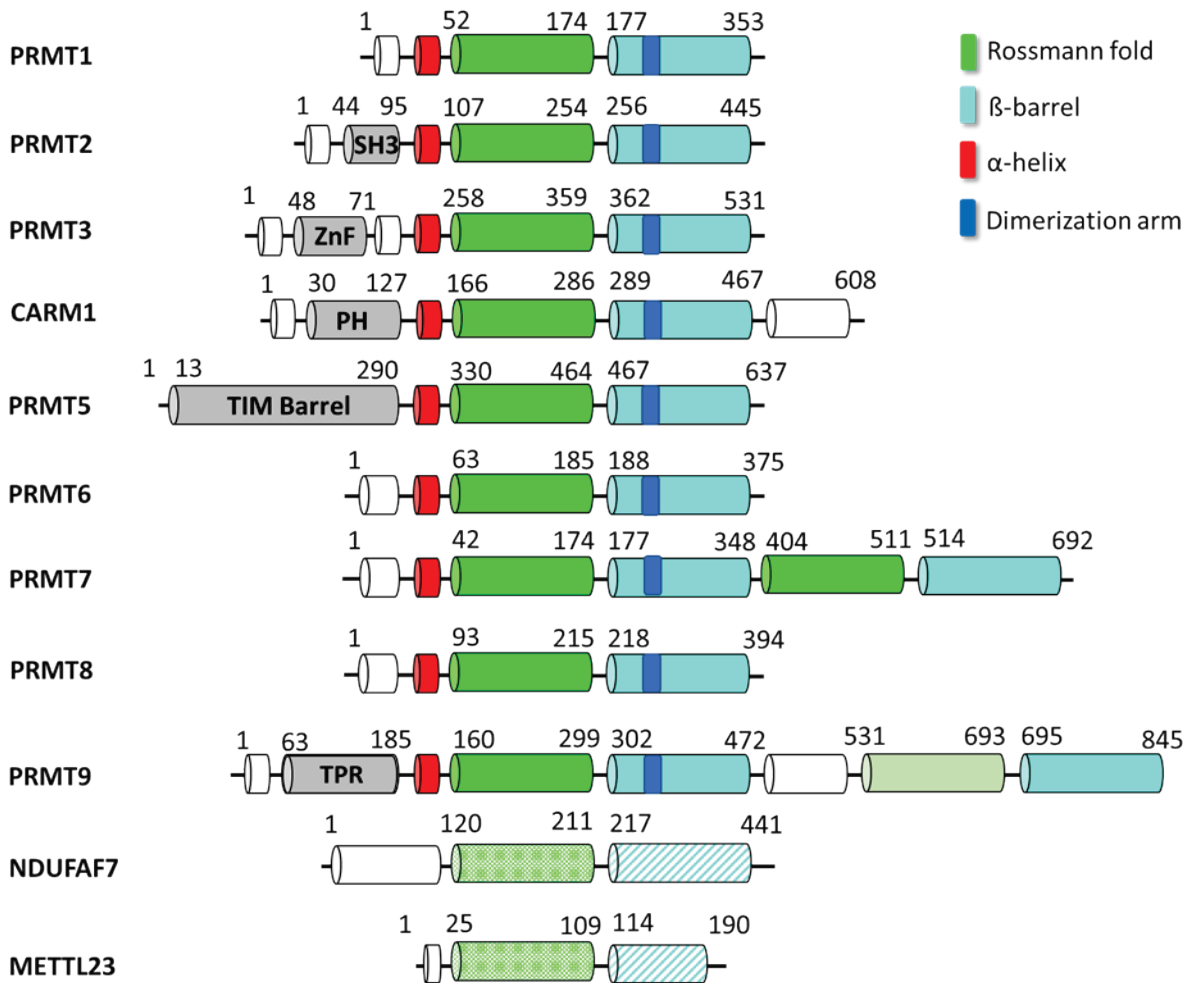


Figure 1-7 The PRMT Family.

There are nine classical PRMT family members, PRMT1-PRMT9 that differ in overall size and some of them carry a unique N-terminal domain. They all share a highly conserved core, consisting of a Rossmann fold (green) followed by a β -barrel domain (cyan) that includes a dimerisation arm (blue). An α -helix (red) is located at the N-terminus of the Rossmann fold that is involved in binding of the cofactor SAM. PRMT7 and PRMT9 have two Rossmann fold/ β -barrel domains as a result of gene duplication. The second Rossmann fold of PRMT7 is not able to bind SAM and thus has no enzyme activity. The domains are assigned based on protein structure alignment. Two novel PRMTs, NDUFAF7 and METTL23, were recently identified but they exhibit a low sequence identity with the classical PRMTs (<20 %).

In addition, two distantly related proteins, F-box only protein (FBXO)10 and FBXO11, that show a low degree of sequence homology (overall sequence identity across the PRMT family is 12-16 %) to some PRMT motifs but do not have the typical PRMT β -barrel fold but a β -propeller core domain, were identified as possible new PRMTs (Cook *et al.*, 2006, Krause *et al.*, 2007). However, FBXO11, which was temporary named PRMT9, lacks important substrate-binding motifs and no enzyme activity has been demonstrated so far, so it should not be included in the PRMT family (Fielenbach *et al.*, 2007). Instead it was shown that FBXO11 is a subunit protein of the

ubiquitin E3 ligase (Evankovich *et al.*, 2017). FBXO10, which was temporary named PRMT11, was identified as a homologue of FBXO11 (Krause *et al.*, 2007) and has E3 ubiquitin ligase activity (Abida *et al.*, 2007). Thus, FBXO10 and FBXO11 are not considered as PRMTs family members and even if the proteins possess methyltransferase activity, it most likely resulted from convergent evolution.

1.4 Structure of PRMTs

Crystal structures of different PRMTs including rat PRMT1, (PDB: 1OR8, Zhang and Cheng, 2003), mouse PRMT2, (PDB: 5FUL, Cura *et al.*, 2017), human PRMT3, (PDB: 4HSG, Liu *et al.*, 2013), human CARM1, (PDB: 2V74, Yue *et al.*, 2007), human PRMT6, (PDB: 5HZM, Wu *et al.*, 2016), mouse PRMT7, (PDB: 4C4A, Cura *et al.*, 2014), human PRMT5:MEP50, (PDB: 4GQB, Antonysamy *et al.*, 2012), human PRMT8, (PDB: 5DST, Toma-Fukai *et al.*, 2016) and human PRMT9, (PDB: 6PDM, to be published), have been solved. An overview of the different structures can be seen in Figure 1-8.

The crystal structures show that the overall fold of PRMTs is very similar and indicates that most PRMTs exist as head-to-tail homodimers with structurally and functionally distinct regions. The catalytic core comprises a methyltransferase (MTase) domain, a β -barrel domain (cyan), and a dimerisation arm (blue). The methyl transferase domain which is located within the N-terminal region of the catalytic core consists of alternating motifs comprised of β -strand- α -helix- β -strand, known as a $\beta\alpha\beta$ -Rossmann fold (green) (Bedford and Clarke, 2009). The Rossmann fold contains the cofactor binding fold that is highly conserved across all SAM-dependent methyltransferases. It consists of 4 α -helices and 5 β -strands (Martin and McMillan, 2002). The β -barrel domain is composed of 10 β -strands and is located at the C-terminus. The two domains are connected by a conserved *cis*-proline residue (Troffer-Charlier *et al.*, 2007).

In the case of Type 1 PRMTs, dimerisation is initiated by interaction between the dimerisation arm (blue), that is composed of three α -helix segments and is inserted between β -strand 6 and β -strand 7 of the β -barrel domain, and the Rossmann fold of another subunit as exemplified by the structure of CARM1 (Schapira and de Freitas, 2014, Cheng *et al.*, 2005).

Dimerisation was thought to be necessary for all PRMTs to possess enzyme activity. However, for mouse and *C. elegans* PRMT7 and for human PRMT9 it was shown that the protein contains two catalytic core units and mainly exists in a monomeric form (Jain and Clarke, 2019, Yang *et al.*, 2015). The crystal structures of mouse PRMT7 (PDB: 4C4A) and human PRMT9 (PDB: 6PDM) in complex with SAH, shown in Figure 1-8, revealed, that, in each case, one of their catalytic core domains folds back onto the other, generating a pseudo-dimer. In the case of PRMT7 it was shown, that only the first domain has a SAM molecule bound in the catalytic domain and thus has enzyme activity, the second carries several mutations that prevent SAM binding (Cura *et al.*, 2014). Surprisingly, the duplication of the second SAM-binding-like domain is missing in the *T. brucei* PRMT7 structure (Wang *et al.*, 2014).

In the case of Type 1 PRMTs a dynamic α -helix (red) is located at the N-terminus of the Rossmann fold. This α -helix folds over and buries the cofactor SAM/SAH (Adams-Cioaba and Min, 2009, Morales *et al.*, 2016, yellow spheres, Figure 1-7). In contrast, in the Type 2 enzyme PRMT5 the β -barrel domain and the Rossmann fold are present, but instead of the N-terminal α -helix, PRMT5 has a short α -helical segment that is followed by a loop and connected to a Triosephosphate isomerase (TIM) barrel domain (shown in light pink in Figure 1-7). This domain interacts with Methylosome protein 50 (MEP50, yellow, Figure 1-7). PRMT5 forms a large hetero-octameric complex with the WD40 protein MEP50 which interacts with binding partners and substrates (Ho *et al.*, 2013). Moreover, PRMT5 does not dimerise via a dimerisation arm, but instead via interactions of the TIM domain and the Rossmann fold and linker region of another subunit (Antonyamy *et al.*, 2012).

All PRMTs have four conserved motifs named motif I, post-I, II, and III (Schluckebier *et al.*, 1995b), as exemplified by the structure of human CARM1 (Figure 1-9). Three of these conserved motifs, which are common to seven β -strand enzymes, are located in the Rossmann fold, named motif I (VLD/EVGxGxG), motif post-I (L/V/lxG/AxD/E) and motif II (F/I/VDI/L/K). Motif III (F/I/VDI/L/K) is located in the β -barrel domain. Additionally, PRMTs have two PRMT-specific motifs, the Threonine-Histidine-Tryptophan (THW) loop and the double E-loop. An additional structural motif IV (YFxxY), which only exists in Type 1 PRMTs, is located in the N-terminal α -helix (Zhang *et al.*, 2000). The location of the motifs (pink) can be seen in Figure 1-9.

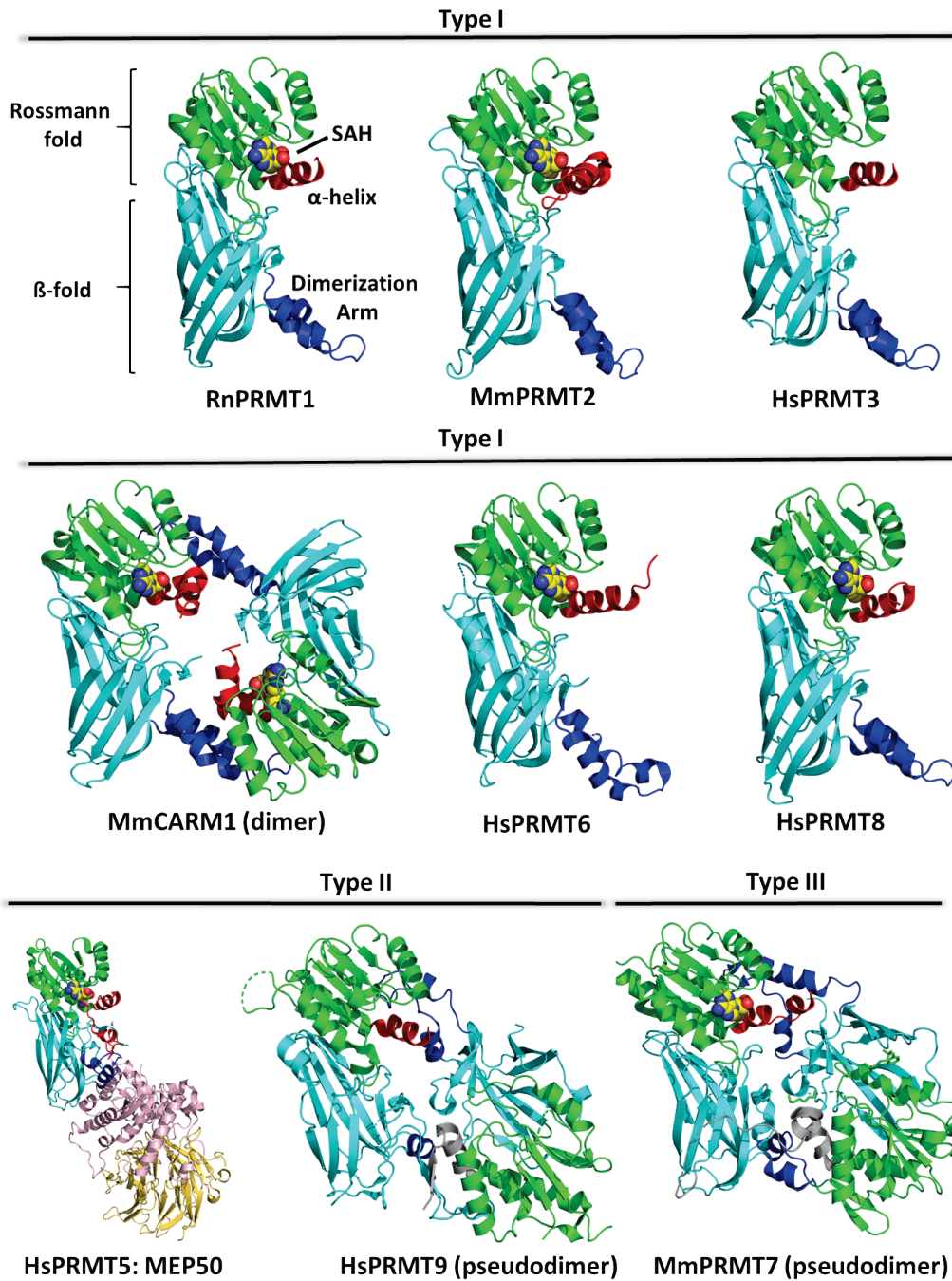


Figure 1-8 Overview of PRMT Crystal Structures.

All PRMTs share a highly conserved catalytic core, which is composed of a $\beta\alpha\beta$ -Rossmann fold (green) and a β -barrel domain (cyan). The dimerisation arm (blue) of the β -barrel domain and the dynamic α -helix (red) located at the N-terminus of the Rossmann fold are important for the formation of subunit contacts. PRMT5 needs MEP50 (yellow) for full enzyme activity and has an N-terminal TIM barrel domain (light pink). PRMT7 and 9 monomers comprise two catalytic core domains (N and C-module) connected by a linker. In the case of PRMT7 only the N-terminal Rossmann fold is able to bind SAM, the second active site in the C-module is inactive. PDB codes: PRMT1: 1OR8; PRMT2: 5FUL; PRMT3: 4HSG; CARM1: 2V74; PRMT5:(MEP50: 3SMQ; PRMT6: 5HZM; PRMT7: 4C4A; PRMT8: 5DST; PRMT9: 6PDM. Abbreviations: Hs= Homo sapiens; Mm= Mus musculus; Rn= Rattus norvegicus.

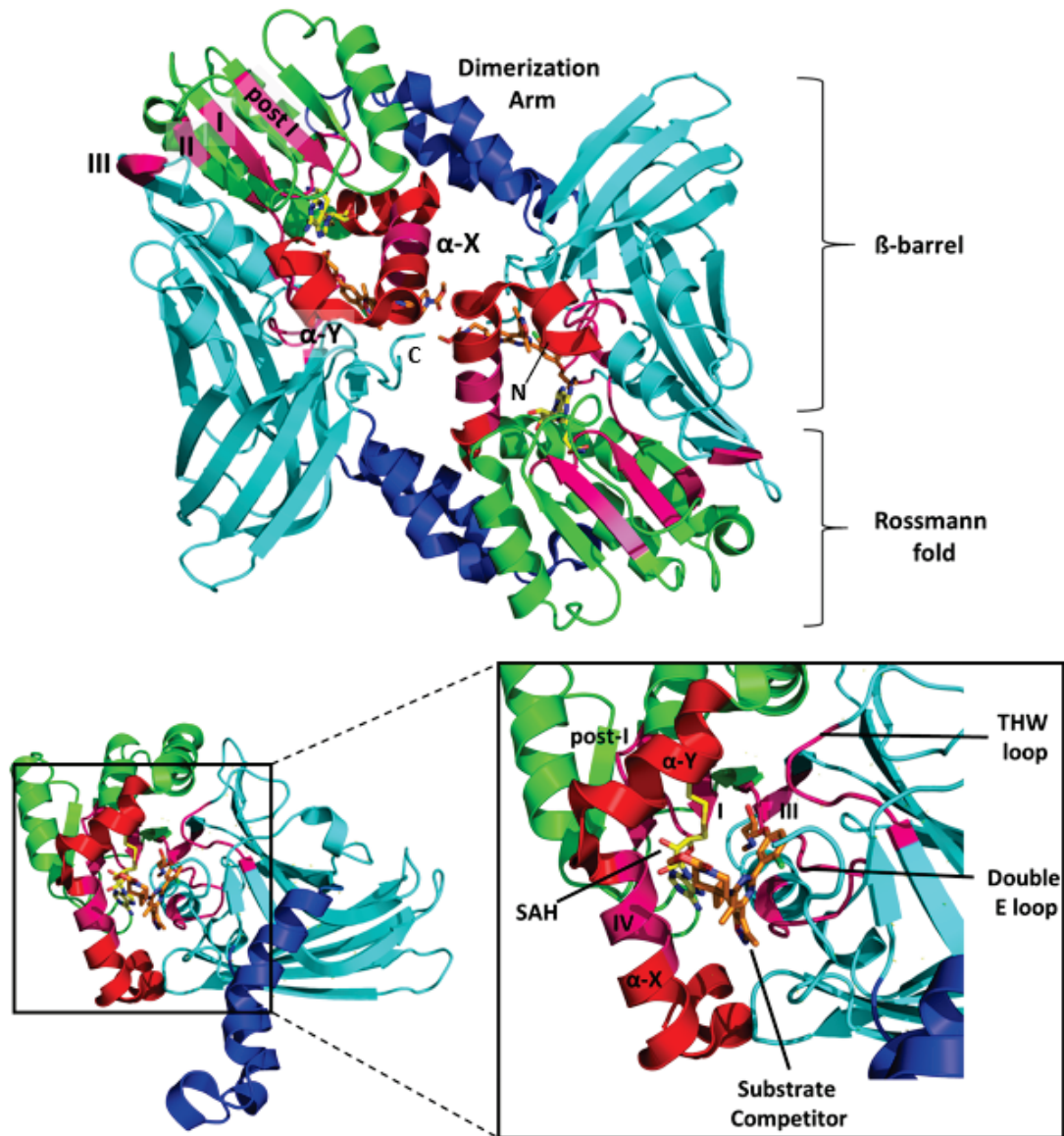


Figure 1-9 Structural Overview of Type 1 PRMTs.

The typical dimeric structure of Type 1 PRMTs is illustrated with the CARM1 structure in complex with SAH (yellow) and a substrate inhibitor (orange) (PDB: 6ARJ). The cofactor binding pocket is located in the Rossmann fold (green). A flexible α -helix (red) at its N-terminus, consisting of two elements α -X and α -Y, folds over the cofactor. The dimerisation arm (blue) of the β -barrel domain (cyan) interacts with the Rossmann fold of the other subunit. The substrate binding pocket is located between the Rossmann and β -barrel domain. All PRMTs have conserved motifs (pink), named motif I, post-I, II, and III that are common to seven β -strand enzymes. In addition, two PRMT specific motifs- the Threonine-Histidine-Tryptophan (THW) loop and the double E-loop exist.

Motif I, post-I and the THW loop are involved in binding of the cofactor SAM. Motif II forms a parallel β -sheet which helps to stabilise motif I, and motif III forms a parallel β -sheet with motif II (Bedford and Clarke, 2009). The THW-loop is adjacent to the double E-loop, which contains two invariant glutamate residues and is involved in binding of the arginine guanidinium group. Together they form the active site. The

THW-loop also helps to stabilise an N-terminal α -helix (red) which is located at the N-terminus of the Rossmann fold. The α -helix consists of two segments, α -X and α -Y, and folds over the cofactor, burying it in the core of the Rossmann fold. This structurally dynamic α -helix that is also involved in the formation of subunit contacts in the active site was observed in many Type 1 PRMT structures including PRMT1 (PDB: 1OR8), PRMT2 (PDB: 5FUL), PRMT3 (PDB: 4RYL), CARM1 (PDB: 2Y1W) and PRMT6 (PDB: 4Y30) (Schapira and de Freitas, 2014). In the case of CARM1, it is thought to trigger the formation of the substrate-binding site at the interface of the Rossmann fold, the β -barrel domain, and the α -helix (Troffer-Charlier *et al.*, 2007).

Secondary structural elements that are located in the same place as the α -helix in Type 1 PRMTs and are involved in SAM binding have also been observed in other SAM-dependent methyltransferases that have a catalytic Rossmann fold including Type 2 and 3 PRMTs, DNA (DNMTs) and RNA Methyltransferases (RNMTs), Catechol-O-methyltransferase (COMT) and DOTL1L (Schapira and de Freitas, 2014). In DNMTs (Figure 1-10 B), it is a α -helix similar to Type 1 PRMTs (Figure 1-10 A/C). In the case of the Type 2 protein PRMT5 (Figure 1-10 D), RNA methyltransferases (Figure 1-10 E), COMT (Figure 1-10 F) and the PKMT DOT1L (Figure 1-10 G), it is a short α -Y element which is followed by a loop and is located between the Rossmann fold and the specific N-terminal domain (Antonysamy *et al.*, 2012, Min *et al.*, 2003). Crystal structures showed that these dynamic structural elements can adopt many different conformations and orientations and they are often disordered in the absence of the cofactor (Troffer-Charlier *et al.*, 2007, Boriack-Sjodin *et al.*, 2015).

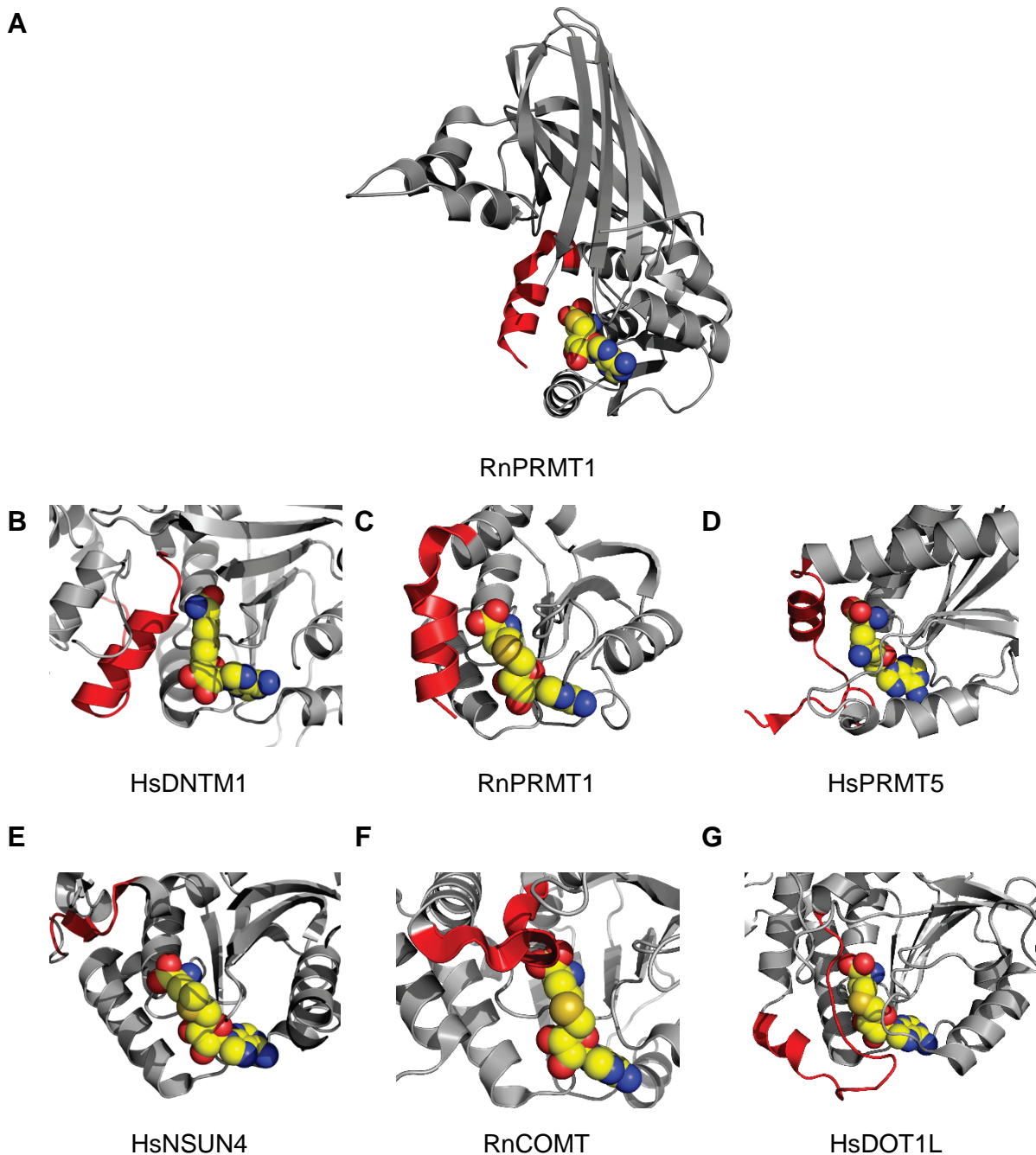


Figure 1-10 Regulatory Elements of the Rossmann Fold.

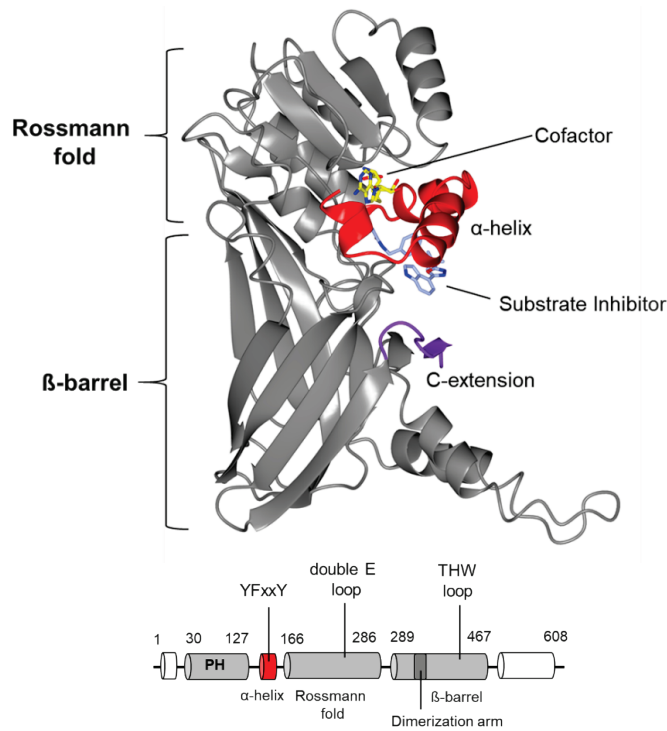
A) Overall MTase structure, exemplified by PRMT1 monomer. Similar to the dynamic α -helix (red) that is located at the N-terminal of the Rossmann fold in PRMTs (compare C and D), other SAM-dependent MTases have similar secondary structural elements (red) at the same position that might influence substrate specificity. PDB codes: HsDNTM1 (3PTA), RnPRMT1 (1OR8), HsDOT1L (3QOX), RnCOMT (4PYN). NSUN4 (4FP9). The carbon atoms in the co-factor SAM structure are shown as yellow spheres, nitrogen blue, oxygen red.

1.5 Active Site and Substrate Specificity of Type 1 PRMTs

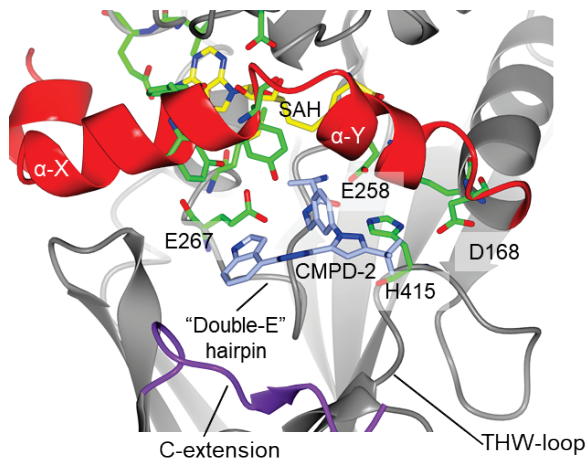
The overall architecture of the active site of Type 1 PRMTs is very similar despite their differences in overall size and sequence. As mentioned before, there are several conserved motifs within the active site that are essential for catalysis. Figure 1-11 shows the active site of CARM1, it can be divided into two regions- the SAM and the arginine-peptide binding site (Schapira and de Freitas, 2014). The conserved key residues include the two invariant glutamate residues (Glu258 and Glu267) of the so-called double E-loop and His414 of the THW-loop. These residues are involved in positioning and binding of the substrate arginine guanidine-group (Zhang and Cheng, 2003). The side chain of Glu267, the carbonyl oxygen of Glu258 and the side chain of His415 bind the guanidine-group via hydrogen bonds and position it in such a way that the S_N2-like attack from the methyl-sulfonium group of SAM can take place.

The additional structural motif, motif IV (Tyr150, Phe151 and Tyr154 in CARM1), which only exists in Type 1 PRMTs, forms π -interactions with the aromatic rings of the cofactor SAM, the tyrosine side chain forms hydrogen bond interactions with one of the E-loop glutamate residues (Glu267). An additional methionine residue (Met163 in CARM1) of the α -helix, helps to keep the substrate arginine residue in place. This arrangement forms a large empty space around one of the two arginine ω -nitrogen atoms, that allows Type 1 PRMTs to either bind unmethylated or mono-methylated arginine for mono-methylation or asymmetrical di-methylation, respectively. In contrast, the other arginine ω -nitrogen forms hydrogen bond interactions with His415 and Glu258 and is in close proximity to Met163 and does not have the space to allow methylation. The SAM binding site also has many conserved residues that are involved in cofactor binding. In the case of HsCARM1 these residues include Glu244 and Val243 which interact with the adenine ring of the cofactor and Glu215 which forms a hydrogen bond to the ribose moiety. The carboxyl group of SAM forms hydrogen interactions with Arg169 and with the side-chain of glutamate residue Glu258 (Schapira and de Freitas, 2014, van Haren *et al.*, 2017).

A



B



C

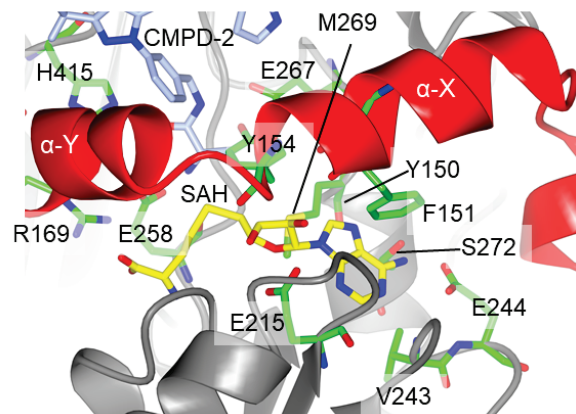


Figure 1-11 Active Site of the Crystal Structure of CARM1.

A) Structural overview and schematic diagram of the modular organisation of full-length HsCARM1. The Rossmann-fold (green) and β -barrel domain (blue) are highly conserved among the PRMT family. The location of the three conserved motifs of Type I PRMTs is also indicated. Cofactor and arginine substrate bind in two distinct pockets which show high sequence conservation. **B)** The conserved residues that are important for binding and positioning of the substrate arginine include the double E-loop residues E258, E267 and H414 of the THW-loop. **C)** The conserved structural motif YFxxY (Y150, F151 and Y154 in CARM1) of Type 1 PRMTs (where x represents any amino acid) stacks with the aromatic rings of the cofactor SAH (yellow) and allows the formation of a hydrogen bond between the Y154 and E267. They generate enough volume around the methyl-accepting nitrogen atom that unmethylated and mono-methylated substrate arginines can be accommodated in the active site (Schapira and de Freitas, 2014). CARM1 PDB code: 2Y1X (SAH and indole inhibitor CMPD-2).

1.6 Catalytic Mechanism of Type 1 Protein Arginine Methyltransferases

The catalytic mechanism of the protein arginine methylation reaction has been studied in different PRMT family members (Schapira and de Freitas, 2014, Tewary *et al.*, 2019). For PRMT1, PRMT5, and PRMT6, it was shown that catalysis follows a rapid-equilibrium random kinetic mechanism, which means that substrate binding and product release occur randomly (Wang *et al.*, 2013b, Obianyo *et al.*, 2008, Obianyo and Thompson, 2012). Moreover, it was shown that PRMT1 methylation is not a processive mechanism, but that the mono-methylated intermediate is released from the active site and needs to rebind for the second methylation step. In the case of CARM1, an ordered mechanism for arginine methylation was first proposed after comparison of the CARM1 crystal structure in apo and holo state complexed with SAH (Yue *et al.*, 2007). According to the authors, the first step is the binding of the cofactor SAM, which introduces large conformational changes and leads to the formation of the substrate binding pocket. After transfer of the methyl-group, the mono-methylated product is then released, and a new SAM molecule can bind. However, a newer kinetic study suggests that CARM1 also methylates H3 peptides in a random kinetic mechanism (Jacques *et al.*, 2016).

The catalytic mechanism of the methyl group transfer of Type 1 PRMTs proceeds via a bimolecular nucleophilic substitution reaction (S_N2) and its proposed mechanism based on structural studies and mutagenesis experiments with PRMT1 is shown in Figure 1-12 (Rust *et al.*, 2011, Zhang *et al.*, 2013). The two invariant glutamate residues Glu144 and Glu153 (Glu267 and Glu258 in CARM1) form hydrogen bonds with the guanidine group of the substrate arginine so that the positive charge is localized to one of the two terminal nitrogen atoms of the guanidine group.

As a consequence, the other ω -nitrogen atom has a lone electron pair and is able to perform a nucleophilic attack on the methyl-sulfonium group of SAM. After transfer of the methyl group from the cofactor to the arginine residue, the proton is abstracted by the E-loop glutamate residue Glu144. The reaction results in the production of the N-methylated arginine and the generation of the by-product SAH (Zhang *et al.*, 2013, Rust *et al.*, 2011).

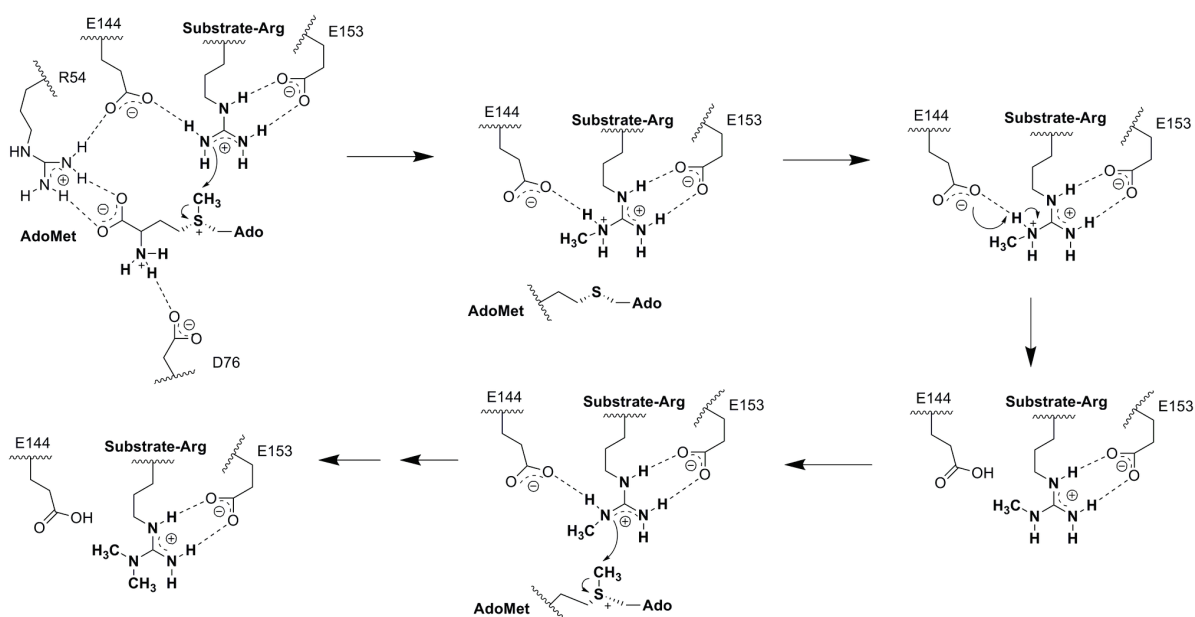


Figure 1-12 Catalytic Mechanism for Type 1 PRMTs, Exemplified by PRMT1.

The two glutamate residues of the E-loop position the methyl accepting nitrogen atom and facilitate the nucleophilic attack of the partially negatively charged nitrogen atom onto the methyl-sulfonium group of the cofactor SAM. After transfer of the methyl-group from the cofactor SAM to the arginine residue, the deprotonation of the arginine occurs. The Glu₁₄₄ residue acts as proton acceptor. Figure taken from (Zhang *et al.*, 2013).

1.7 Consequences of Arginine Methylation

The methylation reaction catalysed by PRMTs changes the molecular mass of the histones by 14 Da or 28 Da in the case of mono-methylation or di-methylation, respectively. The alteration of the side chain causes an increase in steric hindrance (Fuhrmann *et al.*, 2015). Moreover, the methylation causes a decrease in possible hydrogen bond interactions due to the reduction in hydrogen bond donor sites (Bedford *et al.*, 2000). It also increases not only bulkiness but also hydrophobicity of the side chain by withdrawing electrons through hyperconjugation. In contrast, as seen in Figure 1-13, the overall electrostatic charge of the guanidine groups is unchanged (Lorton and Shechter, 2019).

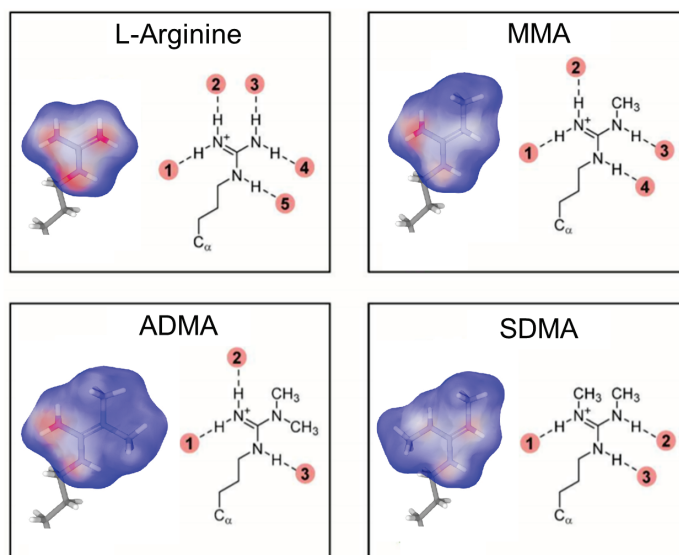


Figure 1-13 Hydrogen Bond Donor Sites and Electrostatic Surface Potential of Unmethylated and Methylated Arginine Side Chains.

Electrostatic surface potentials are contoured from -4 (red) to 4 (blue) $kBT e^{-1}$. Hydrogen bond donor sites are numbered in red. MMA: Mono-methylated Arginine, ADMA: Asymmetrical di-methylated Arginine, SDMA: Symmetrical di-methylated Arginine, C_{α} : α -carbon. Figure adapted, with permission, from Lorton and Shechter (2019).

1.8 Arginine Methyltransferase Substrate Recognition

The first PRMT substrates identified were histone proteins (Chen *et al.*, 1999). Since then, many non-histone proteins have been identified, mainly for PRMT1 and CARM1 (Bedford and Clarke, 2009). However, many other PRMTs are yet to have substrates or interactors identified.

Arginine methylation usually occurs within glycine-, alanine-, and arginine-rich (GAR) domains, which are also called RGG boxes (Bedford and Clarke, 2009). These domains consist of single or multiple RGG/RG motifs (Thandapani *et al.*, 2013). Exceptions are PRMT7 and CARM1, which cannot methylate GAR sequences. Instead, CARM1 prefers proline, glycine, methionine, and arginine-rich sites (PGMs) (Shishkova *et al.*, 2017) and PRMT7 modifies substrates containing RxR motifs, two arginine residues that are separated by a basic residue (Branscombe *et al.*, 2001, Zhang *et al.*, 2013, Yang and Bedford, 2013). PRMT5 can methylate GAR and PGM motifs, and PRMT6 modifies arginine residues within GAR motifs but also non-GAR motifs (Boulanger *et al.*, 2005, Singhroy *et al.*, 2013).

Histone methylation has been seen for PRMT1, PRMT2, PRMT3, CARM1, PRMT5, PRMT6, PRMT7 and PRMT8 (Di Lorenzo and Bedford, 2011, Fuhrmann and Thompson, 2016). An overview of the different histone methylation sites is given in Figure 1-14. The steroid receptor coactivator CARM1 was the first PRMT that was shown to influence gene transcription through asymmetric di-methylation of Arg17 and Arg26 on histone H3 (Schurter *et al.*, 2001). CARM1 also shows low levels of methylation activity towards Arg2 of histone H3 (Chen *et al.*, 1999, Schurter *et al.*, 2001). Methylation of Arg17 lead to transcriptional activation, whereas methylation of Arg26 caused transcriptional repression (Zhang *et al.*, 2017b, Yang *et al.*, 2014). PRMT2 was shown to have weak methyltransferase activity on histone H4 *in vitro* (Lakowski and Frankel, 2009) and is able to asymmetrically di-methylate Arg8 in histone H3 *in vivo* (Dong *et al.*, 2018).

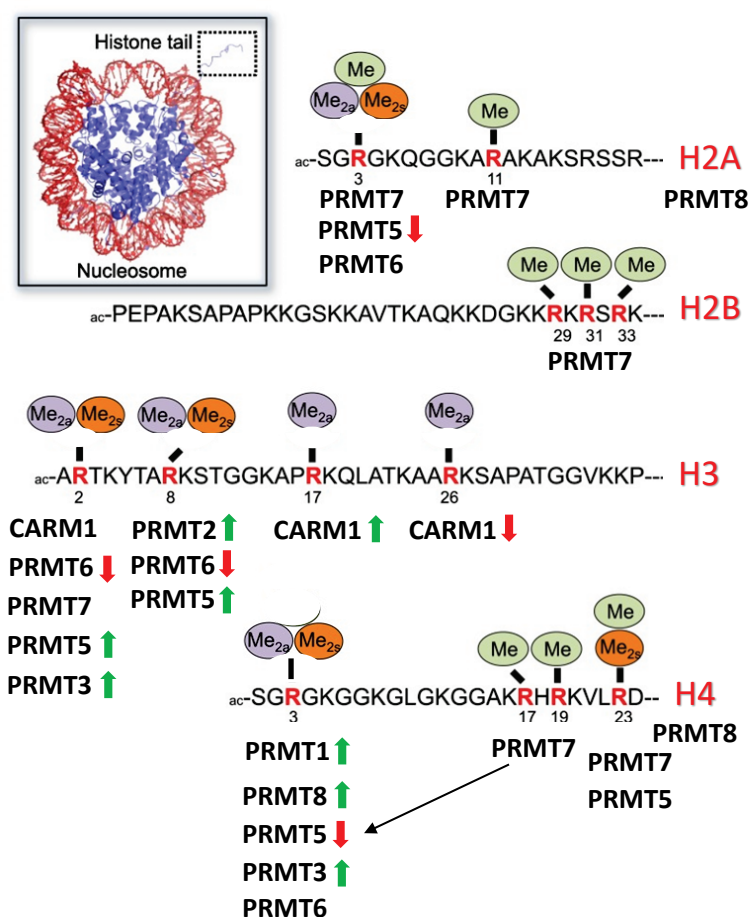


Figure 1-14 Arginine Methylation Sites on N-terminal Histone Tails.

PRMTs methylate arginine residues and, depending on their type, produce mono-methylation (MMA), asymmetric di-methylation (ADMA) or symmetric di-methylation (SDMA). Figure adapted, with permission from Fuhrmann and Thompson (2016)*.

*Further permissions related to the material excerpted should be directed to the ACS. <https://pubs.acs.org/doi/10.1021/acscchembio.5b00942>

PRMT1 and PRMT8 were shown to activate transcription by di-methylation of Arg3 of histone H4 (Li *et al.*, 2010, Di Lorenzo and Bedford, 2011). Moreover, PRMT5 can regulate gene expression via methylation of different arginine residues of histone H2, H3 and H4, which result in either transcription activation or repression (Chen *et al.*, 2017). PRMT6 was shown to methylate Arg2 and Arg8 *in vivo* and is the main methyltransferase on histone H3 Arg2 causing a decrease in transcription (Hyllus *et al.*, 2007, Iberg *et al.*, 2008). The Type 3 enzyme PRMT7 mono-methylates Arg17 and Arg19 of histone H4. Methylation of Arg17 is also thought to activate PRMT5-mediated methylation of Arg3 on the same histone (Jain and Clarke, 2019). PRMT8 methylates histones H2A and H4 *in vitro* (Sayegh *et al.*, 2007).

Very recently putative new PRMT substrates were identified in a study using BioID technology (Roux *et al.*, 2013) for screening of physiologically relevant PRMT-protein interactions (Wei *et al.*, 2019). Some of the identified substrates had new motifs such as SR-, PR-, ER- and DR- rich motifs. The same study also identified potential interaction partners/substrates of PRMTs which were mainly RNA binding proteins that are involved in RNA splicing and translation. Inhibition of selected PRMTs (CARM1 and PRMT1) caused a global change in alternative mRNA splicing and a reduction in mRNA translation (Wei *et al.*, 2019). All motifs have in common that they are short and their sequence is not very complex, which might indicate that the substrate interaction regions of PRMTs are unstructured (Wei *et al.*, 2019).

1.9 Methyl-arginine Binding Proteins- The Readers of Arginine Methylation

Arginine methylation of GAR and PGM motifs by PRMTs leads to the interaction of those motifs with proteins mainly via a Tudor domain but protein binding via PHD and WD40 repeat (WDR) domain-containing proteins has also been reported (Migliori *et al.*, 2012, Selenko *et al.*, 2001). Tudor domains belong to the so-called “royal superfamily” of domains which also include Chromo domains, Malignant Brain Tumor (MBT) domains, PWWP domains, which are named after a conserved Pro-Trp-Trp-Pro motif, and Agenet-like domains. Their royal domain core shares a high degree of structural and sequence similarity (Jin *et al.*, 2009) and they are all able to read protein methylation marks (Figure 1-15) (Maurer-Stroh *et al.*, 2003, Zhao *et al.*, 2009, Botuyan and Mer, 2016).

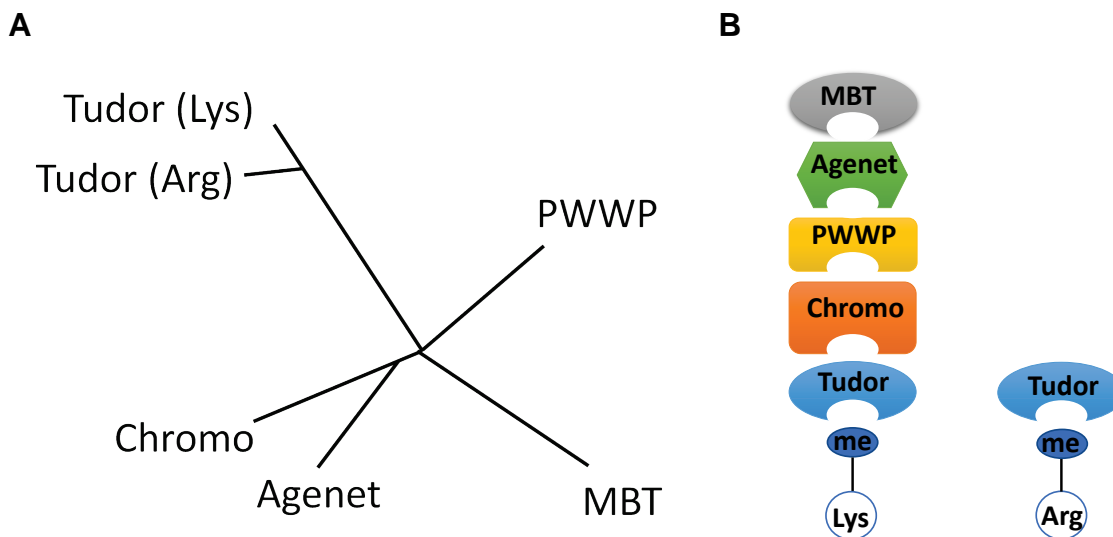


Figure 1-15 Royal Superfamily Domains- The Reader of Methylation Marks.

A) Phylogenetic tree of the family members based on sequence homology of the royal domain cores. **B)** Schematic overview of the superfamily domains and their specificity for lysine or arginine methylation marks. Tudor domains are the only domains that are specialised in the recognition of methyl-arginines. Figure adapted from Chen *et al.* (2011).

The Tudor family is the only member of the royal superfamily that is specialised in the recognition of methylarginines and comprises ~36 proteins in mammals, which can be divided into methylarginine and methyl-lysine binding Tudor proteins (Maurer-Stroh *et al.*, 2003). Discrimination of the two groups is not possible via their sequence but via the structural composition of the binding site. The Tudor domains are approximately 60 amino acids in length and contain a β -barrel core which consist of four anti-parallel β -sheets (Figure 1-16 A). The core region can be flanked by additional structural elements which differ among family members. Moreover, mammalian Tudor proteins can have a single Tudor domain or multiple tandem repeats (Chen *et al.*, 2011).

Most Tudor domains have an aromatic cage at the surface which allows methylarginine or methyllysine docking (Maurer-Stroh *et al.*, 2003, Adams-Cioaba and Min, 2009). Methylarginine binding Tudor domains have a narrower aromatic cage than methyllysine binding Tudor domains which is why binding of the planar methylguanidine group of the arginine is more likely (Liu *et al.*, 2012b). Tudor proteins that bind methyl-arginine include the Survival Motor Neuron (SMN) protein (Friesen *et al.*, 2001), Splicing factor 30 (SPF30) (Côté and Richard, 2005), Tudor domain-containing (TDRD) proteins (Gan *et al.*, 2019) and human spindlin (SPIN) proteins (Zhao *et al.*, 2007).

In humans, 12 TDRD proteins with single or multiple Tudor domains exist. They differ from SMN proteins that contain a single Tudor domain, by having an additional N-terminal α -helix and two β strand extensions (Jin *et al.*, 2009). SMN and SPF30 proteins bind to asymmetrical and symmetrical di-methylated arginines but prefer symmetric di-methylated (SDMAs), whereas the TDRD proteins differ in their arginine methylation preferences (Liu *et al.*, 2012a, Tripsianes *et al.*, 2011). For example, SND1 prefers SDMAs (Liu *et al.*, 2010), whereas TDRD9 prefers asymmetric di-methylated (ADMAs) (Zhang *et al.*, 2017a). SMN proteins are also the only known Tudor proteins that can bind to both methylated arginine and lysine residues with the same aromatic cage (Sabra *et al.*, 2013, Tripsianes *et al.*, 2011). The proto-oncogene spindlin1 (SPIN1) protein can also bind to a tri-methylated lysine residue of histone 3 and at the same time interact with asymmetrically di-methylated arginine residues on the same histone (Arg8) via a different domain (Su *et al.*, 2014, Janecki *et al.*, 2018). The SPF30 Tudor domain is closely related to SMN proteins, and shares over 35 % sequence identity (Talbot *et al.*, 1998). TDRD3 is capable of recognizing mono-methylated arginines but prefers ADMA marks (Yang *et al.*, 2010).

Figure 1-16 shows the NMR structure of the Tudor domain of the SMN protein bound to an asymmetrically di-methylated arginine (ADMA) residue (Figure 1-16 B) and to a symmetrically di-methylated arginine (SDMA) of a peptide (Figure 1-16 C) (Tripsianes *et al.*, 2011). The aromatic cage of the Tudor domain consists of different aromatic residues that close around the planar guanidine group of the di-methyl-arginine. The aromatic rings of the binding site residues bind the guanidine group via cation- π -interactions with the SDMA cation.

Comparison of ADMA and SDMA structures showed that the cationic carbon of SDMA is in a more favourable position for cation- π -interaction compared to the ADMA cation. This results in an increase in cation- π -stabilisation thus favouring binding of SDMA. Moreover, it was shown that methylation of the arginine residue increases cation- π interaction compared to non-methylated arginines, which could explain the high specificity of the Tudor protein for DMAs (Tripsianes *et al.*, 2011).

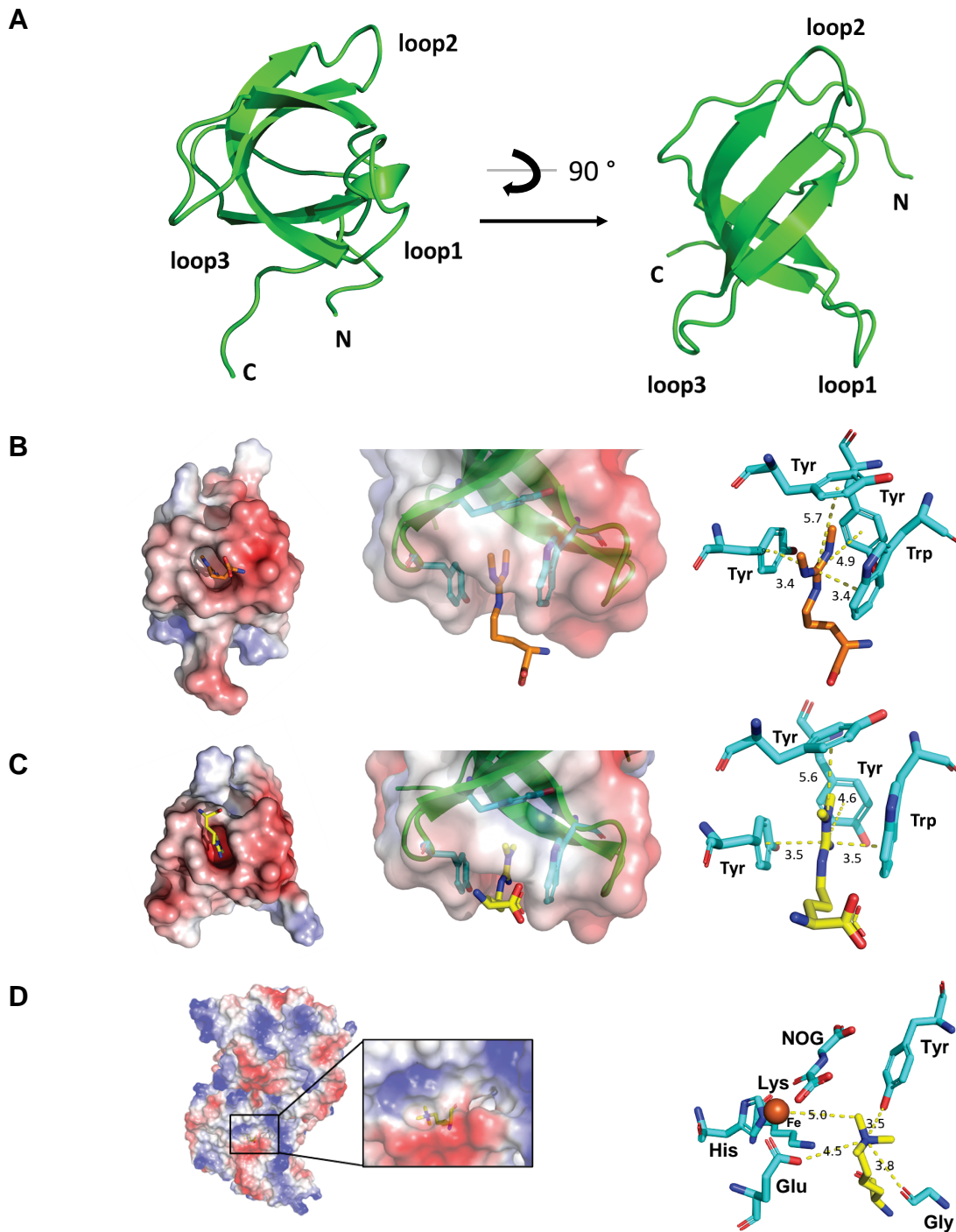


Figure 1-16 Tudor Domain Structure of SMN and its Recognition of SDMA and ADMA.

A) The overall structure of SMN is shown as green cartoon from two different angles (PDB: 4A4E). An aromatic cage is formed by the binding site residues around the guanidine group of the substrate arginine residue. They form cation- π interactions with the SDMA (**B**, orange) and the ADMA (**C**, yellow) cation. SDMA (**B**) is in a more favoured position increasing the electrostatic stabilization compared with ADMA (**C**) and as a result has a higher binding affinity (Tripsianes et al., 2011). **D)** Aromatic cage of JMJD2A bound to Fe(II) and N-oxalylglycine (NOG) and a tri-methylated H3K36 peptide (yellow). Figures were prepared in Pymol using SMN in complex with an ADMA (PDB: 4A4G) or SDMA (PDB: 4A4E) peptide and JMJD2A in complex with H3K36 (PDB: 2P5B).

Figure 1-16 D shows the crystal structure of Jumonji C-containing oxygenase JMJD2A, that can demethylate tri- and di-methylated lysine residues of histone 3. In the structure the protein is bound to Fe(II) and N-oxalylglycine (NOG) and a trimethylated H3K36 peptide, which is the state of the hydroxylation reaction (Chen *et al.*, 2007). The trimethyl group methylated is deeply buried in the catalytic pocket and in a polar environment formed by different residues including serine, tyrosine, glycine, and glutamate.

Methyl-arginine binding Tudor proteins are involved in different processes such as RNA metabolism (TDRD1, 2, 6, 9), regulation of gene transcription (TDRD3, 11) and mRNA splicing (SMN, SPF30) (Gayatri and Bedford, 2014). The subset of Tudor domains that have methyl-lysine binding domains (Figure 1-15 B) and that are able to bind methyl-lysine residues, are mainly involved in chromatin biology via recruitment of other chromatin regulating proteins to their binding sites (Chen *et al.*, 2011).

1.10 Arginine Demethylases, Erasers of Methyl-arginine Marks?

Histone methylations were shown to be a reversible reaction. Enzymes that catalyse the removal of methyl groups on histone arginine or lysine residues are called demethylases (Zhang *et al.*, 2019). Different histone lysine demethylases have been identified but until now no arginine-specific demethylases are known. Lysine-specific histone demethylases (KDMs) can be divided into two families, flavin-dependent KDM1, also called LSD1 (Shi, 2007), and Fe(II)- and 2OG-dependent Jumonji C-domain (JMJD)-containing enzymes, for example KDM2A. Both groups have been intensely studied (Tsukada *et al.*, 2006, Kang *et al.*, 2017).

In contrast, only a few enzymes that can demethylate histone arginine residues, including peptidyl deiminase 4 (PAD4) (Wang *et al.*, 2004) and the dioxygenase JmjC-domain-containing protein 6 (JMJD6) (Chang *et al.*, 2007) have been identified. Their dynamics and regulation are not yet fully understood. No specific arginine demethylase is known so far (Zhang *et al.*, 2019). The Ca²⁺-dependent PAD4 enzyme catalyses the conversion of peptidyl-arginine to peptidyl-citrulline (Bicker and Thompson, 2013). A structural overview of the enzyme and the catalysed reaction are given in Figure 1-17. Citrullination changes the structure and stability of the target proteins and thus influences its function.

Upregulation of different PAD members including PAD4 has been observed in a number of diseases including cancer and Alzheimer's (Chang and Han, 2006, Olsen *et al.*, 2018). The deiminase PAD4 has also been shown to remove the methyl group from mono-methylated arginine residues of histone H3 and H4 by converting them into citrulline and methyl-citrulline *in vitro* and *in vivo* (Kearney *et al.*, 2005, Thompson and Fast, 2006). Its target sites included methylated Arg3 of histone H4, produced by PRMT1, and Arg17 of histone H3, methylated by CARM1 (Wang *et al.*, 2004). The two possible PAD4 reaction demethylation reaction mechanisms are shown in Figure 1-17.

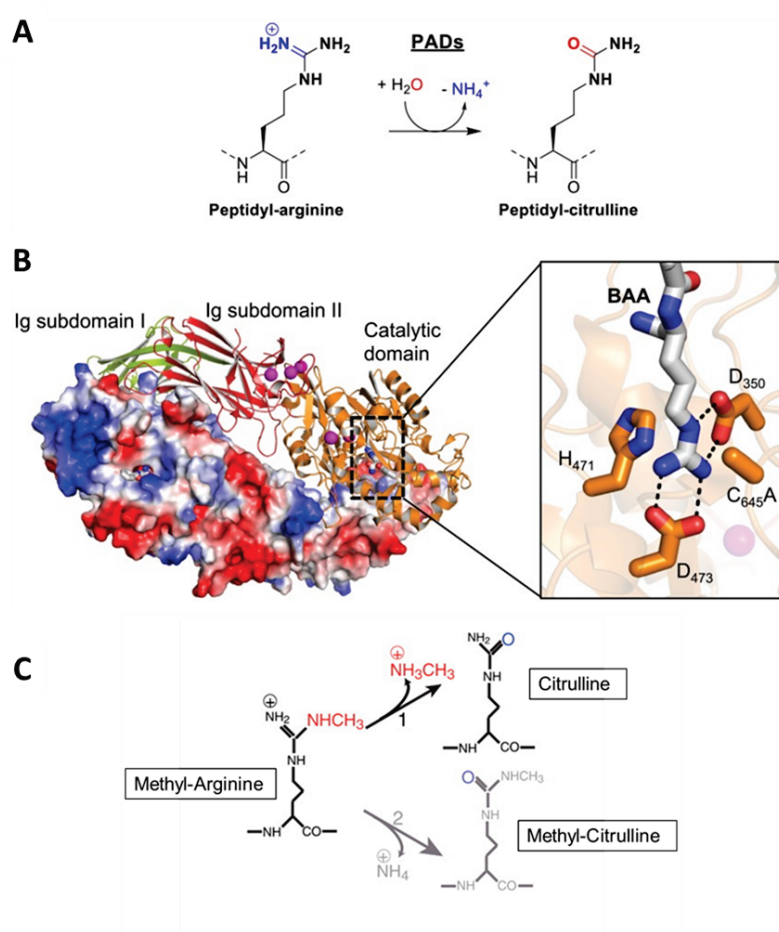


Figure 1-17 Structure and Function of PADs.

A) Ca²⁺-dependent PAD enzymes catalyse the citrullination reaction and convert peptidyl-arginine into peptidyl-citrulline. **B)** Dimeric crystal structure of PAD4 (PDB: 1WDA) with a bound arginine mimicking substrate benzoyl-L-arginine amide (BAA). **C)** PAD4 can also remove mono-methylated arginine residues on H3 and H4 histones. Two possible mechanism are shown: 1) PAD4 removal of the Methyl-imine group from Methyl-arginine, producing Citrulline and Methyl-amine; 2) Removal of the Imine group from Methyl-arginine, producing Methyl-citrulline and ammonium. Figure adapted, with permission, from Fuhrmann and Thompson (2016)*.

*Further permissions related to the material excerpted should be directed to the ACS. <https://pubs.acs.org/doi/10.1021/acscchembio.5b00942>

However, its activity seems to be very weak and other groups have not been able to reproduce the demethylase activity (Raijmakers *et al.*, 2007). Moreover, it was shown that di-methylated arginine residues cannot be demethylated by the enzyme due to steric occlusion of the active site (Holbert and Marmorstein, 2005). The catalysed reaction does not reverse methylation but instead neutral citrulline is produced, thus its role as a specific and relevant arginine demethylase is questionable (Wang *et al.*, 2004, Zhang *et al.*, 2019).

JMJD6 a member of the superfamily of non-haem Fe(II) and 2-oxoglutarate (2OG)-dependent oxygenases was the first reported enzyme capable of direct demethylation of symmetrical and asymmetrical di-methylated arginine residues on human H3 and H4 (Chang *et al.*, 2007). The proposed reaction mechanism is shown in Figure 1-18.

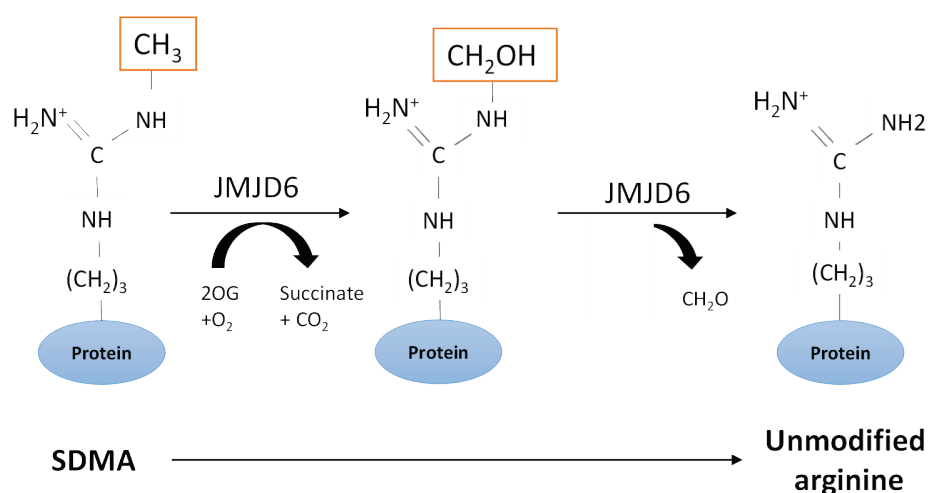


Figure 1-18 Demethylation Reaction by JMJD6.

*Mono-methylated arginine residues are demethylated by JMJD6 in a 2-step reaction. First, the methyl group is hydroxylated with the help of oxoglutarate (2OG). Second, a JMJD6-mediated deformylation produces formaldehyde (CH₂O) and the demethylated arginine residue (Poulard *et al.*, 2016).*

Different groups were able to show that JMJD6 can not only demethylate PRMT1 added methylation marks (Chang *et al.*, 2007, Tikhanovich *et al.*, 2015) but also a methylated arginine residue of ER α (Poulard *et al.*, 2014) and heat-shock 70 kDa protein (HSP70) *in vitro* (Gao *et al.*, 2015). However, even if the demethylase activity of JMJD6 was confirmed, it mainly functions as a lysyl-hydroxylase (Boeckel *et al.*, 2011) and it was shown that it regulates RNA splicing by catalysing the lysyl-hydroxylation of the splicing factor U2 small nuclear ribonucleoprotein auxiliary factor 65-

kD subunit (U2AF65) (Webby *et al.*, 2009). Thus, its role as demethylase is still under debate and more research is needed (Böttger *et al.*, 2015).

Recently, it has been shown that other JmjC-containing proteins of the KDM family that are histone lysine demethylases, are able to demethylate arginine residues *in vitro*, these include KDM4E, KDM3A, KDM5C, KDM6B, KDM2A, and KDM7B0 (Walport *et al.*, 2016, Klose *et al.*, 2007). Furthermore, the lysine demethylase JMJD1B was reported to demethylate mono and di-methylated arginine residues of histone H4 (Li *et al.*, 2018). Further research is needed to confirm their role as arginine-specific demethylases *in vivo*.

1.11 Regulation of PRMT Activity

PRMT activity is regulated by a number of mechanisms that include protein association, post-translational modification, miRNA regulation, subcellular localisation, and ubiquitin-dependent protein degradation. Additionally, alternative splicing generates PRMT isoforms that possess different substrate specificities and subcellular locations. Moreover, PRMT activity can be regulated by modifying or masking of substrate arginines (Guccione and Richard, 2019). The different mechanism will be discussed in more detail below.

1.11.1 Post-translational Modifications

CARM1 and PRMT5 were shown to be regulated by phosphorylation (Guccione and Richard, 2019). Phosphorylation of two conserved serine residues, Ser217 and Ser228 (human numbering), during mitosis caused the transcriptional co-activator CARM1 to be re-located from the nucleus to the cytoplasm. Enzyme activity was also abolished by preventing either binding of the cofactor SAM (Higashimoto *et al.*, 2007) or protein dimerisation, which is essential for enzyme activity (Feng *et al.*, 2009). The kinase responsible for the modification has still not been identified, but it is proposed that phosphorylation of both serine residues allows control of CARM1 activity during different phases of the cell cycle, by inactivating it during mitosis to prevent off-target methylation and activating it in G1 phase (Feng *et al.*, 2009). Phosphorylation of Ser448 of CARM1 by Protein Kinase A (PKA) facilitates CARM1 binding to estrogen receptor α which then leads receptor activation by cAMP (Carascossa *et al.*, 2010). Additionally,

p38 γ MAPK was shown to phosphorylate CARM1 on Ser572 and inhibit its transport to the nucleus and interaction with Pax7, and thus myogenesis (Chang *et al.*, 2018).

Ubiquitin-dependent degradation has been observed for PRMT1, CARM1 and PRMT5. PRMT5 is ubiquitinated by the E3 ubiquitin-ligase CHIP, and PRMT1 by the E3 ubiquitin-ligase E4B (Kim *et al.*, 2014, Lai *et al.*, 2017, Zhang *et al.*, 2016). CARM1 is protected from ubiquitin-dependent degradation by phosphorylation at Thr132 by Glycogen synthase kinase-3 (GSK3) (Li *et al.*, 2017). However, under nutrient-rich conditions it is ubiquitinated by SKP2-containing SCF (SKP1-cullin1-F-box protein) E3 ubiquitin-ligase in the nucleus and thus degraded (Shin *et al.*, 2016).

The protein substrate specificity of CARM1 might be further regulated by O-GlcNAcylation at the C-terminus of CARM1. This modification is proposed to alter CARM1-protein interactions, but further research is needed to identify the CARM1 substrates and their regulation (Charoensuksai *et al.*, 2015, Cheung *et al.*, 2008). Additionally, PRMTs can be regulated by other PRMT family members. For example, CARM1 was shown to methylate an arginine residue of PRMT5 which leads to a decrease in its methyltransferase activity (Chi *et al.*, 2010, Nie *et al.*, 2018). In the case of PRMT1, binding by PRMT2 increases its methyltransferase activity (Pak *et al.*, 2011). Moreover, in the case of PRMT6 (Singhroy *et al.*, 2013), PRMT7 (Geng *et al.*, 2017), PRMT8 (Dillon *et al.*, 2013) and CARM1 (Kuhn *et al.*, 2010), enzyme activity is regulated by auto-methylation. In the case of CARM1, auto di-methylation of Arg551 does not influence enzyme activity but prevents CARM1-activated transcription and pre-mRNA splicing (Kuhn *et al.*, 2010).

1.11.2 PRMT-Binding Partners

PRMTs including PRMT1 and PRMT5 can be regulated by cofactor recruitment. For example, the CCR4-associated factor 1 (CAF1) interacts with PRMT1 and regulates its activity in a substrate-specific manner (Robin-Lespinnasse *et al.*, 2007). Many PRMT-binding proteins have been identified that can change the enzyme activity by inhibition, activation, or influencing PRMT substrate specificity (Guccione and Richard, 2019). CARM1 interacts with at least 10 proteins to form the nucleosomal methylation activator complex (NUMAC) (Xu *et al.*, 2004). If CARM1 is bound to the complex it can methylate nucleosomal histone H3, whereas uncomplexed CARM1 can only methylate

free core histone H3. Additionally, CARM1 mRNA can be negatively regulated by different microRNAs (Lu *et al.*, 2018, Vu *et al.*, 2013, Xu *et al.*, 2013).

1.11.3 Alternative Splicing of PRMTs

Beside different cellular localisation patterns, each PRMT family member has alternatively spliced variant isoforms (Esse *et al.*, 2012). In the case of PRMT1, seven isoforms can be produced by alternating splicing at the N-terminal region of the pre-mRNA (Figure 1-19), (Baldwin *et al.*, 2012, Goulet *et al.*, 2007).

Of all isoforms, PRMT1v1 is the most abundantly expressed form, followed by PRMT1v2 and PRMT1v3 (Scorilas *et al.*, 2000). All of them with the exception of PRMT1v7 possess enzyme activity and have unique N-terminal sequences which influence substrate specificity. The N-terminal domain is thought to be important for protein-protein interactions with the PRMT substrate, thus alterations of the N-terminal sequence might be the reason for the observed altered substrate specificity (Goulet *et al.*, 2007). In contrast to the other isoforms, PRMT1v7 has no α Y helix and the invariant YFxxY motif is missing and thus is most likely not able to bind SAM, which would explain its inactivity (Goulet *et al.*, 2007). PRMT1 mainly occurs in the cytoplasm but also in the nucleus and is a highly mobile protein in both cell compartments as shown by photobleaching experiments (Goulet *et al.*, 2007). In the case of PRMT1v2 it was found that it carries a functional CRM1-dependent nuclear export signal (NES), which leads to its transport from the nucleus to the cytoplasm (Goulet *et al.*, 2007). It was also shown that PRMT1v2 can alternate between nucleus and cytoplasm and that the inactive form accumulates in the nucleus (Herrmann *et al.*, 2005, Herrmann and Fackelmayer, 2009). PRMT1v1 and PRMT1v7 are mostly found in the nucleus whereas the other isoforms are present in cytoplasm and nucleus. Additionally, the isoform mRNA expression profiles differ significantly across different human tissues. For example, PRMT1v4 is only expressed in heart tissue, whereas PRMT1v1 expression was found in spleen, kidney, lung, liver and skeletal muscle. PRMT1v6 was not found in any normal human tissue but in different breast cancer cell lines (Goulet *et al.*, 2007). PRMT1v3 and PRMT1v4 showed a reduced enzyme activity compared to wildtype protein (Goulet *et al.*, 2007). Some of the isoforms such as PRMT1v1 and PRMT1v3 have also been shown to be overexpressed in different breast cancer cell lines. Expression of PRMT1v2 was elevated the most compared to PRMT1v1 and was

also shown to promote the survival and invasiveness of breast cancer cells (Baldwin *et al.*, 2012, Goulet *et al.*, 2007).

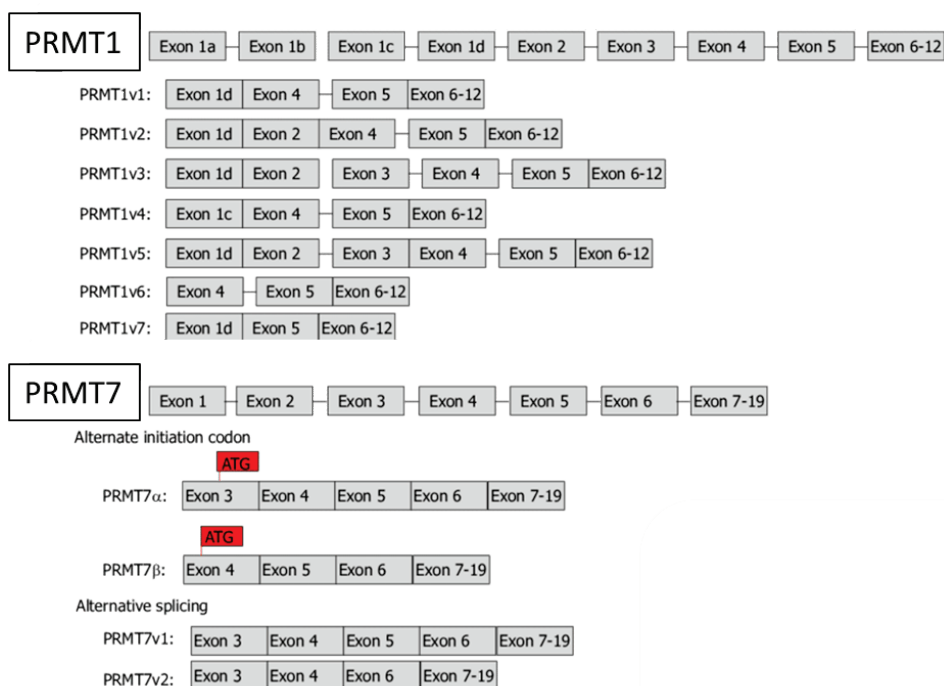


Figure 1-19 Overview of PRMT1 and PRMT7 Isoforms.

Seven variant isoforms were identified for PRMT1. The two PRMT7 isoforms α and β were identified in Chinese hamster cells, in humans one with high sequence similarity to PRMT7 α exists and two others named PRMT7v1 and v2. Figure adapted from (Baldwin *et al.*, 2014).

Two PRMT7 isoforms α and β are expressed in Chinese hamster cells, but in human only one isoform with high sequence identity to PRMT7 α exists (Gros *et al.*, 2006). The existence of two other isoforms (called v1 and v2 for human PRMT7) were predicted to differ in the in-frame deletion of exon 5. Alternative splicing has also been observed for PRMT2 (Zhong *et al.*, 2012), CARM1 (Shlensky *et al.*, 2015) and PRMT7 (Gros *et al.*, 2006). The isoforms of CARM1 and PRMT2 will be discussed in detail in Chapter 5 and 6.

1.11.4 Modification of Substrate Arginines

As previously discussed, PAD enzymes are capable of citrullination, the conversion of peptidyl-arginine residues into peptidyl-citrulline. Arg2, Arg8 and Arg17 of histone H3, and Arg3 of histone H4 are target sites of CARM1 and PRMT1, and their deimination

by PAD4 prevents arginine methylation (Wang *et al.*, 2004, Cuthbert *et al.*, 2004). On the other hand, di-methylation of arginine residues prevents deimination by PAD4 (Hidaka *et al.*, 2005). Taken together, deimination of unmethylated histone arginines by PAD4 might antagonize arginine methylation by PRMTs (Cuthbert *et al.*, 2004).

1.12 Development of PRMT Inhibitors

Genetic alterations of PRMTs, mostly upregulation, have been reported to drive cancer tumorigenesis and to play a role in inflammatory and neurodegenerative diseases, recently reviewed by Jarrold and Davies (2019). Thus, different PRMTs have been the target for drug development over the past 16 years. However, identification of selective inhibitors has been challenging due to the high sequence conservation of the catalytic domain especially in the SAM-binding pocket. This section will give an overview of inhibitor development against different Type 1 PRMTs.

1.12.1 Non-selective PRMT Inhibitors

The first PRMT inhibitor to be identified was SAH that is produced during the methyltransferase reaction by the demethylation of SAM. Compounds that inhibit SAH hydrolyase activity such as adenosine dialdehyde also inhibit PRMT activity by increasing cellular SAH levels. Additional analogues of the cofactor SAM were shown to inhibit PRMT activity, including Sinefungin (SNF) and Methylthioadenosine (MTA) (Cheng *et al.*, 2004). However, these compounds did not show any selectivity for the PRMT family and also inhibited other SAM-dependent enzymes. A structural overview of the early non-selective PRMT inhibitors is shown in Figure 1-20.

The first substrate competitive inhibitor that inhibited PRMT but not PKMT activity was identified in 2004 in an antibody-based fragment library screen and was named arginine methylation inhibitor 1 (AMI-1) (Cheng *et al.*, 2004). However, the compound had poor cell-permeability, no selectivity for a specific PRMT family member and is most likely not orally bioavailable due to its bisanionic structure. Based upon AMI-1, different screens were performed, and analogues designed. However, none of the identified compounds showed increased affinity compared to AMI-1 (Mai *et al.*, 2008, Mai *et al.*, 2007). Later, Feng *et al.* showed that AMI-I and related inhibitors did not directly target PRMTs but their histone substrates and prevent enzyme binding

(Feng *et al.*, 2010). Virtual screening using the published rat PRMT1 structure combined with biochemical assays was used to identify more drug-like PRMT inhibitors with affinities in the 50-100 μM range and with good cell permeability (Spannhoff *et al.*, 2007b, Spannhoff *et al.*, 2007a): RM65 competes for the substrate and cofactor binding pocket, whereas Stilbamidine and Allantodapsone only target the substrate binding site. Many optimisation rounds resulted in compound 9, which possess an IC_{50} of 1.5 μM against human PRMT1 (Bissinger *et al.*, 2011).

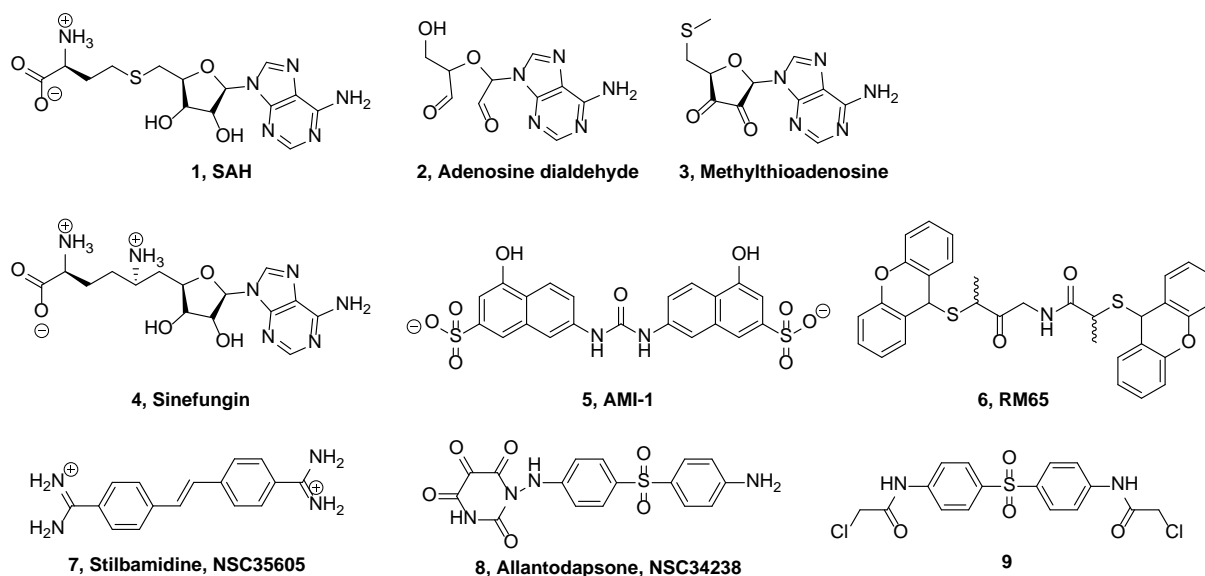


Figure 1-20 Early non-selective direct and indirect PRMT Inhibitors.

1.12.2 PRMT6 and CARM1 Selective Inhibitors

In the case of PRMT6 and CARM1, a number of pyrazole and imidazole derivatives were found to be potent inhibitors (Purandare *et al.*, 2008). An overview of reported selective inhibitors is presented in Figure 1-21. Intensive optimisation studies lead to the development of high affinity and selective indole inhibitors of PRMT6, named CMPD-1 and EPZ020411, and the pyrazole inhibitor of CARM1, CMPPD-2 (Sack *et al.*, 2011). They all showed IC_{50} values in the nanomolar range *in vitro* and co-crystal structures showed that they bind to the substrate binding pocket. As already discussed, the SAM cofactor must be present to promote the structural changes required to form the peptide binding pocket (Sack *et al.*, 2011). Based upon the two PRMT6 inhibitors CMPD-1 and EPZ020411, the first potent and cell-active inhibitor of Type 1 PRMTs, MS023, that is not active on Type 2 or Type 3 PRMTs, PKMTs or other DNA Methyltransferases, was

developed (Eram *et al.*, 2016). The MS023 IC₅₀s against Type 1 PRMTs varied with values of 30 nM for PRMT1, 119 nM for PRMT3, 83 nM for CARM1, and 4 and 5 nM in the case of PRMT6 and PRMT8 reported (Eram *et al.*, 2016). Subsequently, MS049 was discovered as a potent, selective, and cell-active dual inhibitor of CARM1 and PRMT6 with IC₅₀ values of 35-40 μM (Shen *et al.*, 2016).

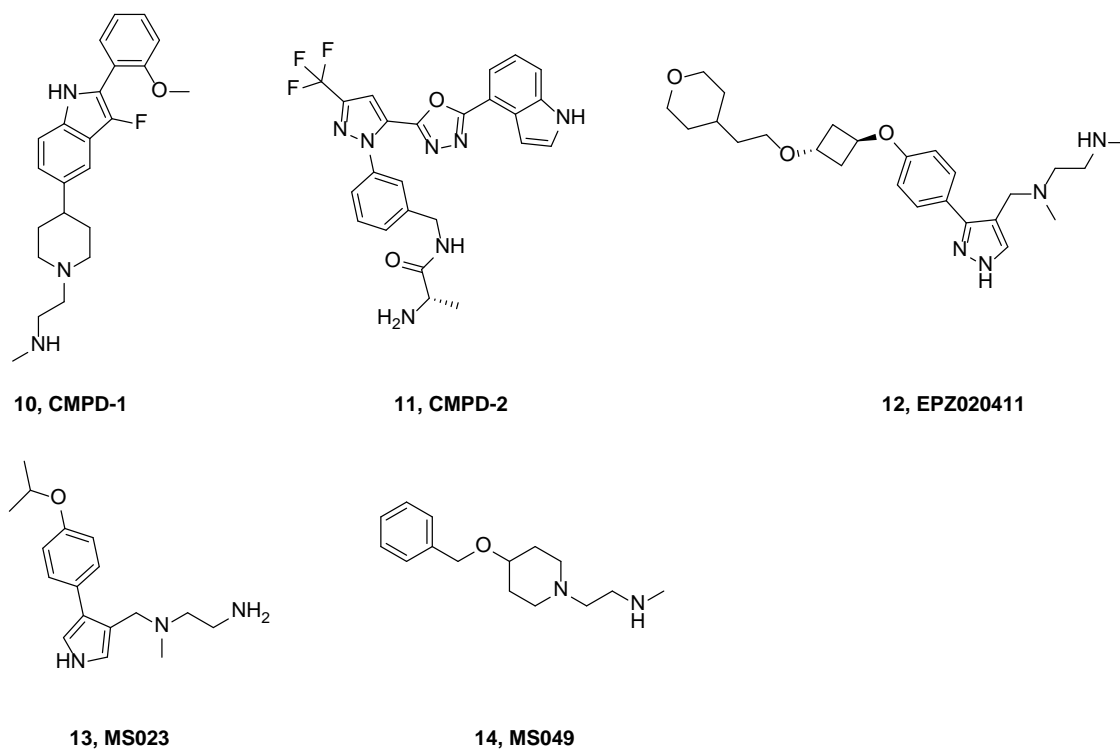


Figure 1-21 Early Inhibitors of CARM1 and PRMT6.

Two selective CARM1 inhibitors, TP-064 (Nakayama *et al.*, 2018) and EZM2302 (Drew *et al.*, 2017), have been identified with high *in vitro* and *in vivo* potency. Their chemical structures are shown in Figure 1-22.

EZM2302 is orally bioavailable, has an *in vitro* IC₅₀ value of 6 nM, and shows in cell activity in preclinical models of multiple myeloma. EZM2302 is based on the compound EPZ025654 that showed high *in vitro* activity with an IC₅₀ of 3 nM but its pharmacokinetic profile was not satisfactory and thus it was not tested *in vivo* (Drew *et al.*, 2017). TP-064 is a small molecule inhibitor of CARM1 that targets the substrate binding site and has an IC₅₀ value of <10 nM. As with other CARM1 inhibitors that target the peptide pocket, inhibition is SAM-non-competitive (Nakayama *et al.*, 2018).

Recently, a new selective chemical probe of CARM1 was identified named SKI-73 (Cai *et al.*, 2019). SKI-73 enters the cell as a prodrug and is then processed

into the active drug SKI-72. Crystallographic studies of a CARM1-SKI-72 complex revealed that it simultaneously binds the SAM and peptide binding sites. In contrast to the two other selective CARM1 inhibitors, SKI-72 binds to the active site is a SAM-competitive, substrate-non-competitive CARM1 inhibitors. in a SAM-competitive manner with an IC_{50} of 43 nM. SKI-72 cell activity was also confirmed. The compound inhibited CARM1-mediated arginine methylation of the substrates BRG1-associated factor 155 (BAF155) and poly(A)-binding protein 1 (PABP1). Moreover, treatment with $\geq 10 \mu\text{M}$ SKI-73 suppressed the invasion capability of MDA-MB-231 in a matrigel invasion assay by 80% but showed no antiproliferation activities. Moreover, LC-MS/MS quantification studies showed that SKI-72 achieves high intracellular concentrations due to its poor membrane permeability and that it remains stable over several days. A single low dose of the pro-drug SKI-73 (2.5-10 μM) was sufficient to occupy 95 % of the SAM binding sites with the active drug and maintain it for at least for 48 h (Cai *et al.*, 2019).

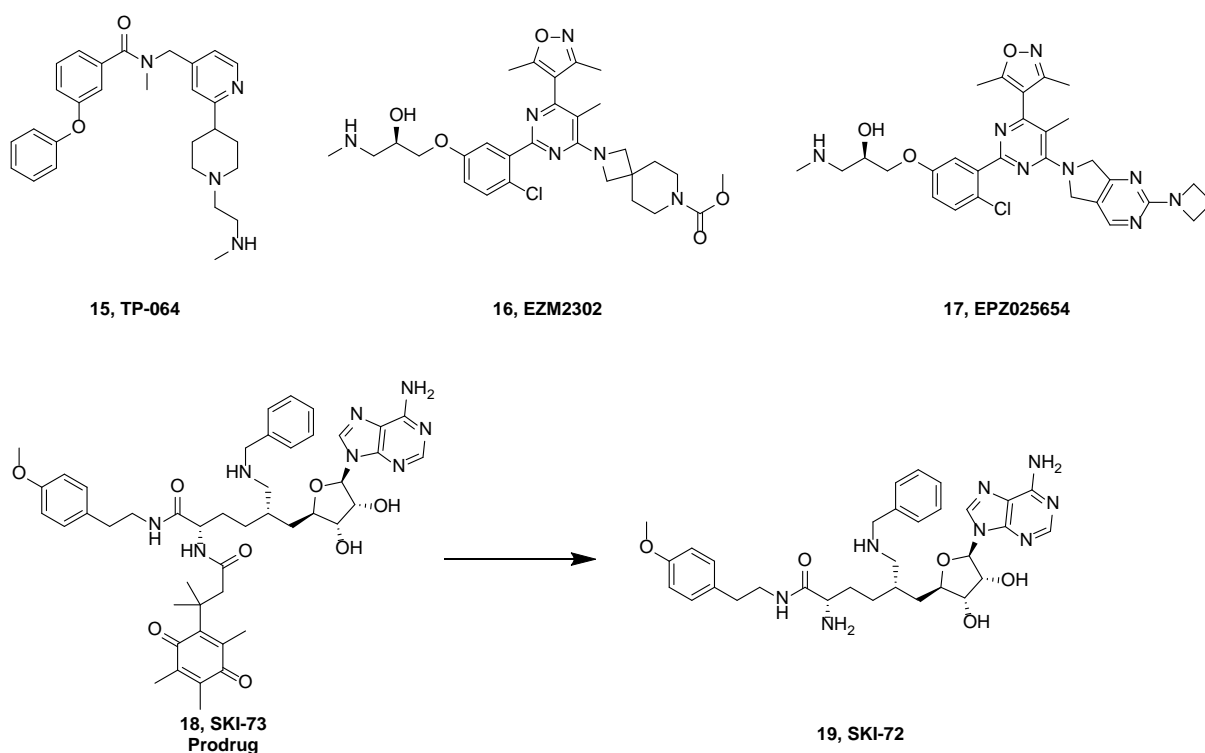


Figure 1-22 Selective CARM1 Inhibitors.

1.12.3 Inhibitors of PRMT1 and PRMT5

A number of active and high affinity compounds that target PRMT1 have been reported with one inhibitor, GSK3368715, being in Phase I clinical trials (Fedoriw *et al.*, 2019). Bisubstrate inhibitors such as the SAM-based inhibitors 20 and 21 have been developed for PRMT1, as has ligand 22 for PRMT5 (Figure 1-23). However, the PRMT1 inhibitors 20 and 22 showed poor selectivity and cell permeability (Dowden *et al.*, 2010, van Haren *et al.*, 2015, Li *et al.*, 2019, Wu *et al.*, 2019).

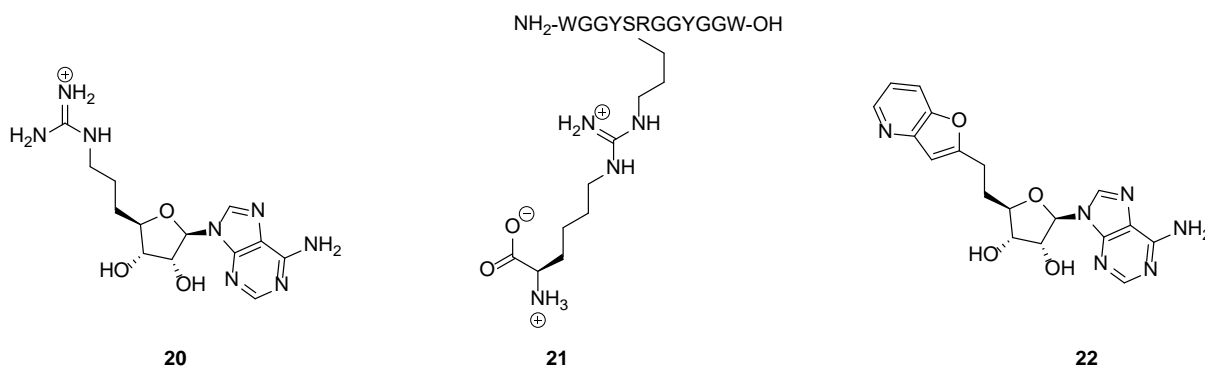


Figure 1-23 Bisubstrate Inhibitors of PRMT1 and PRMT5.

1.12.4 Selective Inhibitors of PRMT5 and PRMT7

Drug development against PRMT5 has been very successful. The first *in vivo*-active PRMT5 tool compound developed was EPZ015666 (Chan-Penebre *et al.*, 2015). The inhibitor is selective, potent, and orally bioavailable and targets the peptide binding site. Moreover, it reduced tumour growth in a leukemic mouse model (Chan-Penebre *et al.*, 2015). Optimisation of the ligand resulted in two compounds, EPZ015938 and GSK591 (Gerhart *et al.*, 2018), which are currently in Phase I clinical trials (Figure 1-24).

The PRMT5 inhibitor JNJ-64619178 is in clinical trials for treatment of non-Hodgkin lymphoma (Wu *et al.*, 2018b). The structures of two other PRMT5 inhibitors PF-06939999 and PRT543 that are in clinical trials are not disclosed.

Most PRMT inhibitors target the substrate arginine binding site. However, in the case of PRMT3, an allosteric inhibitor, SGC707, was reported to inhibit dimerisation

and thus enzyme activity by binding in an allosteric pocket located near the dimerisation arm (Siarheyeva *et al.*, 2012). Moreover, a few PRMT5 and PRMT7 inhibitors have been developed that target the SAM binding site (Figure 1-25), including the already mentioned inhibitors JNJ-64619178 and PF-06939999 (Figure 1-24). The selective and cell active PRMT7 inhibitor SGC8158 is formed *in vivo* from the prodrug SGC3027 via reduction. The compound has an *in vitro* IC₅₀ of <2.5 nM. Binding to the cofactor pocket has been confirmed by crystallography (Szewczyk *et al.*, 2019). The compound LLY-283 developed by Lilly also targets the cofactor binding pocket, and has nanomolar potency *in vitro* and *in vivo* and showed anti-tumor activity in mouse xenografts (Bonday *et al.*, 2018). These reports show that selective and cell-active inhibitors that compete with SAM for the cofactor binding site can be developed despite the high sequence conservation and competition with cellular SAM.

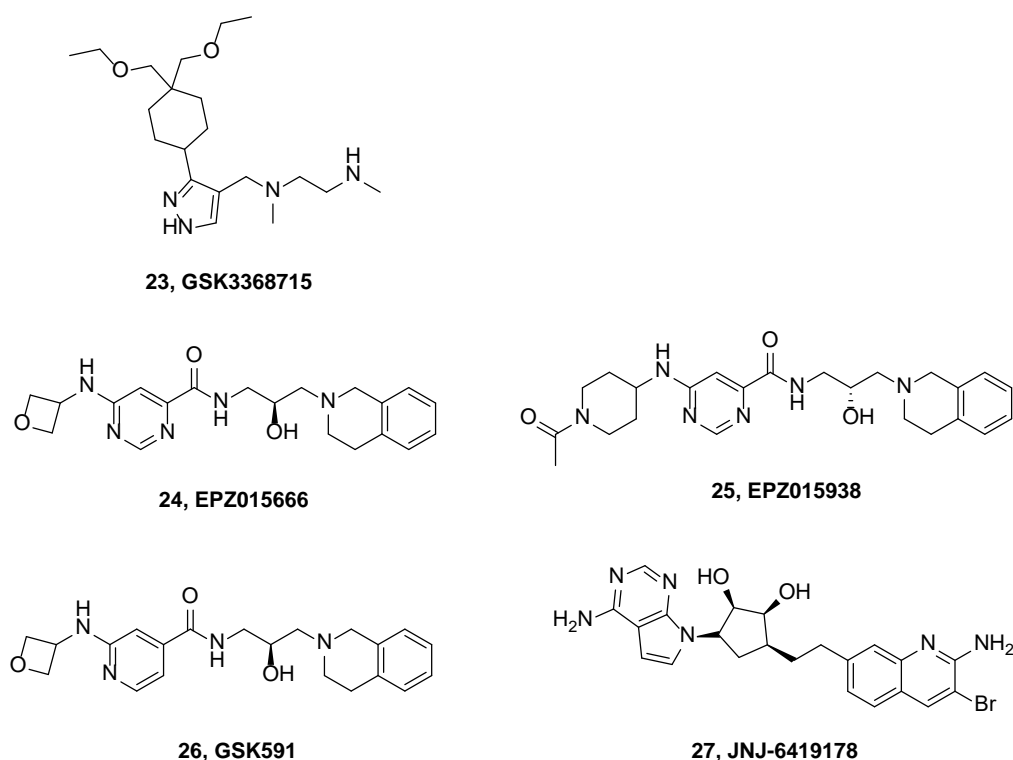


Figure 1-24 PRMT Type 1 and Type 2 Inhibitors in Clinical Trials.

In the case of PRMT2, Cura *et al.* (2017) solved the crystal structure of MmPRMT2 in complex with a known CARM1 inhibitor, Cp1. Cp1 is a SAM analogue that was designed to bind the SAM and peptide binding pocket of type 1 PRMTs (van

Haren *et al.*, 2015). The compound was able to reduce PRMT2 activity *in vitro*. However, it was a more potent CARM1 inhibitor (IC_{50} circa 460 nM) and with an IC_{50} of 16 μ M only slightly more potent an inhibitor of PRMT2 than SAH (IC_{50} of 18 μ M). Thus, so far, no selective PRMT2 inhibitor has been reported.

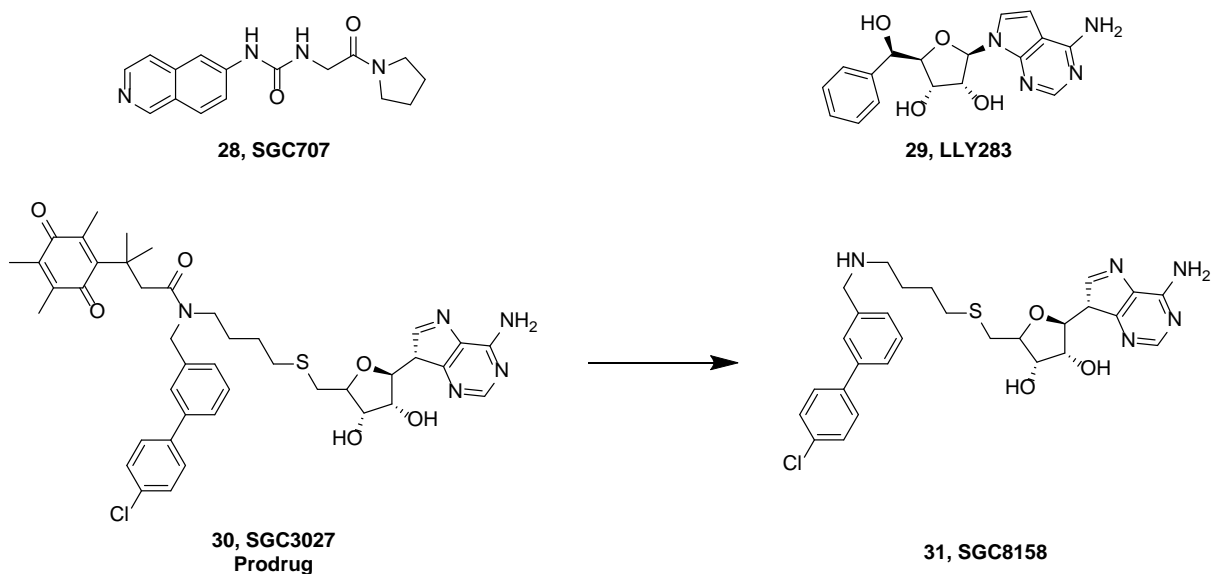


Figure 1-25 Selective PRMT Inhibitors that target the SAM-binding Site or allosteric Sites.

1.13 Research Aim

Alteration of PRMT expression, mostly upregulation, has been observed in various types of human cancer. However, it is still not clear whether PRMT2 has a pro- or anti-proliferative activity and further studies are needed to clarify the role of PRMT2 in cancer. The aim of this project was to further characterise the enzyme PRMT2 via biochemical and structural studies and to develop selective chemical probes for PRMT2, that could act as tools for target validation and provide potential start points for drug discovery. The active site is highly conserved across the PRMT family, which is why finding selective inhibitors is challenging. However, non-conserved residues within the binding-site might offer opportunities to design more selective compounds. In addition to PRMT2, other family members PRMT1 and CARM1, which are upregulated in different cancer types including breast and prostate cancer, were also produced with the aim to use them for selectivity screening of identified PRMT2 inhibitors. In order to perform these steps, the production of soluble and active protein was essential. Due to difficulties with producing sufficient amounts of PRMT2 protein, the second part of the project focussed on CARM1.

Biophysical assays were developed to allow the characterisation and optimisation of the binding of fragments to PRMT2 and CARM1. X-ray crystallography was used to characterise the binding modes of different fragments targeting the peptide binding site of CARM1 and to allow further ligand optimization. In Chapter 2 research methodology is described and explained. Chapter 3 describes the production of recombinant CARM1, PRMT2 and PRMT1 and their characterisation in preparation for biophysical and biochemical studies. The protein structures of PRMT2 that were obtained by X-ray crystallography are described in Chapter 4. Chapter 5 describes the development of different enzyme activity assays to measure and quantify enzyme activity and to compare the potency of different fragments. Finally, in Chapter 6, surface plasmon resonance assays with CARM1 were performed to quantitate fragment and ligand affinities. The binding data was combined with analysis of CARM1-ligand crystal structures to elaborate their mode of interaction at the atomic level.

Chapter 2 Materials and Methods

2.1 Chemicals and Reagents

The reagents and chemicals listed in the following section were all of high analytical standard and from Sigma Aldrich unless stated otherwise.

2.2 Expression Constructs

Different expression constructs were designed for recombinant expression in *Escherichia coli* (*E. coli*) or *Spodoptera frugiperda* (*Sf*) 9 or 21 cells. They are summarized in the following two tables.

Table 2-1 Summary of Protein Constructs for Insect Cell Expression.

Protein	Construct Name	Residue Range	N-terminal Tag ¹
RnPRMT1 UniProtKB: Q63009	RnPRMT1-1	M11-R353	His ₆ , His ₁₀ -MBP
HsCARM1 UniProtKB: Q86X55	HsCARM1-1	S135-S482	His ₆ , His ₁₀ -MBP
HsPRMT2 UniProtKB: P55345	HsPRMT2-1	M1-R433	His ₆ , His ₁₀ -MBP
	HsPRMT2-2	M1-T414	His ₆ , His ₁₀ -MBP
	HsPRMT2-3	G29-R433	His ₆ , His ₁₀ -MBP
	HsPRMT2-4	G29-R433	His ₆ , His ₁₀ -MBP
	HsPRMT2-5	V88-T414	His ₆ , His ₁₀ -MBP
	HsPRMT2-6	V88-R433	His ₆ , His ₁₀ -MBP
	HsPRMT2-7	G104-T414	His ₆ , His ₁₀ -MBP
	HsPRMT2-8	G104-R433	His ₆ , His ₁₀ -MBP
HSPRMT2_E223Q UniProtKB: P55345	HsPRMT2-9	M1-R433	His ₆ , His ₁₀ -MBP
MmPRMT2 UniProtKB: Q9R144	MmPRMT2-1	M1-S448	His ₆ , His ₁₀ -MBP, GST
MmPRMT2 UniProtKB: Q3UKX1	MmPRMT2-9	M1-R445	His ₆ , His ₁₀ -MBP, GST
MmPRMT2_R445W UniProtKB: Q3UKX1	MmPRMT2-10	M1-W445	His ₆ , His ₁₀ -MBP, GST

¹The vector backbone for all constructs is pACEBac1.

Table 2-2 Summary of Protein Constructs for *E. coli* Expression.

Protein	Construct Name	Residue Range	N-terminal Tag¹	Comments
RnPRMT1 UniProtKB: Q63009	RnPRMT1-1	M11-R353	His ₆ , GST	
	RnPRMT1-2	M11-R353	His ₆ , GST	Codon opt.
HsCARM1 UniProtKB: Q86X55	HsCARM1-1	S135-S482	His ₆ , GST	
	HsCARM1-2	S135-S482	His ₆ , GST	Codon opt.
MmPRMT2 UniProtKB: Q9R144	MmPRMT2-1	M1-S448	His ₆ , GST	Codon opt.
	MmPRMT2-2	M1-G424	His ₆ , GST	Codon opt.
	MmPRMT2-3	L42-S448	His ₆ , GST	Codon opt.
	MmPRMT2-4	L42-G242	His ₆ , GST	Codon opt.
	MmPRMT2-5	L100-G424	His ₆ , GST	Codon opt.
	MmPRMT2-6	L100-S448	His ₆ , GST	Codon opt.
	MmPRMT2-7	D116-G424	His ₆ , GST	Codon opt.
	MmPRMT2-8	D116-S448	His ₆ , GST	Codon opt.
HsPRMT2 UniProtKB: P55345	HsPRMT2-1	M1-R433	His ₆ , GST	
	HsPRMT2-2	M1-T414	His ₆ , GST	
	HsPRMT2-3	G29-R433	His ₆ , GST	
	HsPRMT2-4	G29-R433	His ₆ , GST	
	HsPRMT2-5	V88-T414	His ₆ , GST	
	HsPRMT2-6	V88-R433	His ₆ , GST	
	HsPRMT2-7	G104-T414	His ₆ , GST	
	HsPRMT2-8	G104-R433	His ₆ , GST	
	HsPRMT2-9	M1-R433	His ₆ , GST	Codon opt.
	HsPRMT2-10	M1-T414	His ₆ , GST	Codon opt.
	HsPRMT2-11	G29-R433	His ₆ , GST	Codon opt.
	HsPRMT2-12	G29-R433	His ₆ , GST	Codon opt.
	HsPRMT2-13	V88-T414	His ₆ , GST	Codon opt.
	HsPRMT2-14	V88-R433	His ₆ , GST	Codon opt.
	HsPRMT2-15	G104-T414	His ₆ , GST	Codon opt.
	HsPRMT2-16	G104-R433	His ₆ , GST	Codon opt.
SfRSF1	SfRSF1-1	M1-Y77	His ₆	Codon opt.

¹The vector backbone for all constructs is pET3dM.

2.3 *E. coli* Strains

Different *E. coli* strains were used for protein expression and cloning depending on their individual characteristics. The strains that were used and their properties are shown in Appendix B. The bacterial growth media was made up from pre-mixed, pre-buffered dry powder, which was dissolved in deionized water and sterilised by autoclaving before use. The different media and their compositions can be seen in Table 2-3, the selection antibiotics in Table 2-4.

Table 2-3 *E. coli* Growth Media.

Bacterial Growth Media¹	Formulation
Luria-Bertani Broth (LB) Base (Invitrogen)	10 g/L Tryptone, 5 g/L Yeast Extract, 5 g/L NaCl
Terrific Broth (TB) Medium (Melford)	12 g/L Casein digested Peptone, 24 g/L Yeast Extract, 9.4 g/L K ₂ HPO ₄ , 2.2 g/L KH ₂ PO ₄
Auto Induction 2xYT Medium (Foremedium)	16 g/L Tryptone, 10 g/L Yeast Extract, 6.8 g/L K ₂ HPO ₄ , 3.3 g/L (NH ₄) ₂ SO ₄ , 7.1 g/L Na ₂ HPO ₄ , 0.5 g/L Glucose, 2.0 g/L α -Lactose, 0.15 g/L MgSO ₄
LB Agar	25 g/L LB Broth Base (Invitrogen), 15 g/L Agar (Melford)
2xYT Medium	16 g/L Tryptone (Melford), 10 g/L Yeast Extract (Melford), 5 g/L NaCl (Sigma)
S.O.C Medium (Thermo Fisher)	0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 10 mM MgSO ₄ , 20 mM Glucose

¹The media were prepared in water and sterilised via autoclaving prior to use.

Table 2-4 Antibiotics and their used Working Concentration for Bacterial Selection.

Antibiotic¹	Formulation
Ampicillin	100 μ g/mL
Kanamycin	50 μ g/mL
Chloramphenicol	34 μ g/mL
Tetracycline	12.5 μ g/mL
Gentamycin	7.0 μ g/mL

¹The antibiotics were prepared at 100x stock concentration and stored at -20 °C or 4° C in the case of Gentamycin. They were thawed and diluted prior to use. Except for Chloramphenicol, which was prepared in ethanol, all others were dissolved in water.

2.4 Nucleic Acid Techniques

2.4.1 DNA Concentration Measurement

Nucleic acid quantifications were conducted using a NanoDrop 200 Spectrophotometer (ThermoFisher). The 260/280 nm absorbance ratio was used to determine the DNA purity. A ratio value of 1.8 is accepted as pure DNA, a ratio of 2.0 as pure RNA.

2.4.2 Primer Design

The 3' and 5' primers for the different constructs were designed in such a way that they could be used for InFusion and/or restriction (NcoI/Spel) cloning into any of the modified pET3dM vectors (Novogen) for *E. coli* protein expression or pACEBac1 vectors for

insect cell expression (Berger *et al.*, 2004). The sequence of the 5'-primer is TTCCAGGGGCCCATGGNN-gene-of-interest, where NN represents any base and is used to keep the sequence in frame; the NcoI restriction site is underlined. The sequence for the 3'-primer is GTTAGCAGCCACTAGT-STOP-gene-of interest; the SpeI restriction site is underlined. Table 2-5 lists the designed primers.

Table 2-5 Summary of the Designed Primers.

Primer Name	Primer Sequence (5'-3')¹
HsPRMT2_M1_F	TTCCAGGGG <u>CCC</u> ATGGCAACATCAGGTGAC
HsPRMT2_G29_F	TTCCAGGGG <u>CCC</u> ATGGGAGTACAGCCAGAGGAGTTTGTGG
HsPRMT2_V88_F	TTCCAGGGG <u>CCC</u> ATGGTGGATGAGTACGACCCCGAG
HsPRMT2_G104_F	TTCCAGGGG <u>CCC</u> ATGGGCAGCTATGGA ACTCTGAAACTC
HsPRMT2_O_M1_F	TTCCAGGGG <u>CCC</u> ATGGCAACCAGCGGTGATTG
HsPRMT2_O_G29_F	TTCCAGGGG <u>CCC</u> ATGGGTGTT CAGCCGGAAGAATTTGTTG
HsPRMT2_O_V88_F	TTCCAGGGG <u>CCC</u> ATGGTGGATGAATATGATCCGGAAG
HsPRMT2_O_G104_F	TTCCAGGGG <u>CCC</u> ATGGGTAGCTATGGCACCCCTGAAACTG
RnPRMT1_O_M11_F	TTCCAGGGG <u>CCC</u> ATGGAAGTTAGCTGTGTC
RnPRMT1_M11_F	TTCCAGGGG <u>CCC</u> ATGGAGGTTTCTGTGGC
MmPRMT2_M1_F	TTCCAGGGG <u>CCC</u> ATGGAAGCACCGGGTGAAGGTC
MmPRMT2_L42_F	TTCCAGGGG <u>CCC</u> ATGGAAGTGCAGCCGGAAGAATTTGTTG
MmPRMT2_L100_F	TTCCAGGGG <u>CCC</u> ATGGTGGAGAATATGATCCGGAAGATAC
MmPRMT2_D116_F	TTCCAGGGG <u>CCC</u> ATGGATAGCTATGGCACCCCTGAAACTG
HsCARM1_S135_F	TTCCAGGGG <u>CCC</u> ATGGGCTCTGTGTT CAGCGAGCGG
HsCARM1_O_S135_F	TTCCAGGGG <u>CCC</u> ATGGGCAGCGTGT TTAGCGAACGTAC
HsPRMT2_T412_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAAGTGACAGCCAGCTCAG
HsPRMT2_R433_R	GTTAGCAGCC <u>ACTAGT</u> TCACTATCTCCAGATGGGGAA-
HsPRMT2_O_T412_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAGGTA ACTGCCAGCTCAG
HsPRMT2_O_R433_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAGCGCCAATCGGAAAACTT
RnPRMT1_O_R353_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAACGCATACGATAATCGGTG
RnPRMT1_R353_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAGCGCATCCGGTAGTCGGTG
MmPRMT2_G424_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAACCCAGTTCAC-
MmPRMT2_S448_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAGCTATCACCACCGCTAC-
HsCARM1_S482_R	GTTAGCAGCC <u>ACTAGT</u> TCACTATGAGGGCGTTGTGCCCGTG
HsCARM1_O_S482_R	GTTAGCAGCC <u>ACTAGT</u> TCACTAACTCGGGGTTGTACCGG-
MmPRMT2_M1_F2	TTCCAGGGG <u>CCC</u> ATGGAGGCACCAGGAGAAGGTCCCTG
MmPRMT2_R445_R2	GTTAGCAGCC <u>ACTAGT</u> TCACCTCCAGATCGGAAAGAC
MmPRMT2_W445_R2	GTTAGCAGCC <u>ACTAGT</u> TCACCACCAGATCGGAAAGAC

¹The anneal sequence is shown in bold, NcoI/SpeI restriction sequences are underlined and the start/stop codons are shown in italics. The O indicated primers used for the codon optimised version of the gene.

2.4.3 PCR Amplification of DNA Inserts

Oligonucleotides were re-suspended in nuclease-free water to a final concentration of 100 pmol/ μ L and mixed by vortexing for 30 s. PCR reactions were set up in 1 \times Phusion GC Buffer (New England Biolabs) with 3 % (v/v) DMSO using sterile PCR tubes or 96-well PCR plates. The total reaction volume was 25 μ L, consisting of 200 μ M dNTPs, 1 unit/reaction Phusion HotStart DNA Polymerase (Invitrogen), 0.5 μ M forward and reverse primer, and a final template DNA concentration between 100-200 ng.

The PCR amplification of the target DNA was carried out using a PCR machine (Applied Biosystems 2720 Thermal Cycler, Thermo Fisher) using the cycle times shown in Table 2-6. The annealing temperature was calculated for each primer pair using the NEB T_m online calculator tool (<http://tmcalculator.neb.com/#!/main>). The extension time was calculated from the size of each amplicon (15 s/kb).

Table 2-6 PCR Reaction Cycles for DNA-Insert Amplification.

Initial De-naturation		Amplification Cycles						No. Cycles	Final Extension		Hold Temp
Temp	Time	Temp	Time	Temp	Time	Temp	Time		Temp	Time	
98 °C	30 s	98 °C	10 s	a °C	30 s	72 °C	b s	30	72 °C	10 s	4 °C

(a) Annealing temperature 45-72 °C dependent on primer T_m

(b) 15 s/kb extension time

2.4.4 Preparative Restriction Digest

Restriction digests were set up in sterile PCR tubes with a final volume of 50 μ L. The reaction composition is provided in Table 2-7. The reaction was incubated overnight at 37 °C.

Table 2-7 Preparative Restriction Digest Components.

Component	Final Concentration
10 \times CutSmart Buffer (New England Biolabs)	1 \times
Vector	50-100 ng/ μ L
NcoI-HF (20 U/ μ L) (New England Biolabs)	12-25 Units per Reaction
SpeI-HF (20 U/ μ L) (New England Biolabs)	12-25 Units per Reaction
Nuclease free water	to 50 μ L

2.4.5 InFusion Cloning

The NcoI/SpeI-digested vectors and the PCR-amplified inserts were PCR purified using either the QIAGEN PCR Purification Kit or the NucleoSpin PCR CleanUp Kit according to the manufacturer's instructions. The InFusion cloning reactions with a total volume of 5 μ L were set up in sterile PCR tubes as detailed in Table 2-8.

After mixing, the reaction was incubated at 50 °C for 15 min and then placed on ice. 1 μ L of each InFusion cloning reaction was transformed into 25 μ L Stellar competent cells (Clontech) according to standard procedures (Section 2.5), adding 250 μ L pre-warmed SOC broth (Clontech, 37 °C). 150 μ L of each reaction was spread on 15 mL LB/agar plates or 40 μ L onto 5 mL plates containing the appropriate antibiotic. All plates were incubated overnight at 37 °C. The next day, single colonies were used to inoculate 5 mL LB containing the appropriate antibiotics. The cultures were grown overnight, and plasmids were isolated using the QIAprep Spin Miniprep Kit according to the manufacturer's instructions.

Table 2-8 InFusion Reactions for Purified PCR Fragments.

Component	Final Concentration
5 x InFusion HD Enzyme Premix (Clontech)	1x
Linearized Vector	50-100 ng/reaction
Purified PCR Fragment	<0.5 kb: 5–25 ng/reaction 0.5 to 10 kb: 25–50 ng/reaction >10 kb: 25–100 ng/reaction
Nuclease Free Water	to 5 μ L
Total	5 μ L

In some cases, InFusion Cloning was performed using Cloning Enhancer Treatment (Clontech). Prior to setting up the InFusion cloning reactions 2 μ L Cloning Enhancer (Clontech) was added to 5 μ L of unpurified PCR reaction. The sample was incubated for 15 min at 37 °C, followed by incubation at 80 °C for 15 min using a PCR machine (Applied Biosystems 2720 Thermal Cycler, Thermo Fisher). Afterwards, the InFusion reactions were set up as described in the above section except that 1 μ L of PCR reaction was used (regardless of DNA concentration).

2.4.6 Ligation of PCR Amplicons into Expression Vectors

The PCR-purified PCR amplicons were digested overnight at 37 °C with NcoI/Spel in CutSmart buffer (New England Biolabs). The following morning, the digests were purified using the QIAGEN PCR Spin Purification kit according to the manufacturer's instructions. The ligation reactions were set up in sterile micro-centrifuge tubes. 50 ng of purified NcoI/Spel-cut vector was used in a total reaction volume of 20 µL in 1× T4 DNA ligase buffer (New England Biolabs) with a molar ratio of insert: vector of 3:1. After the addition of 1 µL of T4 DNA ligase (New England Biolabs), nuclease free water was added up to 20 µL. The ligation reaction was incubated at room temperature for 1 h. 2.5 µL of each ligation reaction, or sterile water (negative control), or 1 µL of 0.1 ng/µL pUC19 (positive control) were then used to transform 25 µL Stellar competent cells (Clontech) as described in Section 2.5. 125 µL of each transformation reaction were plated onto 15 mL LB/Agar plates with 100 µg/mL ampicillin. The plates were incubated overnight at 37 °C.

2.4.7 Site-directed Mutagenesis by PCR

Site-directed mutagenesis (SDM) was used to introduce point mutations to the HsPRMT2 plasmid. The primers that include the point mutation, are shown in Table 2-9. PCR amplification of the target DNA was carried out using a PCR machine (Applied Biosystems 2720 Thermal Cycler, Thermo Fisher) following the protocol described in Section 2.4.3 using the cycle times shown in Table 2-10. The remaining template DNA was digested by adding 0.5 uL Dpn1 (NEB) to the reaction and incubating for 1 h at 37 °C. After the sequence was verified by DNA sequencing (Section 2.6), the target DNA was cloned into pACEBac1_GST, pACEBac1_His₆ and pACEBac1_His₁₀_MBP vectors using Infusion cloning (Section 2.4.5). 1 µL of each reaction was then transformed into 25 µL *E. coli* Stellar cells (Section 2.5).

Table 2-9 Primers for Site-Directed Mutagenesis of HsPRMT2.

Primer Name	Primer Sequence (5'-3')
HsPRMT2_E223Q_F	CTGAGCTAGTACTTGACTTTGTCGTCCTCCAGG
HsPRMT2_E223Q_R	CCTGGACGGACGACAAAGTCAAGTACTAGCTCAG

F = Forward primer, R = Reverse primer.

Table 2-10 Reaction Cycles for SDM by PCR.

Initial De-naturation		Amplification Cycles						No. Cycles	Final Extension		Hold Temp
Temp	Time	Temp	Time	Temp	Time	Temp	Time		Temp	Time	
98 °C	30 s	98 °C	10 s	65 °C	30 s	72 °C	60 s	30	72 °C	10 s	4 °C

2.5 Heat-Shock Transformation of Competent *E. coli* Cells

25 µL of competent *E. coli* cells were thawed on ice for 20 min and then 1 µL of plasmid DNA was added to a sterile 1.5 mL microcentrifuge tube. The cells and DNA were incubated on ice for 30 min, followed by a heat shock for 45 s at 42 °C. Afterwards, the cells were transferred back on ice for 2 min to allow for recovery. 225 µL pre-warmed SOC broth (Clontech, 37 °C) were added to each tube and the reaction was shaken at 350 rpm in an Eppendorf Thermomixer for 1 h at 37 °C. 100 µL of each transformation reaction were plated onto 15 mL LB/Agar plates containing the appropriate antibiotics for selection. The plates were inverted and incubated at 37 °C overnight.

2.6 DNA Sequencing

For DNA sequencing, 5 mL LB overnight cell cultures with the appropriate antibiotics were prepared in duplicate which were inoculated with one independent clone picked from a transformation plate. The DNA was isolated using the Qiagen Miniprep Kit following the manufacturer's instructions. The sequence of every cloned insert was verified via DNA sequencing in both directions (Eurofins sequencing service). The DNA stocks were stored at -20°C.

2.7 Agarose Gel Electrophoresis

0.4 g high purity agarose (Sigma-Aldrich) were transferred to a 100 mL Duran bottle and 40 mL 1×TAE buffer (Biorad) and 4 µL SYBRSafe Dye (Thermo Fisher) were added. The bottle was heated in the microwave for 2-3 min until the agarose was completely dissolved. The agarose solution was allowed to cool for 5 min and poured into the gel tray. The gel was left on the bench until it was completely set. The gel was then

placed into a gel tank filled with 1×TAE buffer and the comb was carefully removed. 10 µL of each sample were mixed with 2 µL 6× loading dye (New England Biolabs) and the whole sample was loaded into the well. A molecular weight ladder was also loaded into a separate lane. The gel was run at 90 V for 30-45 min and DNA- bands were visualized using the BioRad Imager using the UV transillumination option.

2.8 Recombinant Protein Expression in *E. coli* Cells

For *E. coli* protein expression, one 100 mL Erlenmeyer flask with 25 mL of sterile medium containing the appropriate antibiotic was inoculated with 2-3 picked colonies from a transformation plate that was prepared the previous day and grown overnight at 37 °C and 220 rpm in an Infors HT Multitron shaking incubator (Infors). The next morning, each 5 L Erlenmeyer flask (Corning) containing 500 mL sterile medium and the appropriate antibiotic was inoculated with 10 mL pre-culture. Protein expression was induced at an OD₆₀₀ of approximately 0.4 or 0.6 depending on the tested construct by adding 0.05-1 mM IPTG and the cells were grown overnight at 220 rpm at 18 °C. The cells were harvested the following morning by centrifugation using an Avanti J-26 XP centrifuge with an JLA 8.1000 rotor (Beckman Coulter) at 6000 ×g for 30 min at 4 °C. The supernatant was discarded without disturbing the cell pellet, which was then transferred to a plastic bag or falcon tube and stored at -20 °C until purification.

2.9 Recombinant Protein Expression in Insect Cells

Recombinant protein expression was performed in *Sf9* and *Sf21* cells using the Multi-Bac Expression System (Geneva Biotech). The genes of interest were cloned into the multiple cloning sites (MCS) of pACEBac1 vectors using the standard cloning procedures described in Section 2.4.5. The pACEBac1 plasmid was transformed into DH10EMBacYFP cells (Geneva Biotech) as described in Section 2.9.2 and cells containing the plasmid were identified via Blue/White Screening (Section 2.9.2) and afterwards isolated (Section 2.9.4). The bacmid DNA was transfected into *Sf9/Sf21* cells to produce the V0 virus, as detailed in Section 2.9.5.

2.9.1 Maintenance of Sf9 and Sf21 Cell Lines

Sf9 and Sf21 insect cell lines were used for recombinant protein expression. Cells were grown in SF 900 II SMF medium (Lonza) at 27 °C in a non-humidified incubator while shaking at 150 rpm (Infors HT Multitron shaking incubator). 300 mL of suspension cultures were seeded at a cell density of 0.8×10^6 cells/mL in 1 L Erlenmeyer cell culture flasks (Sigma Aldrich) and passaged after 2-3 days when they reached the mid-log phase of growth (cell density $\sim 2-4 \times 10^6$ cells/mL).

2.9.2 Transformation and Blue/White Screening of pFastBac1 DNA into DH10EMBacYFP Cells

25 μ L of DH10EMBacYFP cells (Geneva Biotech) cells were thawed on ice for 10 min and 1 μ L of pFastBac1 DNA containing the DNA sequence of interest (10 ng/ μ L) was added. The mixture was incubated on ice for 30 min, followed by a heat shock for 45 s at 42 °C. Afterwards, the cells were transferred back on ice for 2 min to allow for recovery. 200 μ L pre-warmed S.O.C broth (Clontech, 37 °C) were added to each tube and the reaction was shaken at 200 rpm for 5-7 h at 37 °C in an Infors HT Multitron shaking incubator (Infors). For each transformation dilutions in SOC medium were prepared (1:40, 1:100, 1:500) and 100 μ L of each dilution was plated out onto pre-warmed X-gal/IPTG agar plates (Section 2.9.3). The plates were inverted and incubated at 37 °C for 48-72 h. Blue-white screening was used for identification of colonies carrying the recombinant bacmid. The bacmid of DH10EMBacYFP cells carries a gentamycin resistance gene and a gene encoding the LacZ α peptide, which complements the lacZ deletion mutation on the chromosome and allows the production of functional β -galactosidase. Cells with unmodified bacmid can produce functional β -galactosidase, hydrolyse X-gal and generate an insoluble blue pigment. The resulting colonies appear blue. Insertion of the mini-T7 transposon, which includes the target protein gene, into the bacmid DNA Tn7 attachment site, prevents the expression of the LacZ α peptide and functional β -galactosidase cannot be produced. The resulting colonies still have their gentamycin resistance but are white. In contrast, cells that did not take up the pFastBac1 plasmid are not gentamycin resistant and do not grow on the plates.

From each plate, two white clones were chosen and 2 mL of LB cultures for each supplemented with 50 μ g/mL Kanamycin, 10 μ g/mL Tetracycline, and 7 μ g/mL

Gentamycin were prepared in 15 mL falcon tubes and incubated at 37 °C and 160 rpm overnight.

2.9.3 Preparation of X-gal/IPTG Agar Plates for Blue/White Screening

LB agar (37g/L) plates were prepared with a final concentration of 50 µg/mL Kanamycin, 10 µg/mL Tetracycline, 7 µg/mL Gentamycin, 40 µg/mL IPTG, 60 µg/mL chromogenic substrate X-gal (5-Bromo-4-Chloro-3-Indolyl β-D- Galactopyranoside) to allow for blue-white screening.

2.9.4 Isolation of Bacmid DNA

Cells were harvested by centrifugation at 3400 g for 5 min at 4 °C. 300 µL P2 buffer (Qiagen) was added to labelled Eppendorf tubes. The supernatant of each cell pellet was removed, and the pellet was re-suspended in 300 µL P1 buffer (Qiagen) and transferred into the Eppendorf tube and mixed by inverting 2-3 times. After incubation at RT for a maximum of 5 min, 300 µL of pre-chilled N3 buffer (Qiagen) were added, and then slowly mixed by inverting. The solution was incubated for 10 min on ice and spun at 13 000 g for 10 min in a bench top centrifuge at 4 °C. In a fresh Eppendorf tube 800 µL of propan-2-ol were added and the supernatant was slowly added. The tube was inverted a few times, incubated for 10 min on ice and then spun at 13 000 g at 4 °C for 15 min. Afterwards the supernatant was removed without disturbing the DNA pellet. 1 mL of 70 % ethanol was added and then spun at 13 000 g at 4 °C for 15 min. Afterwards, as much ethanol as possible was removed and the pellet air dried for 2 min at RT. 40 µL of sterile water was added to the pellet without mixing and it was stored at -20 °C. The DNA pellet went into solution during thawing prior to transfection into Sf9/Sf21 cells.

2.9.5 Transfection of Bacmid DNA into Sf9/Sf21 Cells

Transfections were prepared in 24 deep well blocks with 5 mL volume per well and in duplicate. 100 µL of SF-900 medium (Lonza) and 50 µL of transfection reagent (Cellfectin II Reagent, Thermo Fisher) were placed in each well. 10 µL of bacmid DNA was added to each well, the wells were then covered with a breathable seal and incubated

for 30 min at RT in a laminar flow hood. In the last 10 min of the incubation time insect cells at 0.6×10^6 cells/mL were prepared. 5 mL of the cell suspension was then added to each well and the transfection mixture was incubated at 27 °C, 180 rpm for 7 days. The V0 virus was harvested by centrifugation of the plate at 3400 g at 4 °C for 10 min (Allegra 25R centrifuge with TA-10-250 Fixed-Angle Aluminium Rotor, Beckman Coulter) and the virus from duplicated wells was pooled into 15 mL falcon tubes. Successful transfection was evaluated using fluorescence microscopy (Nikon Eclipse TE2000 U) to analyse the expression of the yellow fluorescence protein (YFP) the gene for which is located on the bacmid. Fluorescence of the cells indicates that the transfection was successful. The virus was further amplified by addition of 1.5 mL V0 to 50 mL of cells at 0.6×10^6 cells/mL. The cells were grown at 27 °C, 180 rpm for 3 days and harvested as described above for V0. The V0 or V1 tubes were then wrapped in aluminium foil and stored at 4 °C. For large scale insect cell expression, a V2 virus preparation was freshly prepared for each expression as described in Section 2.15.

2.10 Small- Scale Expression Test in Sf9 Insect Cells

The small-scale expression screen was conducted in 24 deep well blocks in a laminar flow hood in a final volume of 3 mL. All screens were carried out in duplicate. First 180 µL of V1 virus preparation was added to each well. Then 3 mL of Sf9 cells at 3×10^6 cells/mL were added and the wells were sealed using sterile porous membrane lids (Sigma Aldrich). The deep well plate was incubated for 2 days at 27 °C and 170 rpm in a shaking incubator. The cells were harvested at 3400 g and 4 °C for 10 min and the supernatant discarded. The plates were sealed with a non-porous membrane and stored at -80 °C prior to expression analysis.

2.11 Small- Scale Expression Test in Sf21 Insect Cells

An expression screen of 11 different viral stocks (V1) produced in Sf9 cells was performed in Sf21 cells. 100, 200, or 400 µL of V1 were added to 3 mL Sf21 cells at 3×10^6 cells/mL in Sf900 II media (Thermo Fisher) in a 24 deep well block. The cells were incubated for 48 h at 27 °C and 140 rpm in a shaking incubator. The cells were harvested at 3400 g for 10 min at 4 °C and the supernatant discarded. The pellets were stored at -80 °C.

2.12 Small- Scale Expression Test in *E. coli* Cells

The small-scale expression screen was conducted in 24 deep well blocks in a final volume of 3 mL. All screens were carried out in duplicate. First overnight cultures were prepared by adding 3 mL of LB medium supplemented with the appropriate antibiotic to each well. The medium was inoculated with 3-4 picked colonies and the wells were sealed with a breathable membrane (Sigma Aldrich) and incubated overnight at 37 °C and 220 rpm in an Infors HT Multitron shaking incubator (Infors). The next day 3 mL of the expression medium was added to duplicated wells of a 24-well block and 30 µL of each appropriate overnight culture added. The wells were sealed and incubated at 37 °C and 220 rpm. In the case of AIM medium, they were incubated for 6 hours, after which the temperature was dropped to 18 °C overnight. Twenty-four hours after inoculation, the cells were harvested by spinning the plates at 4000 rpm (S5700 Swinging-Bucket Aluminium Rotor, Allegra 25R centrifuge, Beckman Coulter) for 10 min at 4 °C. The supernatant was discarded, and the pellets stored at -80 °C prior to expression analysis.

2.13 24-well Expression Screen Analysis using Spin Columns

The pellets were thawed in a warm water bath and resuspended in 750 µL of lysis buffer by pipetting up and down. The 24-well blocks were incubated for 30 min at 10 °C and 220 rpm in a shaking incubator and frozen at -80 °C. The 24-well blocks with the frozen lysate were then thawed in a warm water bath at 25 °C and mixed on a shaker at 180 rpm for 10 min. Cells were lysed by a freeze-thaw cycle at -80 °C and a total lysate sample of 30 µL for SDS-PAGE analysis was taken. Next, the 24-well blocks were centrifuged at 4 °C for 30 min at 3500 g and the soluble section carefully transferred to a new 24 well block. A 30 µL soluble fraction sample for SDS-PAGE was taken. 100 µL of 50 % (v/v) affinity resin slurry pre-equilibrated in binding buffer was added to each well, sealed and incubated at 10 °C at 200 rpm for 1 h in a shaking incubator. In the case of His-tagged protein expression, Ni Sepharose 6 Fast Flow (Sigma Aldrich) was used as affinity resin, for GST-tagged proteins Glutathione Sepharose 4B (GE Healthcare). The protein purification was conducted using 400 µL spin columns (Thermo Fisher) according to the manufacturer's instructions. In the last step the protein was eluted from the column with 250 µL of elution buffer. A 30 µL sample was taken for SDS-PAGE. To remove the tags, 29 µL of the elution fractions was mixed

with 1 μ L 3C protease to a final concentration of 0.2 mg/mL and incubated overnight at 4 °C. 10 μ L 4 \times loading buffer was added to all SDS-PAGE samples, and SDS-PAGE analysis was performed after heating up the samples to 95 °C for 3 min.

Table 2-11 Buffer Composition for Multiple Parallel Expression Screening.

Binding buffer His-tagged proteins	40 mM HEPES, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM TCEP
Binding buffer GST-tagged proteins	40 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM TCEP
Lysis buffer	Binding buffer + 1 mg/mL lysozyme (Sigma Aldrich) + 1 Complete Protease Inhibitor Tablet- EDTA Free (Roche)/50 mL+ 10 μ g/mL DNaseI + 50 μ g/mL RNaseA + 5 mM MgCl ₂
Elution buffer His-tagged proteins	Binding buffer + 500 mM imidazole pH 8.0
Elution buffer GST-tagged proteins	Binding buffer + 20 mM reduced glutathione

2.14 PhyNexus 24-well Expression Screen Purification

Cells were lysed by a freeze-thaw cycle. First the pellets were removed from the -80 °C freezer and 1 mL of lysis buffer was added to each well and they were incubated on a plate shaker at 1100 rpm for 30 min at RT. The cells were then frozen at -80 °C for 20 min, before thawing them in a warm water bath for 10 min at 25 °C. The blocks were spun at 3724 g for 30 min at 4 °C in a bench top centrifuge. 500 μ L of each lysate was then pipette in a well of a 96 well plate, without disturbing the pellet. A 30 μ L lysate sample was taken for later analysis for SDS-PAGE. The purification was then conducted using the fully automated protein purification system PhyNexus MEA 2 (Biotage). For His-tagged protein purification, 1000 μ L tips with 20 μ L of IMAC affinity resin per column were used (PhyNexus, #PTR 91-20-03). For GST-tagged proteins 1000 μ L tips with glutathione sepharose were used (PhyNexus, #PTR 91-20-03). 15 μ L of each elution sample was mixed with loading dye and analysed by SDS-PAGE.

Table 2-12 Buffer Composition for PhyNexus 24-well Purification.

Binding buffer His-tagged proteins	40 mM HEPES, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM TCEP
Binding buffer GST-tagged proteins	40 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM TCEP
Lysis buffer	Binding buffer + 1 mg/mL lysozyme (Sigma Aldrich) + Complete Protease Inhibitor Tablets- EDTA Free (Roche) + 2.5 units Benzonase/mL (Novogen)
Elution buffer His-tagged proteins	Binding buffer + 500 mM imidazole pH 8.0
Elution buffer GST-tagged proteins	Binding buffer + 20 mM reduced glutathione

2.15 Large-Scale Expression in Insect Cells

On the first day a new virus stock preparation was set up, named V2. 250 mL mid-log Sf9/Sf21 cells at 1.5×10^6 cells/mL were seeded with 1 mL of V1 virus in 1 L Erlenmeyer cell culture flasks (Sigma Aldrich). The cells were then incubated in a shaking incubator for 3 days at 27 °C and 150 rpm. At the same time, sufficient insect cells for large scale expression were prepared. The required volume of mid-log cells was scaled up in 1 L Erlenmeyer cell culture flasks with 500 mL cells per flask, by dilution with pre-warmed medium at 0.8×10^6 cells/mL. The cells were incubated for 3 days at 27 °C and 150 rpm. Cells were then counted and infected at $2.5\text{-}2.8 \times 10^6$ cells/mL by the addition of 25 mL short virus preparation (whole cell culture) to each flask. The cells were incubated at 27 °C at 170 rpm and harvested (3400 g for 15 min at 4 °C) after 48 or 72 hours. The cells were stored at -80 °C.

2.16 Protein Purification

All buffers and protein samples were filtered through 0.2/ 0.45 µm filters (Sigma Aldrich) prior to purification, and all purification steps were conducted at 4 °C on an ÄKTA Pure Purification system, unless stated otherwise.

2.16.1 Cell Lysis

The pellet from the cell culture was thawed on ice and gently re-suspended in 5 mL cold lysis buffer per gram of cell pellet. *E. coli* cells were disrupted by sonication on ice for a total time of 5-7 min (20 s “on”; 40 s “off”) at 30 % amplitude using a sonicator (Sonics Vibra-Cell VCX 500 / VCX 750). Insect cells were lysed by a thaw-freeze cycle at -80 °C or by sonication. The cell lysate was then centrifuged for 1 h at 48 500 (JA 25.50 Fixed-Angle Aluminum Rotor, Avanti J-26 XP centrifuge, Beckman Coulter) or 100 000 ×g (JA-30.50 Ti Fixed-Angle Titanium Rotor, Avanti JXN-30 centrifuge, Beckman Coulter) at 4 °C and the cleared lysate transferred to a clean Falcon tube and filtered through a 0.45 µm syringe filter (Millipore).

2.16.2 Expression and Purification of His₆-RSF1

His₆-RSF1 was expressed in *E. coli* BL21 (DE3) and BL21 (DE3) Star cells. A 50 mL pre-culture containing 50 µg/mL ampicillin was inoculated with one picked colony and grown overnight at 37 °C and 120 rpm. 1L of LB media supplemented with 50 µg/mL ampicillin was then inoculated with 10 mL of the pre-culture. Cells were grown at 37 °C and 120 rpm until the optical density at 600 nm reached 0.6. Cell expression was induced by addition of 0.4 mM IPTG. Cultures were harvested by centrifugation for 20 min at 4500 g at 4 °C using an Avanti J-26 XP centrifuge with JLA 8.1000 rotor (Beckman Coulter) after an incubation time of 3 h at 37 °C. The cells were resuspended in lysis buffer and lysed by sonication as described in Section 2.16.1. After centrifugation, the soluble fraction was applied to a gravity column containing 2 mL of Ni-NTA affinity resin pre-equilibrated in binding buffer. The column was then washed with 10 CV wash of binding buffer (20 mM Tris pH 8.0, 250 mM NaCl, 1 mM TCEP). The bound protein was eluted from the column by applying 5 × 1 mL elution buffer (binding buffer + 250 mM imidazole). Following SDS-PAGE, elution fractions containing the protein were combined and dialysed into dialysis buffer (20 mM Tris pH 8.0, 100 mM NaCl, 0.5 mM TCEP) overnight at 4 °C. Simultaneously, the His₆-tag was removed by the addition of His-3C-protease in a ratio of 1:50. On the next day, the protein solution was applied to a gravity column containing 1 mL of Ni-NTA affinity resin pre-equilibrated in dialysis buffer. The RSF1 containing flow-through was collected, followed by a 3 mL wash. Both flow-through and wash were combined, and the protein concentrated to 5 mL at *circa* 1 mg/mL. The sample was then applied onto a HiLoad Superdex 200

16/60 column (Sigma Aldrich) pre-equilibrated in SEC buffer (20 mM Tris pH 8.0, 500 mM NaCl, 0.5 mM TCEP) and run overnight at a flow rate of 0.5 mL/min at 4 °C. RSF1-containing fractions were identified by SDS-PAGE analysis, pooled, and concentrated to *circa* 0.5 mg/mL. RSF1 was then aliquoted and stored at -80 °C.

2.16.3 Expression and Purification of His₁₀-MBP-CARM1

His₁₀-MBP-CARM1 was expressed in *Sf9* or *Sf21* cells as described in Section 2.15. The pellets were thawed under cold water, re-suspended in lysis buffer, and lysed using a freeze-thaw cycle at -80 °C. The cells were mixed using a magnetic stirrer until homogenous. After centrifugation at 48 500 g for 1 h at 4 °C, 5 mL Talon resin was added to the supernatant and incubated overnight on a roller at 4 °C. The sample was then applied onto a gravity column. After a 10 CV wash with binding buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 % (v/v) glycerol and 1 mM TCEP), the protein was eluted with 3 CV of 100 % elution buffer (binding buffer + 500 mM imidazole). The His₁₀-MBP tag was removed by adding His-3C-protease in a ratio of 1:50 and incubating overnight at 4 °C. The protein was dialysed into dialysis buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol and 1 mM TCEP) and separated from the tags via subtractive Ni NTA IMAC (5 mL HisTrap HP column, Sigma Aldrich). Flow-through and wash were combined, concentrated to 7 mg/mL and applied onto a HiLoad 16/600 Superdex 75 pg column (Sigma Aldrich) that was pre-equilibrated in SEC buffer (40 mM HEPES pH 8.0, 600 mM NaCl, 5 % (v/v) glycerol and 0.5 mM TCEP). Protein containing fractions were identified by SDS-PAGE and concentrated to *circa* 7 mg/mL. The protein was stored in 40 µL aliquots at -80 °C.

2.16.4 Expression and Purification of GST-MmPRMT2-10

GST-MmPRMT2-10 was expressed in *Sf9* cells as described in Section 2.15. The pellets were re-suspended in lysis buffer and lysed by sonication. The lysate was clarified via centrifugation at 100 000 xg for 1 h at 4 °C. 2 mL of glutathione sepharose 4b resin was added to the soluble fraction, followed by a 2 h incubation at 4 °C. The sample was then applied to a gravity column and washed with 10 CV binding buffer (50 mM Tris pH 8.0, 250 mM NaCl, 5 % (v/v) glycerol, 10 mM DTT). In order to remove the GST-tag, 10 mL binding buffer with 100 µL GST-3C-protease (*circa* 1 mg/mL) was

added to the column and incubated overnight at 4 °C while rotating. The flow-through and a 5 CV wash were combined and concentrated to 5 mL at 2 mg/mL. The sample was then applied onto a HiLoad Superdex 200 16/60 column (Sigma Aldrich) pre-equilibrated in SEC buffer (20 mM Tris pH 8.0, 100 mM NaCl, 5 mM TCEP) at a flow rate of 0.5 mL/min. Protein-containing fractions were identified by SDS-PAGE, combined and diluted 1:1 with buffer (20 mM Tris pH 8.0, 5 mM TCEP) to obtain a final NaCl concentration of 50 mM. In the next step IEX chromatography was conducted using a HiTrap Q FF column (Sigma Aldrich) pre-equilibrated in IEX A buffer (20 mM Tris pH 8.0, 50 mM NaCl, 5 mM TCEP) at a flow rate of 3 mL/min. Following loading, proteins were eluted over an elution gradient of 0-50 % IEX buffer B (20 mM Tris pH 8.0, 1 M NaCl, 5 mM TCEP). MmPRMT2-10-containing fractions were combined and concentrated to 3.5 mg/mL, aliquoted and stored at -80 °C.

2.16.5 Expression and Purification of His₆-RnPRMT1-1

His₆-RnPRMT1 was expressed in *E. coli* BL21(DE3) cells in 2YT medium. A 50 mL pre-culture containing 50 µg/mL ampicillin was inoculated with one picked colony and grown overnight at 37 °C and 200 rpm. 1L of LB media supplemented with 50 µg/mL ampicillin was then inoculated with 6 mL of the pre-culture. Cells were grown at 37 °C and 180 rpm until the optical density at 600 nm reached 0.4. Cell expression was induced by addition of 0.4 mM IPTG. Cultures were harvested by centrifugation for 20 min at 4500 g at 4 °C after an overnight incubation at 20 °C. After lysis and sonication of a cell pellet from 0.9 L *E. coli* cells, the protein was purified using a 5 mL HisTrap HP column (Sigma Aldrich) pre-equilibrated in IMAC A buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 % (v/v) glycerol, 0.5 mM TECEP). After a 10 CV wash, the bound protein was eluted using a gradient of from 0-100 % IMAC B buffer over 10 CV (binding buffer supplemented with 500 mM imidazole). PRMT1-containing elution fractions were identified by SDS-PAGE and pooled. The His-tag was removed from the protein by incubation with His-3C-protease and the sample dialysed into 40 mM HEPES, pH 8.0, 600 mM NaCl, 5 % (v/v) glycerol and 0.5 mM TCEP. Tagged and cleaved protein were separated by subtractive IMAC on a 5 mL HisTrap HP column. Flow-through and wash were combined and concentrated to 3 mL and loaded onto a buffer Superose 6 10/300 GL column (Sigma Aldrich) pre-equilibrated in

dialysis at a flow rate of 1 mL/min. The protein was identified by SDS-PAGE and concentrated to 1 mg/mL before storing it at -80 °C.

2.16.6 Expression and Purification of HsPRMT2_E223Q

His₁₀-MBP- and His₆-tagged HsPRMT2_E223Q was expressed in 300 mL of *Sf9* cells as described in Section 2.16.6. The pellets were thawed under cold water, re-suspended in 50 mL lysis buffer, and lysed using a freeze-thaw cycle at -80 °C. The cells were mixed using a magnetic stirrer until homogenous. After centrifugation at 48 500 g for 1 h at 4 °C, 3 mL Talon resin was added to the supernatant and incubated for 20 min on a roller at 4 °C. The sample was then applied onto a gravity column. After a 10 CV wash with binding buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 % (v/v) glycerol and 1 mM TCEP), the protein was eluted with 100 % elution buffer (binding buffer + 500 mM imidazole). The eluted tagged protein was cleaved with His-3C-protease overnight and dialysed into 40 mM HEPES pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol and 1 mM TCEP, followed by subtractive TALON IMAC. Protein containing fractions were concentrated to 1 mg/mL and stored at -80 °C.

2.16.7 Expression and Purification of His₁₀-MBP-HsPRMT2-6

His₁₀-MBP-HsPRMT2-6 was expressed in 2.5 L of *Sf9* cells as described in Section 2.15. The pellets were thawed under cold water, re-suspended in lysis buffer, and lysed using via sonication. After centrifugation at 48 500 g for 1 h at 4 °C, 5 mL Talon resin was added to the supernatant and incubated for 30 min on a roller at 4 °C. The sample was then applied onto a gravity column. After a 10 CV wash with binding buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 % (v/v) glycerol and 1 mM TCEP), the protein was eluted with 3 CV of 50 % elution buffer (binding buffer + 500 mM imidazole). Elution fractions were combined and the His₁₀-MBP tag was removed by His-3C-protease cleavage (2.16.1) overnight at 4 °C. The protein was dialysed into dialysis buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol and 1 mM TCEP) and separated from the tags via subtractive Talon IMAC (5 mL gravity column). Flow-through and wash were combined, concentrated to 2 mL and applied onto a Hiload Superdex 200 16/60 (Sigma Aldrich) that was pre-equilibrated in SEC buffer (40 mM

HEPES pH 8.0, 600 mM NaCl, 5 % (v/v) glycerol and 0.5 mM TCEP). Protein containing fractions were identified by SDS-PAGE, combined, and diluted 1:6 with dilution buffer (40 mM HEPES pH 8.0, 5 % (v/v) glycerol and 1 mM TCEP) to a final NaCl concentration of 50 mM. The sample was then loaded onto a pre-equilibrated Mono Q 5/50 GL column at 1 mL/min. The protein was eluted with a NaCl gradient from 0-1 M over 30 CV at 0.5 mL/min. Protein containing fractions were identified by SDS-PAGE, concentrated to 3.6 mg/mL and stored in 40 μ L aliquots at -80 °C.

2.16.8 Expression and Purification of His₆-HsPRMT2-1

His₆-HsPRMT2-1 was expressed in 1.2 L of Sf9 cells for 48 h as described in Section 2.9. Cultures were harvested by centrifugation for 10 min at 3400 g at 4 °C. The cells were lysed by sonication and centrifuged as described in Section 2.16.1. The protein was purified using a 4 mL Talon column pre-equilibrated in IMAC A buffer (40 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 % (v/v) glycerol, 0.5 mM TECEP). The column was incubated for 15 min at 4 °C. First the flow-through was collected, then after a 10 CV wash, the bound protein was eluted using 3 \times 1 mL IMAC B buffer (binding buffer supplemented with 500 mM imidazole). HsPRMT2-containing elution fractions were identified by SDS-PAGE and pooled. The His-tag was removed from the protein by incubation with His-3C-protease and the sample dialysed into 40 mM HEPES, pH 8.0, 600 mM NaCl, 5 % (v/v) glycerol and 0.5 mM TCEP. Tagged and cleaved protein were separated by subtractive IMAC on a 2 mL Ni²⁺-column. Flow-through and wash were combined and concentrated to 0.5 mL and loaded onto a Superdex 200 ag 10/300 column (Sigma Aldrich) pre-equilibrated in dialysis buffer at a flow rate of 0.5 mL/min. The protein was identified by SDS-PAGE and concentrated to 5.5 mg/mL before storing it at -80 °C.

2.17 Protein Analysis

2.17.1 SDS- Polyacrylamide Gel Electrophoresis (PAGE)

The purity of the recombinant proteins was assessed by SDS-PAGE after every purification step. SDS-PAGE was carried out according to Laemmli (1970). 10 or 12 % (w/v) pre-made SDS-PAGE gels (Biorad) or Mini-PROTEAN TGX Stain-Free Precast Gels (Biorad) were used. Before the sample was loaded onto the gel, 4 \times sample loading

buffer (C.B.S. Scientific) containing 100 mM dithiothreitol (DTT) was added and the tubes were heated in a hot block for 5 min at 95 °C. Gels were run for ~45 min at 180 V until the dye front reached the gel bottom. All gels were run using a premixed 10× Tris/Glycine/SDS running buffer (Biorad) that was diluted prior to use.

Dilute protein samples were concentrated using StrataClean Resin (Agilent Technologies). Following resuspension by brief vortexing, 1 µL of slurry was added per 100 µL sample and vortexing repeated. The sample was then incubated at RT for 5 min and vortexed again and the resin pelleted (11 766 xg for 5 min). The supernatant was discarded, and the resin re-suspended in 10 µL 1× sample loading dye supplemented with reducing agent (100 mM DTT). Protein was eluted by heating to 95 °C for 10 min. The entire supernatant was then subject to SDS-PAGE (approximately 40 min at 180 V) and gels subsequently stained with Instant blue (C.B.S. Scientific) or analysed by western blotting (Section 2.17.2).

2.17.2 Western Blotting Analysis

GST-tagged proteins were identified by western blot analysis. Proteins were separated by SDS-PAGE, then transferred onto a PVDF membrane using a Wet/Tank Blotting System (Biorad) using transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol). Transfers were performed for 1 h at 100 V and ice packs were used for cooling to mitigate the produced heat. After the protein transfer, the membrane was washed briefly in TBS/T buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 1 % (v/v) TWEEN 20) and then blocked by incubation in 5 % (w/v) non-fat dry milk in TBS/T buffer for 1 h at RT. The membrane was washed in TBS/T buffer and then incubated with anti-GST-antibody (GST (91G1) Rabbit mAb #2625, Cell Signalling) was diluted 1:1000 in 5 % (w/v) bovine serum albumin (BSA) in TBS/T buffer. The membrane was incubated with the primary antibody overnight at 4 °C. The next morning, the membrane was washed 3× 15 min in TBS/T buffer. The secondary antibody (anti-rabbit IgG, HRP-linked antibody #7074, Cell Signalling) diluted 1:2000 in 5 % (w/v) BSA in 10 ml of TBS/T buffer and the membrane was incubated for 1 h at RT. Afterwards the membrane was washed four times with TBS/T for 15 min and the protein was detected using the Amersham ECL Prime Western Blotting Detection Reagent kit (GE Healthcare) according to the manufacturer's instructions. The chemo-luminescent protein western blots were visualised by X-ray film.

2.17.3 Protein Quantification

Protein concentrations were measured by absorbance measurements at 280 nm using a NanoDrop 200 spectrophotometer (ThermoFisher) according to Lambert-Beer's Law:

$$A = \epsilon \times b \times c \quad \text{Equation 3}$$

A= Absorbance value; ϵ = Wavelength-dependent molar extinction coefficient; b= Path length; c = Analyte concentration

Molar extinction coefficients at 280 nm were calculated using the ProtParam online tool on the ExPASy Server (<http://web.expasy.org/protparam>).

2.17.4 Intact Protein Analysis using Mass Spectrometry

The accurate mass of purified proteins was confirmed by liquid chromatography combined with mass spectrometry (LC-MS). Sample preparation and LC-MS experiments were performed by Dr Claire Jennings.

2.17.5 Limited Proteolysis

First the enzymes were prepared. Endoproteinase Asp-N from *Pseud fragi* (Roche 11054589001) was resuspended in water to 0.5 mg/mL. Chymotrypsin from *bovine* pancreas (Roche 11418475001) was resuspended to the same concentration in 0.1 mM HCL. HsPRMT2-1 at *circa* 1.16 mg/mL was mixed with each protease at two different protease to protein ratios, 1:100 and 1:400 and incubated at three different temperatures, 4 °C, RT and 37 °C. The final reaction volume was 20 μ L. 5 μ L samples were taken at specific time points, after 1 h, 2.5 h and 24 h. A control sample was prepared by mixing 10 μ L of HsPRMT2-1 with 2.5 μ L loading dye. Samples were immediately frozen in dry ice to inactivate the protein and stored at -80 °C until analysis. The samples were mixed with 15 μ L buffer and 5 μ L gel loading buffer, heated at 95 °C for 10 minutes and analysed by SDS-PAGE on a 10 % (w/v) gel. Significant bands that could represent protein fragments were identified and further analysed via MS analysis.

MS Analysis was performed by Dr Rachel Rowlinson (AstraZeneca, UK). The protein bands were reduced, alkylated, and digested with chymotrypsin or endoproteinase Lys-C from *Lyso enzymogenes* (Roche 1047825). The resultant digests were loaded onto a Dionex U3000 NanoLC for ESI LC MS analysis using the Qstar Elite mass spectrometer. The mass spectrometry data was searched against the provided sequences with the APUser database using the Mascot search engine.

2.17.6 Thermal Stability Assay using the Thermofluor-iCycler

For thermal stability screening a standard buffer screen from AstraZeneca was used, the composition can be found in Appendix C. The assay was performed in a 384-well plate format. At first the optimal protein concentration of PRMT2 and PRMT1 were determined by testing different reaction volumes ranging from 6-12 μL , protein concentrations from 1.1- 20 μM and different Sypro Orange dye (5000 \times concentrated stock solution in DMSO, Sigma Aldrich) concentrations (8 or 16 \times). The dye is excited at a wavelength of 470 nm and emits light at 570 nm. The final reaction had a volume of 6 μL with a final concentration of 8 \times SYPRO Orange dye and a final protein concentration of 3.5 μM RnPRMT1 or 2.1 μM in the case of HsPRMT2. The assay plate was incubated for 45 min at 4 $^{\circ}\text{C}$ before the assay was conducted using a Thermofluor-iCycler iQ Real Time Detection System (Bio-Rad). After an initial hold at 20 $^{\circ}\text{C}$, the temperature was increased up to 80 $^{\circ}\text{C}$ in 0.01 $^{\circ}\text{C}/\text{s}$ increments, and the SYPRO Orange fluorescence emission signal at 570 nm at the end of a 30 s hold at each temperature was read. Measurements were performed in duplicate and an additional no protein control was conducted (only buffer and dye). The average of the duplicates was taken, and the no protein control subtracted. The thermal melt temperature (T_m) was then calculated by plotting the first derivative of the fluorescence emission ($-\text{dF}/\text{dT}$) as a function of temperature using Excel.

2.17.7 Analytical SEC using the Ettan System

Analytical SEC of RnPRMT1-1 without and in the presence of added SAM or SAH was performed at RT using a Superdex 75 10/300 GL column (Amersham Biosciences, GE Healthcare) pre-equilibrated in binding buffer (40 mM HEPES, pH 8.0, 300 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol) on an Ettan chromatography system (Amersham

Biosciences, GE Healthcare). 20 μ L RnPRMT1-1 at *circa* 1.2 mg/mL were incubated with either 25 μ M SAH or SAM for 30 min on ice. Afterwards, the samples were spun for 10 min at 13 000 g before loading them onto the column at a flow rate of 0.04 ml/min. For preparation of a calibration curve, the molecular weight markers (Bio-rad), shown in Table 2-13, were freshly prepared according to the supplier's instructions and 10 μ L of each was run at a flow rate of 0.04 mL/min under the same conditions as mentioned above. The calibration curve was produced by plotting the logarithm of the molecular weight of the markers against their elution volume.

Table 2-13 Gel Filtration Standard Components and their Molecular Weight.

Molecular weight marker protein	Molecular weight (Da)
Thyroglobulin (bovine)	670,000
g-globulin (bovine)	158,000
Ovalbumin (chicken)	44,000
Myoglobin (horse)	17,000
Vitamin B12	1,350

2.18 *In vitro* Methyltransferase Glo Assay

The MTase Glo™ assay (Promega) was used to measure the enzyme activity of different PRMTs. The assay was performed as a two-step assay in a 384-well micro plate (Thermo Fisher) format. The total reaction volume was 4 μ L.

2.18.1 Determination of the Optimal Enzyme Concentration

An enzyme titration curve of different PRMTs was generated to determine the optimal PRMT amount to use in the MTase Glo™ Assay. First the 2x substrate mix was prepared with 2 or 20 μ M SAM (Promega) in reaction buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl₂, 0.1 mg/mL BSA, 1 mM DTT). The substrate was either full-length histone H3.1 (NEB, # M2503S) or histone H4 (NEB, # M2503S). Both were supplied in 20 mM sodium phosphate (pH 7.0), 300 mM NaCl, 1 mM EDTA and 1 mM DTT and were stored at -20 °C. The final assay substrate concentration was 1 or 5 μ M. Next, 100 μ L of the PRMT at the desired maximal concentration was prepared in reaction buffer and a two-fold serial dilution with 12 points prepared including a non-

enzyme control. 2 μL of the prepared 2 \times substrate mix was added to each well, followed by 2 μL of the prepared PRMT titration. The assay plate was centrifuged using an MPS 1000 Mini PCR Plate Spinner (Labnet International) for 30 s at 1000 rpm, shaken for 10 min on a microplate shaker at 200 rpm, and then incubated for 30 min at RT.

Table 2-14 Components of the 2 \times Substrate Reaction Mix.

Component	Volume	Concentration	Final Assay Conc.
Histone H3 (65 μM) Histone H4 (89 μM)	3.69/ 18.45 μL 2.69/ 13.45 μL	2 or 10 μM	1 or 5 μM
4 \times Reaction buffer	30 μL	1 \times	1 \times
1 mM SAM ¹ or 100 μM SAM	2.4 μL	20 μM or 2 μM	10 μM or 1 μM
Water	Up to 120 μL		

¹The 1 mM and 100 μM SAM stock was included in the MTase Glo Assay Kit and was aliquoted and stored at -20 $^{\circ}\text{C}$. The 4 \times reaction buffer was stored in aliquots at -20 $^{\circ}\text{C}$. BSA and DTT were added fresh prior to use.

During the incubation time the 5 \times MTase Glo reagent was freshly prepared by thawing the 10 \times MTase Glo reagent (Promega) on ice and mixing the required volume 1:1 with water. 1 μL of 5 \times Methyltransferase-Glo Reagent was then added, the plate centrifuged again for 30 s at 1000 rpm, mixed, and incubated for another 30 min at RT. Finally, 5 μL room-temperature MTase Glo Detection Solution (Promega) was added, centrifuged for 30 s at 1000 rpm, mixed, and incubated for 30 min at RT, before recording the luminescence with a plate reader luminometer. An enzyme activity curve was obtained by plotting the luminescence against the enzyme concentration using GraphPad Prism and Excel. The 10 \times MTase Glo Reagent was aliquoted and stored at -80 $^{\circ}\text{C}$, the MTase Detection Solution at -20 $^{\circ}\text{C}$.

Prior to performing the MTase assay, a SAH standard curve was freshly prepared to correlate the luminescence signal and SAH concentration. A two-fold serial dilution of SAH with 12 points ranging from 0-1 μM or 0-10 μM in reaction buffer was prepared. First, a 1 μM or 10 μM SAH solution with a final volume of 200 μL was prepared by mixing 50 μL of 4 \times reaction buffer, 13.3 μL of 15 μM SAH (Promega) and 136.7 μL water. 4 μL of the SAH titration were transferred to the assay plate at the start of the experiment (in duplicate). The SAH stock was stored at -20 $^{\circ}\text{C}$. The SAH standard curve was obtained by plotting the luminescence against the SAH concentration using GraphPad Prism.

2.18.2 Determination of K_m Values for PRMT Substrates

To obtain K_m values for the histone substrates, a histone titration was performed while keeping SAM at a constant concentration. The reaction volume was 4 μ L and reactions were performed at RT. First, a substrate dilution solution was prepared with 20 μ M SAM (Promega) in 1 \times reaction buffer. The enzyme was then diluted in reaction buffer to twice the assay concentration. Next, a 2 \times substrate solution in 1 \times reaction buffer was prepared with 20 μ M SAM and 65 μ M histone H4. A two-fold serial dilution of the 2 \times substrate solution with 12 points ranging from 0-65 μ M histone H4 in substrate dilution solution was prepared. Next, 2 μ L of the prepared 2 \times substrate dilution series was added to each well of the 384-well micro plate, followed by 2 μ L of the prepared PRMT solution. The reaction was performed following the same protocol as described under 2.18.1. The K_m value was calculated using the GraphPad Prism software.

2.19 MALDI-TOF MS-Based Assay

Enzyme activity and screening of the PRMT inhibitors were performed using a label-free MALDI-TOF MS-based assay by which the enzymatic methylation of arginine residues can be quantified.

The assay was performed in 384-well plates with a final reaction volume of 5 μ L consisting of varying concentration of CARM1 (0-800 nM), 10 μ M SAM, and 1 μ M histone peptide in assay buffer (16 mM Tris pH 8.0, 40 mM NaCl, 0.8 mM EDTA, 0.8 mM DTT, 0.32 μ g/mL BSA). The human histone H3 (residues 22-45) and H4 (residues 2-22) peptides were supplied with an N-terminal acetyl and C-terminal amide group and synthesised by Severn Biotech Ltd. (UK). The second H3 (22-45) peptide was supplied by peptides& elephants GmbH (Germany).

The reaction was started by the addition of 1 μ M H3 peptide to the reaction mixture and stopped by addition of 1.2 μ L 10 % trifluoroacetic acid (TFA) to obtain a final TFA concentration of 2 %. Reactions were performed at RT and 37 $^{\circ}$ C and with reaction times spanning from 0-60 min. Analysis of the methylated histone peptides was performed by Matrix-assisted laser desorption ionization-time-of-flight MALDI-TOF mass spectrometry as described (Guitot *et al.*, 2017). Reaction aliquots were mixed 1:1 (v/v) with the matrix α -cyano-4-hydroxycinnamic acid (CHCA) and 100 nL were spotted onto a MALDI plate using a Mosquito liquid handler, crystallised and analysed

in the positive ion reflector mode on a AB 4700 MALDI-TOF mass spectrometer (Applied Biosystems). The area of [M+H]⁺ signals of the methylated (A_{met}) and the unmethylated (A_{unmet}) histone peptides were measured and the ratio calculated:

$$R = \frac{A_{met}}{A_{unmet} + A_{met}} \times 100 \quad \text{Equation 6}$$

2.20 Surface Plasmon Resonance Assay

SPR experiments were conducted using a BIACORE S200 instrument (GE Healthcare) at 20 °C.

2.20.1 Calculation of the Maximal Response

The maximal response can be calculated if the protein and analyte is known with the following formula:

$$R_{max} = R_{ligand} \times \frac{Mr_{analyte}}{Mr_{ligand}} \quad \text{Equation 7}$$

However, this assumes a 1:1 binding and that all bound ligands are in their active state and all binding sites are accessible. This is not the case in practice, but theoretical and practical calculated R_{max} should still agree.

2.20.2 Steady-State Affinity Determinations

The affinity between the immobilised ligand and the analyte can be obtained by equilibrium analysis. Binding affinity is defined as the strength of the binding interactions between two molecules. Binding affinity is often measured by the equilibrium dissociation constant K_D . The smaller K_D , the higher the binding affinity of the ligand for its binding partner.

The equilibrium dissociation constant K_D is defined as:

$$K_D = \frac{[L] \times [A]}{[LA]} = \frac{k_d}{k_a} \quad \text{Equation 8}$$

K_D can be directly calculated from the response levels at steady-state or equilibrium phase of the interaction (R_{eq}) using the following equation:

$$\text{For } [A] \ll K_D \rightarrow R_{eq} = \frac{[A] \times R_{max}}{[A] + K_D} \quad \text{Equation 9}$$

At this phase, the association and dissociation rates are equal. The dissociation constant is obtained by measuring the equilibrium response levels over a range of different analyte concentrations. R_{eq} is then plotted against the analyte concentration and fitted to a simple saturation binding model to determine K_D . At the beginning, the equilibrium response is proportional to the analyte concentration but is limited by the maximum response R_{max} :

$$\text{For } [A] \gg K_D \rightarrow R_{eq} = R_{max} \quad \text{Equation 10}$$

It is important that only data is used where the response is at or close to equilibrium at all concentrations for a reliable measurement. In order to obtain a robust curve fit, the analyte concentration range must be large enough (at least twice the K_D) otherwise the reported values are not reliable.

2.20.3 CARM1 Fragment Screening

For binding studies of ligands that target the CARM1 substrate binding site, 20 μL of CARM1 (*circa* 7 mg/mL) were mixed with 5 μL of 5 mM SAH stock and pre-incubated on ice for 30 min, before immobilisation on the CM5 chip (GE Healthcare) surface by amine coupling following the manufacturer's instructions.

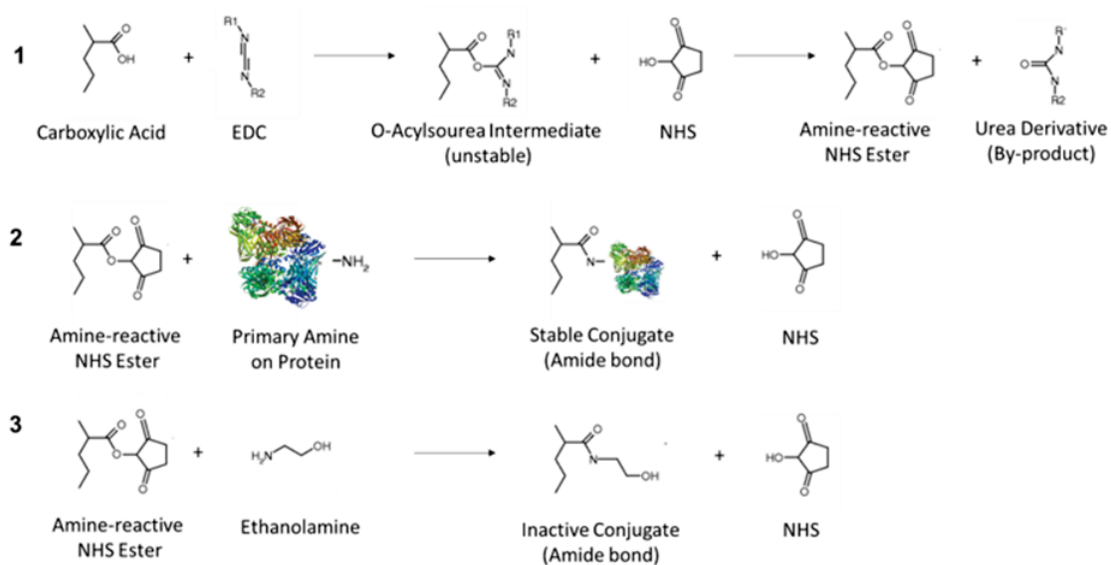


Figure 2-1 Ligand Immobilisation via Amine Coupling.

In the first step the carboxylic acid groups of the biosensor surface are activated with EDC and NHS. After surface activation, the ligand is covalently linked to the surface by an amide bond that forms between a primary amine of the protein (at a pH below its pI) and the carboxylate on the sensor surface. In the last step the remaining NHS Esters are deactivated by ethanolamine. R1: $-\text{CH}_2\text{CH}_3$; R2: $-(\text{CH}_2)_3\text{N}^+\text{H}(\text{CH}_3)_2\text{Cl}^-$. Adapted from Fischer (2010).

The protein is covalently immobilised on the activated chip surface through its primary amine groups. The principle is shown in Figure 2-1. The surface was activated by applying 0.2 M EDC/ 50 mM NHS onto the chip at a flow rate of 10 $\mu\text{L}/\text{min}$ for 420 s, followed by an ethanolamine (1 M ethanolamine-HCL pH 8.5) wash step. After surface activation, 10 μL CARM1 was mixed with 190 μL of 10 mM acetate buffer at pH 5.8 and injected onto one flow cell until the immobilised amount reached 4000- 7000 response units. Finally, ethanolamine is flown over the chip surface at 10 $\mu\text{L}/\text{min}$ for 420 s to deactivate the remaining NHS esters. Another flow cell that was used as a reference cell and has no ligand immobilised was treated in the same way but with buffer instead of the ligand.

Affinity determination experiments were performed using single cycle kinetics. Compounds were dissolved in 100 % DMSO at 100 mM and serially diluted in SPR buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 50 μM SAH, 0.05 % (v/v) Tween) to a final DMSO concentration of 1 % using an Echo acoustic dispenser (Labcyte, UK) and injected over the chip at 30 $\mu\text{L}/\text{min}$ with a contact time of 60 s and dissociation time of 180 s. Between different samples a wash with 50 % (v/v) DMSO

was performed. Eight-point solvent corrections curves for DMSO with a final concentration of 0-1.6 % (v/v) DMSO were prepared to account for any bulk shifts caused by differences in DMSO content of samples. Solvent correction curves were generated and applied to the results using the BIACORE Evaluation Software (GE Healthcare). The sensograms were reference subtracted and analysed using the BIACORE Evaluation Software and fitted to a 1:1 binding model for the determination of the equilibrium dissociation constants (K_D).

2.21 Isothermal Scanning Calorimetry

ITC experiments were performed using a MicroCal PEAQ-ITC Instrument (Malvern Panalytical) at 20 °C.

2.21.1 CARM1 Preparation

Before the ITC experiments were performed, CARM1 was buffer exchanged into buffer containing 25 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP using a PD-10 Desalting Column (GE Healthcare) at 4 °C according to the manufacturer's instructions. The column was equilibrated with 10 CV ITC buffer. The protein was eluted with 1 mL buffer. For experiments that were run in the presence of SAH, 20 μ M CARM1 was incubated with 200 μ M SAH for 20 min after the buffer exchange.

2.21.2 Analysis of Ligand Binding

Ligand titration experiments were performed at 20 °C. The cell contained CARM1 protein with or without SAH, the syringe SAH, SAM or inhibitor. The first injection of 0.5 μ L was followed by 19 2 μ L injections at a stirring speed of 750 rpm and 120 s spacing between them. Ligand titration were performed with 20 μ M CARM1 in the cell and 200 μ M ligand in the syringe. The SAH titration was performed with 15 μ M CARM1 in the cell and 230 μ M SAH in the syringe. A heat of dilution control was performed by titration of ligand into the sample cell containing only buffer and SAH where applicable.

2.21.3 Data Analysis

The ITC raw data was analysed and plotted using the Origin software (OriginLab 7.0). The baseline correction was also performed in Origin and manually corrected if needed. The data was then baseline corrected and the heat of dilution subtracted. A one-site binding model was used.

2.22 X-Ray Crystallography

Protein crystals of the catalytic domain of human CARM1 (CARM1-1) and mouse PRMT2 were obtained using the sitting drop vapor diffusion technique. Crystallisation conditions were screened using different sparse matrix screens, including Index (Hampton Research), Pact (Molecular Dimensions), Structure (Molecular Dimensions) and JCSG+ (Molecular Dimensions). The screens were set up in 96-well 2-drop MRC crystallization plates (Molecular Dimensions) using a Mosquito LCP Liquid Handler (TTP LabTech) dispensing 2 droplets of 100 nL protein solution mixed with 100 nL and 200 nL of 80 μ L reservoir solution. The plates were sealed with a clear seal film (Hampton Research) and stored at 20 °C and/or 4 °C. Initial crystal “hits” were optimised by screening different pH, precipitant concentrations, protein concentrations, salt, and additives conditions.

2.22.1 CARM1 Crystallisation

For the generation of CARM1 crystals with SAH, the protein was used at a concentration of 7 mg/mL in the final SEC buffer (40 mM HEPES pH 8.0, 600 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol). For each plate, 32 μ L of CARM1 was mixed with 5.1 μ L of 5 mM SAH and incubated for 30 min at RT. CARM1 crystals were prepared using the hanging drop vapour diffusion method and set up in 96-well 2-drop MRC Crystallisation Plates (Molecular Dimensions) using a Mosquito LCP Liquid Handler (TTP LabTech). 2 droplets of 100 nL protein solution mixed with 200 nL of 80 μ L reservoir solution were dispensed. The plates were sealed with a clear seal film (Hampton Research) and stored at 20 °C. Crystals appeared after a few days.

2.22.2 Crystal Harvest

Crystals were harvested using CryoLoops (Hampton Research) with the inner diameter matching the crystal size. CARM1 and PRMT2 crystals were cryoprotected before harvesting if no cryoprotectant was present in the growth conditions. For cryoprotection, each crystal was transferred into 10 μ L droplets consisting of 20 % (w/v) PEG400 or 25 % (w/v) EG in reservoir solution. Afterwards they were transferred to a dewar with liquid nitrogen for flash freezing.

2.22.3 DMSO Tolerance Testing of CARM1 Crystals

Determination of the highest tolerable DMSO concentration for CARM1 crystals was conducted by soaking existing CARM1 crystals with SAH in drops consisting of different DMSO concentration and 20 % PEG400 in Index D6 reservoir solution (0.1 M BIS-TRIS pH 5.5, 25 % (w/v) Polyethylene glycol 3,350). Tested DMSO concentrations were 0, 5, 10, 15 and 20 % (v/v). Three crystals were soaked for each condition, incubated overnight, and harvested the next morning as described.

2.22.4 Ligand Soaking into CARM1 Crystals

100-500 mM ligand stocks were prepared in 100 % DMSO. The ligands were then soaked into existing CARM1 with bound SAH crystals that were grown in 0.1 M BIS-TRIS pH 5.5, 25 % (w/v) Polyethylene glycol 3,350 (Index Screen condition D6). The crystals were transferred into 10 μ L droplets consisting of 20 % (w/v) PEG400, 10-50 mM compound and reservoir solution using cryo-loops (Hampton Research). After 24 h incubation at RT (20 °C) the CARM1 crystals were harvested with cryo-loops and flash-frozen in liquid nitrogen as described above.

2.22.5 Co-crystallization of SNF and CARM1

In order to obtain CARM1 crystals in complex with the known PRMT inhibitor SNF, co-crystallisation was used. 200 μ M CARM1 in final SEC buffer (40 mM HEPES pH 8.0, 600 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol) was mixed with 1 mM of the ligand prior to crystallisation and incubated for 1 h at RT. Co-crystallisation trials were set up in 96-well 2-drop MRC crystallisation plates (Molecular Dimensions) using a Mosquito

LCP Liquid Handler (TTP LabTech) dispensing 2 droplets of 100 nL protein solution mixed with 100 nL and 200 nL of 80 μ L reservoir solution. The plates were sealed with a clear seal film (Hampton Research) and stored at 20 °C.

2.22.6 Co-crystallization of Ligands and CARM1

Co-crystallisation of CARM1 with ligands was used when soaking experiments were not successful. CARM1 was diluted in SEC buffer (40 mM HEPES pH 8.0, 600 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol) to a concentration of 1 mg/mL and a final volume of 256 μ L. 256 μ L of the CARM1 solution was then mixed with 6.4 μ L of 100 mM ligand stock in 100 % DMSO resulting in a final DMSO concentration of 2.5 %. In some cases, SAH was added to the protein at a final concentration of 0.8 mM. After incubation at 4 °C overnight, CARM1 was concentrated back to 5-7 mg/mL using Amicon Ultra-2 Centrifugal Filter Units with a molecular weight cut-off of 30 kDa (Merck). Crystallisation plates were set up at 20 °C as described above.

2.22.7 Data Collection and Analysis

X-ray diffraction data were collected at 100 K on beamline, I03 or I04 at the Diamond Light Source (Didcot, UK). Crystal structures were determined within the CCP4 software suite by using the graphical user interface CCP4i2 (Potterton *et al.*, 2018). Data processing was performed using Xia2 and the structures were solved by molecular replacement using MOLREP or Phaser with previously solved structures as search models, as implemented within autoPROC (Vonrhein *et al.*, 2011). Ligand libraries and coordinates were generated from the corresponding SMILES string and fitted to density in COOT (Emsley and Cowtan, 2004). The structures were refined by repetitive cycles of refinement and model building using REFMAC5 and COOT (Murshudov *et al.*, 1997). Structure validation was also performed in COOT using the available tools including Ramachandran plots outliers. Structural figures and protein superposition were performed in CCP4mg and Pymol (McNicholas *et al.*, 2011, Janson *et al.*, 2016).

Chapter 3 Recombinant Protein Expression and Characterisation of Different Type 1 PRMTs

3.1 Introduction

The first aim of the project was the production of different members of the PRMT family. PRMT2 was needed in large amounts to conduct enzyme characterisation, crystallisation, and binding studies. Other PRMT family members including PRMT1 and CARM1 were needed for later selectivity assays. At first, the expression host *Escherichia coli* (*E. coli*) was chosen because it has many advantages compared to other protein expression strategies. *E. coli* cells are very fast-growing with a doubling time of 20 min under optimal conditions and high cell densities are possible (LB media, 37 °C ~ 1×10^{10} cells/mL) (Sezonov *et al.*, 2007). Moreover, different growing media exist, and they are not expensive and readily available (Rosano and Ceccarelli, 2014). Moreover, transformation of plasmid DNA into *E. coli* cells is straightforward and protein expression can be tightly controlled at both the level of transcription and of translation. However, there are also some disadvantages. These include (i) eukaryotic post-translational modifications such as phosphorylation or acetylation are not possible. The absence of these modifications can cause improper folding (leading to protein accumulation in inclusion bodies (Carrio and Villaverde, 2002), inactivity, or poor expression levels of the target protein (Sahdev *et al.*, 2008). (ii) Codon usage of the bacterium often differs from the target protein. An imbalance in the pool of charged tRNAs can cause problems such as premature transcription or translation termination, frameshifts, or low protein expression. However, *E. coli* strains exist which carry helper plasmids that express the needed tRNAs and target DNA codon optimisation for *E. coli* expression is also possible (Sørensen and Mortensen, 2005).

In the initial step different protein constructs for all three enzymes, PRMT1, PRMT2 and CARM1 were designed. For HsCARM1 and RnPRMT1 the constructs were chosen according to the literature precedents. However, from the literature research it became clear that solubility issues might arise during *E. coli* expression of PRMT2. To address this problem, a large number of HsPRMT2 and MmPRMT2 constructs were designed.

3.2 Design of Protein Constructs for Recombinant Expression

Different expression constructs of human and mouse PRMT2 were designed by identifying and removing flexible regions at the N- and C-termini. Removing flexible regions can result in higher soluble protein expression in *E. coli*. Moreover, protein heterogeneity may be reduced, which was shown to decrease crystallisation probability by interfering with directional nucleation and crystal lattice formation (Holcomb *et al.*, 2017). For human PRMT2, both the wild-type and the codon-optimized sequences were employed, whilst for mouse PRMT2, only expression of the codon optimised version was available. For PRMT1, one literature construct comprising amino acids 11-353 of the rat protein (Figure 3-1) was chosen, because this fragment was previously identified as a very stable fragment via limited proteolysis (Zhang and Cheng, 2003).

Moreover, it was successfully expressed in *E. coli* cells, showed similar enzyme activity and oligomerization behaviour to the full-length protein, and was more easily crystallised than the full-length protein (Zhang and Cheng, 2003). For CARM1, a literature construct comprising amino acid residues 135-482 of the human protein was chosen (Figure 3-1), because it gave the highest yield when expressed in *E. coli* cells and no loss in enzyme activity compared to the full-length protein (Sack *et al.*, 2011). The full-length CARM1 protein was not chosen because it cannot be expressed in *E. coli* cells in a soluble form (Chumanov *et al.*, 2011). The chosen HsCARM1 and RnPRMT1 literature constructs were reported to have 100 % enzyme activity (Sack *et al.*, 2011, Zhang and Cheng, 2003).

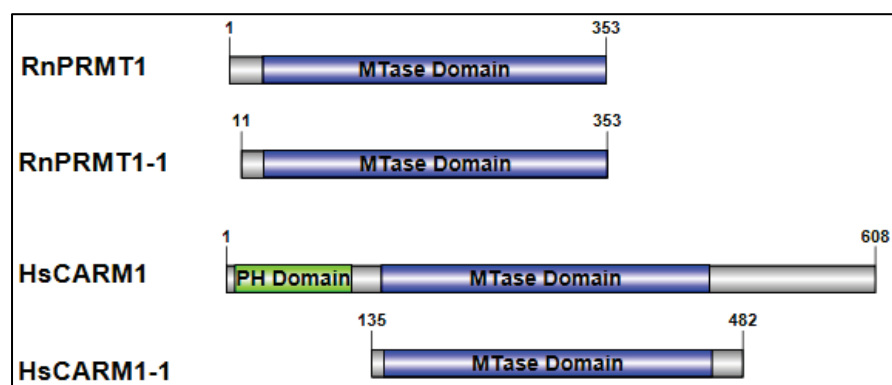


Figure 3-1 Schematic Illustration of Rat PRMT1 and Human CARM1 and Expression Construct Selection.

For RnPRMT1 the literature construct comprising residues 11-353 was chosen. For HsCARM1 a construct comprising residues 135-482 that is missing the N-terminal PH domain was designed.

In contrast, previous studies of human PRMT2 activity were contradictory. Some studies did not see any activity for wild type PRMT2 (Scott *et al.*, 1998, Qi *et al.*, 2002). Another study, reported PRMT2 to be only active if the SH3 domain is present (Lakowski and Frankel, 2009). Thus, different PRMT2 constructs were designed with and without the SH3 domain. Eight different expression constructs were designed for each human and mouse PRMT2 DNA template (Figure 3-2 and Figure 3-3).

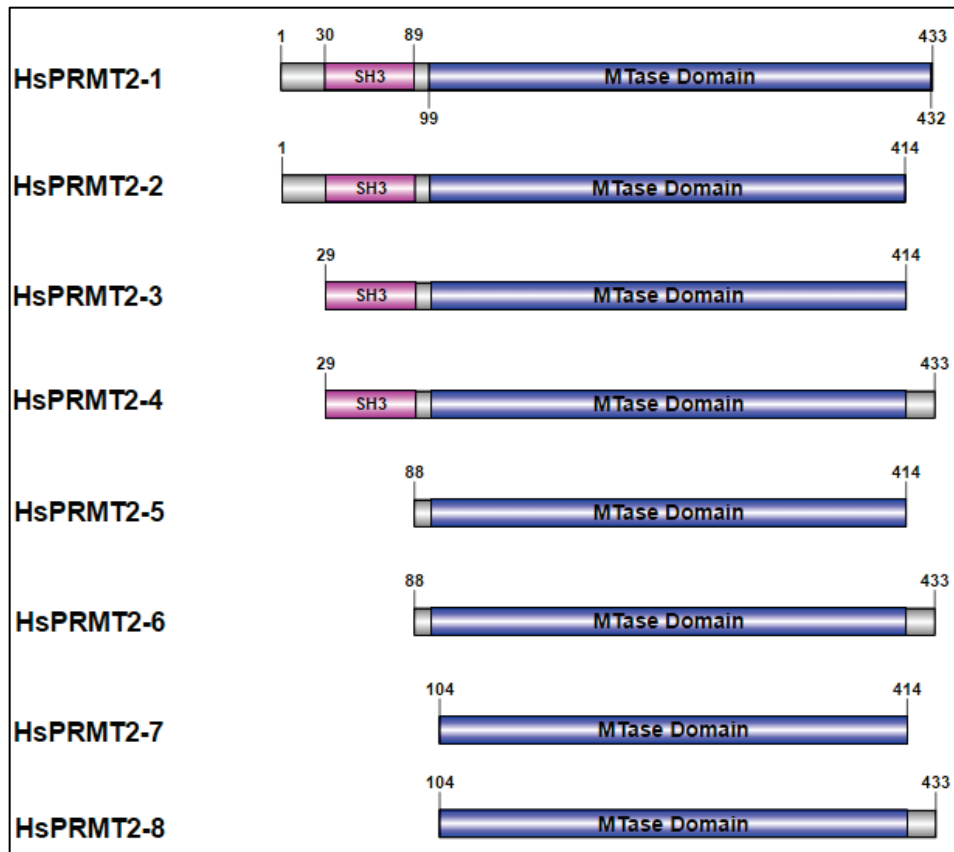


Figure 3-2 Expression Constructs Designed for Human PRMT2.

Eight different human PRMT2 constructs were designed, including full-length human PRMT2 and shorter constructs missing the N-terminal SH3 domain, which is believed to be necessary for enzyme activity.

The SH3 domain boundaries were identified via multiple sequence alignments and secondary structure predictions. The secondary structure of PRMT2 was predicted using the structure prediction server PSIPRED (McGuffin *et al.*, 2000).

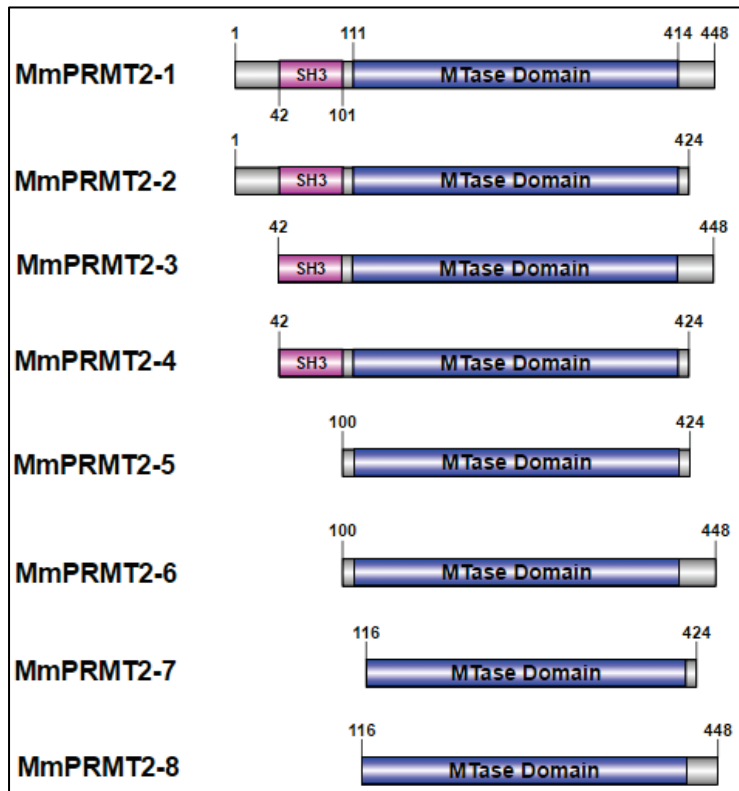


Figure 3-3 Designed Expression Constructs for Mouse PRMT2.

Similar to the human PRMT2 expression constructs, 8 different mouse PRMT2 constructs were designed.

The Disorder Prediction Server DISOPRED2 (Ward *et al.*, 2004) was used to identify disordered intrinsic regions of PRMT2, the results for HsPRMT2 are shown in Figure 3-4.

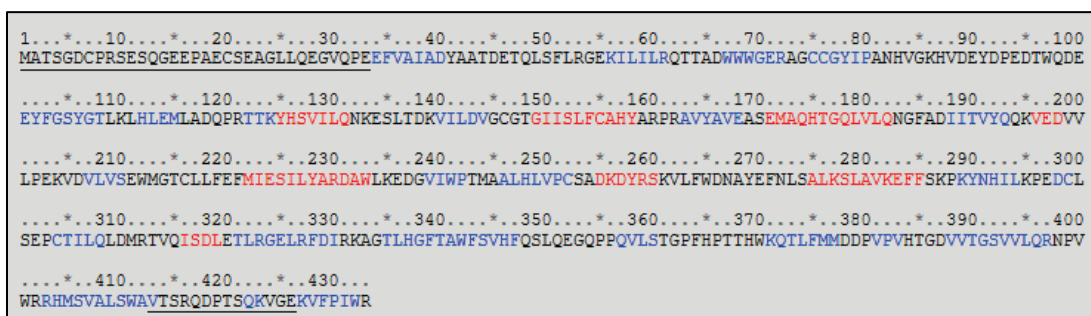


Figure 3-4 Secondary Structure Prediction for Human PRMT2 with the PSIPRED Server.

Loop regions (black), strand (blue) and helical regions (red) were predicted by PSIPRED (McGuffin *et al.*, 2000). Disordered regions were predicted by DISOPRED2 (Ward *et al.*, 2004) and are underlined. No internal disordered domains were predicted but the N- and C-termini were predicted to be disordered.

No internal disordered sequences were predicted, but large disordered regions were identified at the N- and C-termini. Disordered regions of proteins are known to have negative effects on the protein's solubility and ability to crystallise (Deller *et al.*, 2016). Thus, it is important when planning protein expression constructs that are needed for structural studies, to identify those regions and, if possible, exclude them from the construct. The software RONN (Yang *et al.*, 2005) was additionally used to detect natively disordered regions in mouse and human PRMT2. An exemplary result of this analysis using the human PRMT2 sequence is shown in Figure 3-5. In agreement with the results from the PSIPRED server, natively disordered regions were found at the N- and C-terminus of PRMT2 but not in the internal region.

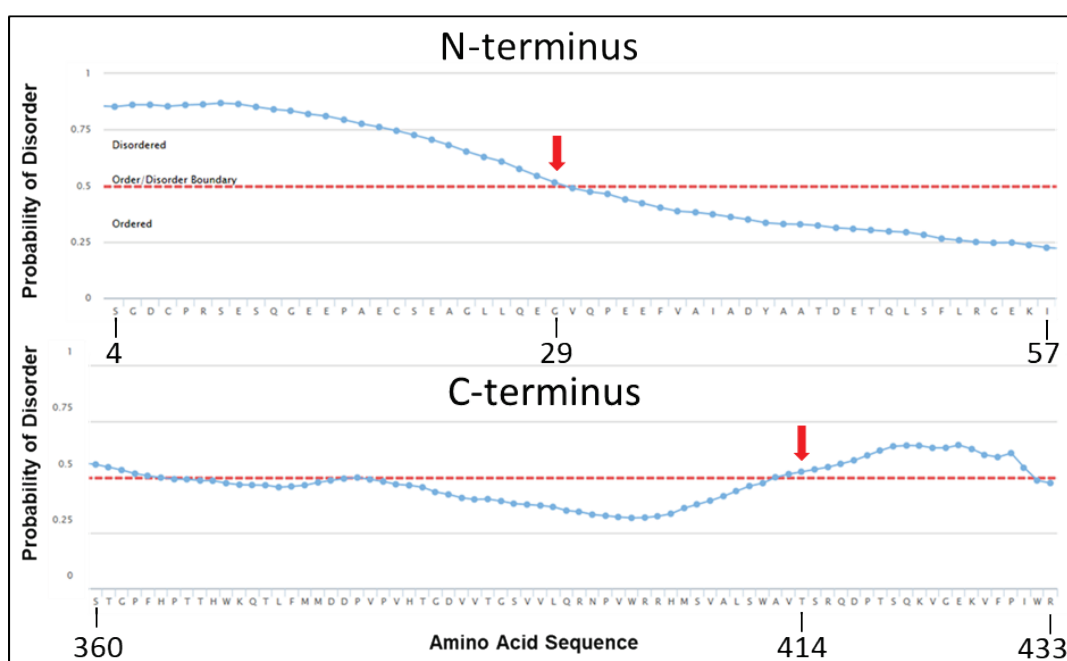


Figure 3-5 *In silico* Prediction of Disordered Regions of Human PRMT2 using RONN.

Disordered region predictions were performed using RONN (Yang *et al.*, 2005). The upper graph shows the probability of disorder for the N-terminal region of the protein (HsPMRT2, UniProtKB: P55345), the lower graph for the C-terminal region. The red line indicates the order/disorder boundary. Both regions show a high disorder probability. The red arrows indicate the start/end of the designed constructs.

Based on these results, human PRMT2 constructs were designed to start at residue Gly29, to include the SH3 domain, or residues Val88 and Gly104, excluding the SH3 domain and thus a disordered region. The designed constructs end at residue Thr414 to remove the disordered protein regions at the C-terminus or alternatively at

Arg433. In the case of mouse PRMT2, the constructs started either at residue Leu42, which includes the SH3 domain, or at Leu100 or Asp116, omitting the SH3 domain. For the C-terminal ends, residues Gly424 and Ser448 were chosen based on disordered domain predictions.

For expression tests in *E. coli*, the DNA-sequences were sub-cloned into modified pET3d expression vectors expressing recombinant PRMT with N-terminal removable fusion tags (cleavable with 3C protease), kindly provided by Dr Richard Heath. Given literature precedent, all constructs were expressed with a Glutathione-S-transferase (GST)-tag and a polyhistidine (His₆)-tag. The removal of the expression tags after purification is necessary especially if performing crystallographic studies or protein-ligand interaction studies, as they can influence enzyme activity and structure.

3.3 Small Scale Expression Tests of PRMT1, PRMT2 and CARM1 in *E. coli* Cells

Small scale expression tests were used to screen all designed PRMT constructs for soluble expression and to identify the optimal expression conditions for later large-scale expression. Small scale expression tests have the advantage that they allow high-throughput expression and that all constructs are tested in parallel which minimizes the variability. Initially, *E. coli* BL21(DE3)Gold cells were used as an expression host, and at a later stage other cell lines including *the E. coli* strains ArcticExpress, Rosetta(DE3)plysS, DH5 α , and BL21(DE3) were screened. They were chosen due to literature precedents or because they carry features which can help to overcome solubility issues. In addition to host strains, different growth media and temperatures were screened. A summary of the different strains and their properties can be found in Appendix B.

3.3.1 Small-Scale Expression Test in *E. coli* BL21(DE3)Gold

In total, small-scale expression tests of 56 different constructs carrying either an N-terminal GST- or His₆-tag were conducted. A summary of the constructs is given in Table 3-1. The bands at ~40 kDa in the post cleavage samples correspond to GST-3C-protease.

All constructs were screened for soluble protein expression in *E. coli* BL21(DE3)Gold cells, which were grown at 37 °C in Auto Induction Medium (AIM) for 6 h and then shifted to 18 °C overnight, as described in detail in Section 2.12. AIM medium was developed by Studier *et al.* in 2005. He defined a new complex medium consisting of glucose, lactose and glycerol that allows autoinduction of *E. coli* cells in late log phase without the need of IPTG. AIM is especially useful for small-scale expression tests as it is difficult to monitor and induce all wells at the correct time. First glucose is used as an energy source which the *E. coli* cells prefer over lactose. Once the glucose is depleted, in mid-to-late log phase, metabolism switches to using lactose. The lactose intake induces *lacUV5*- controlled protein expression and thus expression of the target protein starts in late log phase at high cell density (Rosano and Ceccarelli, 2014, Studier, 2014).

Table 3-1 Summary of PRMT Constructs Tested for Expression in *E. coli* BL21(DE3)Gold.

Protein Name	AA Range	Expression Vector	N-terminal Tags
HsPRMT2-1	1-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-2	1-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-3	29-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-4	29-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-5	88-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-6	88-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-7	104-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-8	104-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-9	1-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-10	1-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-11	29-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-12	29-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-13	88-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-14	104-433	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-15	104-414	pET-3dM	His ₆ -3C; GST-3C
HsPRMT2-16	104-433	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-1	1-488	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-2	1-424	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-3	42-488	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-4	42-424	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-5	100-424	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-6	100-488	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-7	116-424	pET-3dM	His ₆ -3C; GST-3C
MmPRMT2-8	116-448	pET-3dM	His ₆ -3C; GST-3C
RnPRMT1-1	11-353	pET-3dM	His ₆ -3C; GST-3C
RnPRMT1-2	11-353	pET-3dM	His ₆ -3C; GST-3C
HsCARM1-1	135-482	pET-3dM	His ₆ -3C; GST-3C
HsCARM1-2	135-482	pET-3dM	His ₆ -3C; GST-3C

The recombinant proteins were purified via their purification tag using spin columns as described in Section 2.13. After purification, the fusion tags were removed via 3C-protease cleavage. Paired samples of each protein (HsPRMT2-1-16, MmPRMT2-1-8, RnPRMT1-1 and -2 and HsCARM1-1 and -2) with the GST- or His₆-tag and without their fusion tag were analysed by SDS-PAGE. Figure 3-6 shows an exemplary SDS-PAGE analysis of GST-tagged constructs. From all tested constructs, only two, GST-HsPRMT2-2 and GST-MmPRMT2-1 yielded faint bands visible in the eluted fractions after batch purification (Figure 3-6, gel C and D, black arrow). However, the bands for the tagged and untagged proteins were not of the expected molecular weight (73/77 and 47/51 kDa respectively), instead they were running at a higher molecular weight (Figure 3-6 gel B, lanes 2/3 and gel C, lanes 3/4). A single band below 70 kDa is visible in all samples and is most likely chaperone.

A negative and positive control (Cyclin T1 (0-259), 26 kDa) was tested in parallel to ensure that the protocol was working. A band corresponding to the GST-tagged positive control can be clearly seen at the expected molecular weight of 52 kDa (Figure 3-6 gel D; lane 8/9) and also the shift after GST-tag removal (Figure 3-6, gel D; lane 10/11). The bands at ~40 kDa in the post cleavage samples correspond to GST-3C-protease.

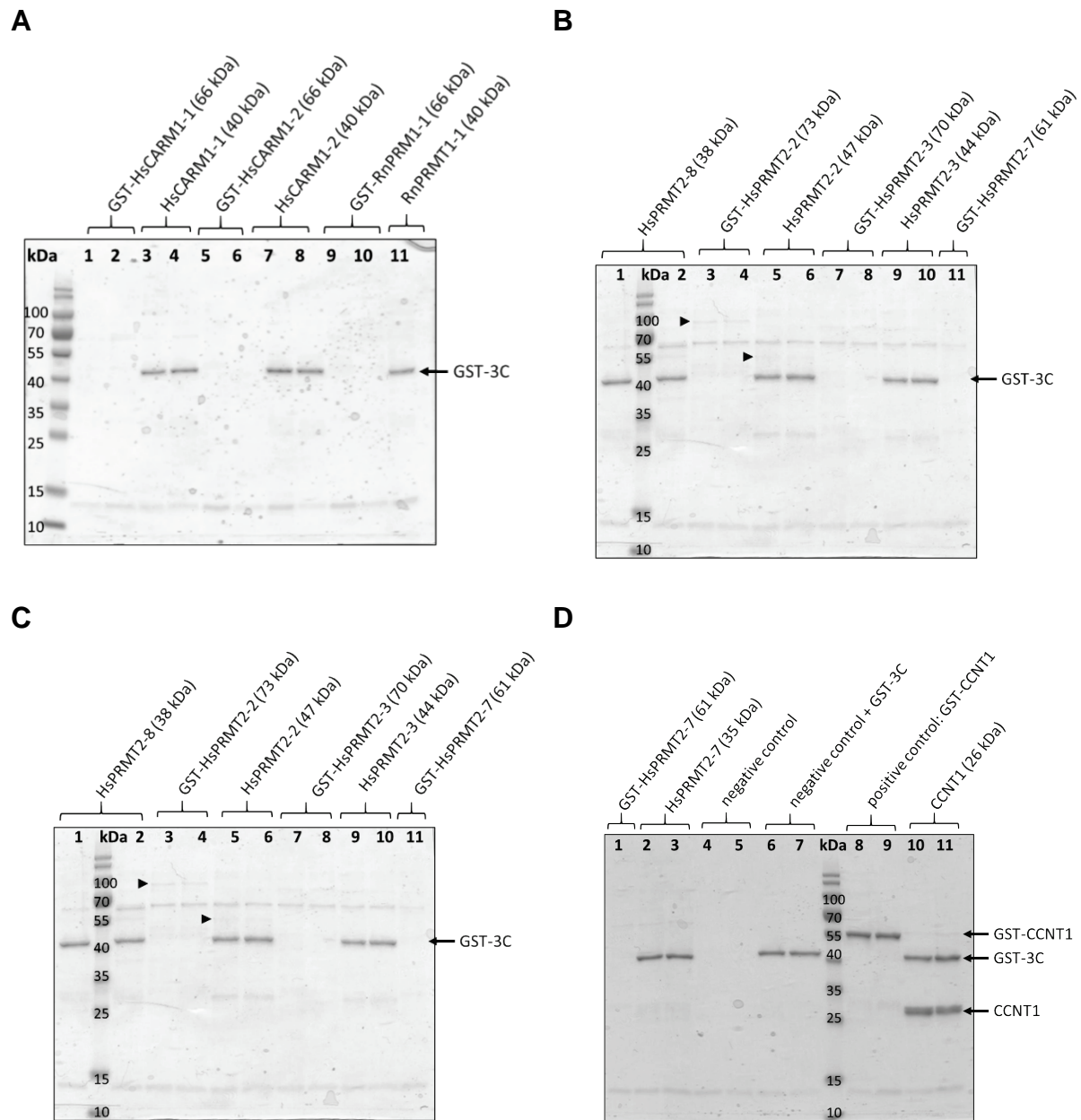


Figure 3-6 Small Scale Expression Test of Different GST-tagged Constructs in *E. coli* BL21(DE3)Gold.

Constructs are identified above each lane. After batch purification, one sample was analysed by SDS-PAGE, another sample was cleaved with 3C-protease overnight and then subject to SDS-PAGE. The arrows indicate bands that shift after tag removal. Pre-stained protein ladder: PAGE Ruler Pre-stained (ThermoFisher). Protein visualised by InstantBlue staining.

In order to analyse whether the absence of bands in the elution fractions of most constructs, indicating low protein yields, was a result of low expression levels or if the protein was produced in an insoluble form, a lysate and a cleared lysate/soluble fraction sample for each tested construct was analysed by SDS-PAGE. Figure 3-7 shows an exemplary gel: in all lysate samples (L) an intense protein band is visible at the expected molecular weight, indicated by a black arrow. However, the bands disappear

after clarification of the cell lysate (S) indicating that the recombinant proteins are produced at high levels but in an insoluble form.

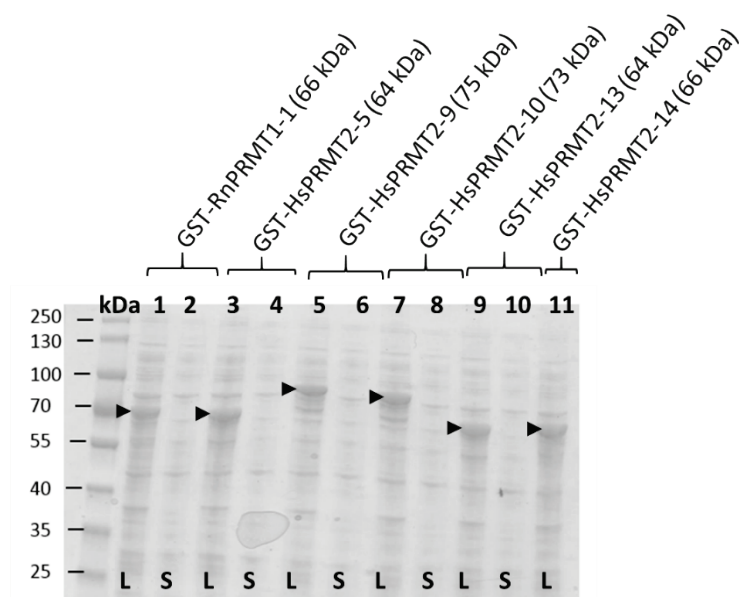


Figure 3-7 Small Scale Expression Test of different GST-tagged Constructs in *E. coli* BL21(DE3)Gold.

Expected MWs are GST-RnPRMT1-1, 65.7 kDa; GST-HsPRMT2-5, 63.5 kDa; GST-HsPRMT2-9, 75.2 kDa; GST-HsPRMT2-10, 73.0 kDa; GST-HsPRMT2-13, 63.5 kDa; GST-HsPRMT2-14, 65.7 kDa. Protein ladder: PAGE Ruler Prestained (ThermoFisher).

3.3.2 Small Scale Expression Test in *E. coli* BL21(DE3)pLysS and DH5 α

The expression tests conducted in *E. coli* BL21(DE3)Gold cells showed that most recombinant proteins are expressed at high levels but in an insoluble form. Changing the host strain can help to overcome solubility issues and can help with the production of potential toxic proteins. In the next step, a number of constructs were tested in a different *E. coli* strain, BL21(DE3)pLysS. This strain carries the pLysS plasmid and produces the T7 lysozyme which reduces basal expression levels in the non-induced state and allows the expression of toxic protein (Studier, 1991). Full-length HsPRMT2 was also tested in *E. coli* DH5 α cells. This strain is most suitable for cloning and subcloning and not protein expression, but the protein was reported to be successfully expressed in this strain (Lakowski and Frankel, 2009). Additionally, different media including LB (Lysogeny Broth) and Terrific Broth (TB) were also screened (Studier, 2005). AIM, LB

and TB media all differ in their composition (Table 2-3) including amounts of carbohydrates and divalent cations. TB for example has more than double the amount of tryptone than LB medium, allowing higher cell densities.

Selected PRMT1 and PRMT2 constructs were expressed in BL21(DE3)pLysS cells. The tested constructs were GST-tagged RnPRMT1-1, HsPRMT2-1, HsPRMT2-2, and MmPRMT2-1. Lysate and soluble fraction samples were analysed by SDS-PAGE (Figure 3-8). However, protein expression was still not detected in the soluble fraction and only very faint bands or no detectable bands were observed in the whole cell fraction. The rest of SDS-PAGE analysis results can be found in Appendix D.

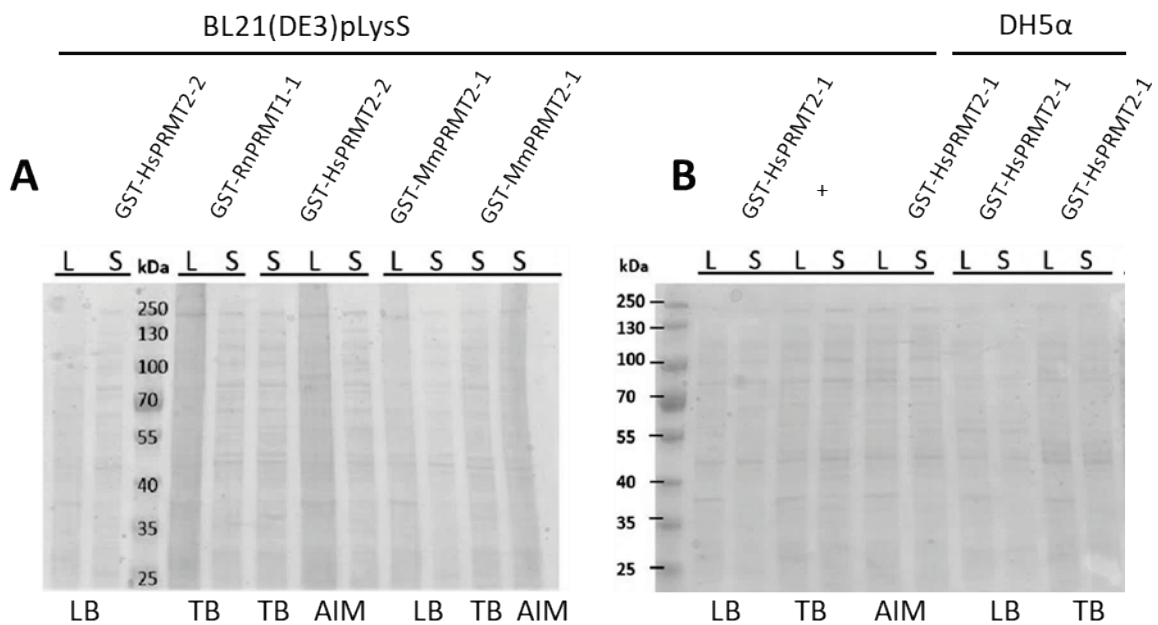


Figure 3-8 SDS-PAGE Analysis of Small Scale Expression Test of different GST-tagged Constructs in *E. coli* DH5α and BL21(DE3)pLysS.

Constructs were grown in *E. coli* BL21(DE3)pLysS in TB, LB and AIM medium as indicated for 5 h at 37 °C and then transferred to 18 °C overnight. Lysate (L), soluble fraction (S). Protein ladder: PAGE Ruler Pre-stained (ThermoFisher). Expected MWs are GST-HsPRMT2-2, 73.0 kDa; GST-RnPRMT1-1, 65.6 kDa; GST-MmPRMT2-1, 76.6 kDa; GST-HsPRMT2-1, 75.2 kDa.

3.3.3 Small Scale Expression Test in *E. coli* Rosetta(DE3)pLysS and ArcticExpress

Changing the growth medium or strain did not improve the expression of soluble recombinant protein. In the next step, another two host strains were tested for expression with the aim to increase protein solubility. Several constructs were tested in the *E. coli*

BL21(DE3)Gold derivative ArcticExpress. These cells express the cold-adapted chaperone proteins chaperonin 10 (Cpn10) and 60 (Cpn60) from the bacterium *Oleispira antarctica* which possess high protein refolding activity even at low expression temperatures. This activity is in contrast to the native *E. coli* chaperone complex GroEL-GroES which shows less than 30 % refolding activity at 12 °C (Horwich *et al.*, 2006). Expression tests in the *E. coli* strain Rosetta(DE3)pLysS were also conducted. Rosetta strains enable the expression of eukaryotic proteins that contain rare codons by supplying the tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a separate plasmid. As a LysS strain, they express T7 lysozyme by which basal expression levels are reduced.

SDS-PAGE analysis of selected constructs in both strains can be seen in Figure 3-9. In the whole cell samples, some bands with the expected molecular size are visible, indicated by black arrows. However, only very faint or no protein bands were detectable in the soluble fractions. The only exception is GST-HsPRMT2-2 expressed in *E. coli* Rosetta(DE3)pLysS cells in AIM (Figure 3-9 B, Lane 12). The size of the band in whole cell and soluble fraction, highlighted by a red square, is in good agreement with the expected molecular weight of 73 kDa.

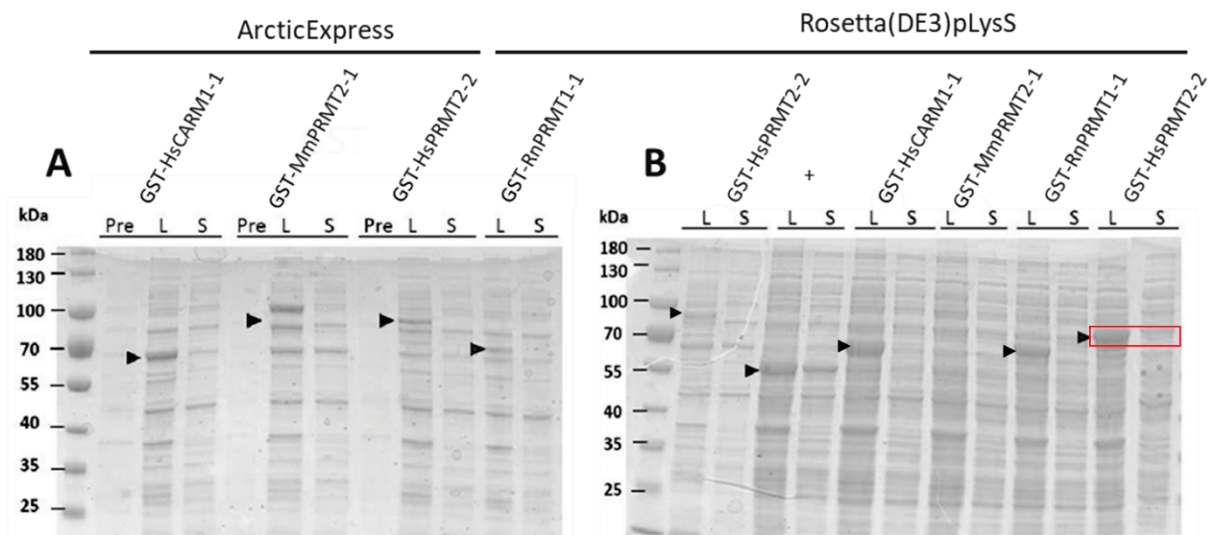


Figure 3-9 SDS-PAGE Analysis of Small-Scale Expression Test of different GST-tagged Constructs.

Lysate (L) and soluble fraction (S) samples from small-scale expression tests conducted in *E. coli* ArcticExpress grown in LB (A) and in *E. coli* Rosetta (DE3) pLysS grown in AIM (B) were analysed for solubility of the expressed protein. Protein ladder: PAGE Ruler Prestained (ThermoFisher). Expected MWts are GST-HsCARM1-1, 65.3 kDa; GST-MmPRMT2-1, 76.6 kDa; GST-HsPRMT2-9, 75.2 kDa; GST-RnPRMT1-1, 65.7 kDa; GST-HsPRMT2-2, 73.0 kDa.

In summary (Table 3-1), despite extensive screening of several constructs for each protein target and a wide range of expression conditions, conditions were not identified that generated significant amounts of soluble protein. In a few cases (for example see Figure 3-7), the target protein was expressed in the lysate but was subsequently found to be insoluble. The multi-domain organisation of these constructs suggested that protein denaturation and refolding would not be successful.

However, small scale expression trials do not always track with subsequent levels of protein expression in larger volume cultures. Before ceasing to work on these constructs, three constructs that had shown the most promising protein expression in the small-scale trials (GST-HsPRMT2-2, GST-HsPRMT2-9, and His₆-RnPRMT1-1) were selected for testing at a larger scale.

Table 3-2 Summary of Small-Scale Expression Tests for Rat PRMT1 and Human CARM1 Constructs in the Whole Cell Fraction or the Soluble Fraction.

Name	Tag	Expression		Soluble	
		WT	Codon optimised	WT	Codon optimised
RnPRMT1	GST	✓	-	✗	-
	His ₆	✓	✓	✗	✗
HsCARM1	GST	✓	✓	✗	✗
	His ₆	✓	✓	✗	✗
HsPRMT2-1	GST	✗	✓	✗	✗
	His ₆	✓	✓	✗	✗
HsPRMT2-2	GST	✓	✓	✓	✗
	His ₆	✓	✓	✗	✗
HsPRMT2-3	GST	✓	✗	✗	✗
	His ₆	✓	-	✗	-
HsPRMT2-4	His ₆	✓	✓	✗	✗
	GST	✓	✓	✗	✗
HsPRMT2-5	GST	✓	✓	✗	✗
	His ₆	✓	✓	✗	✗
HsPRMT2-6	GST	✓	✓	✗	✗
HsPRMT2-7	GST	✓	✓	✗	✗
	His ₆	✓	-	✗	-
HsPRMT2-8	GST	✓	✓	✗	✗
	His ₆	✓	✓	✗	✗
MmPRMT2-1	GST	✓		✗	
	His ₆	✓		✗	
MmPRMT2-2	GST	✓		✗	
	His ₆	✓		✗	
MmPRMT2-3	GST	✗		✗	
MmPRMT2-4	His ₆	✗		✗	
MmPRMT2-5	GST	✓		✗	
	His ₆	✓		✗	
MmPRMT2-6	GST	✓		✗	
	His ₆	✓		✗	
MmPRMT2-7	GST	✓		✗	
MmPRMT2-8	GST	✓		✗	
	His ₆	✓		✗	

Green ticks represent high expression levels of the tested constructs, black ticks indicate low expression levels, and red crosses indicate that no protein was detected in the analysed fraction. No protein bands were detectable in the soluble fractions in any of the tested *E. coli* strains.

3.4 Large Scale Expression Tests of Selected Constructs in *E. coli*

The small-scale expression tests did not show significant levels of soluble protein for any of the tested constructs. However, it might be possible that the expression and solubility of the recombinant proteins is influenced by differences in aeration and culture conditions of small volume cultures compared to large scale culture volumes. Moreover, lysis and protein purification methods used for the small-scale expression tests are different to large scale experiments (Section 2.12). As a result, some proteins might be insoluble when expressed on a small scale but could be expressed in a soluble form when grown in larger culture volume. To test this hypothesis, a large-scale expression test with the codon-optimized full-length HsPRMT2 (HsPRMT2-9) in Rosetta(DE3)pLysS was performed. Additionally, GST-HsPRMT2-2 and GST-HsPRMT2-9 were expressed in *E. coli* DH5 α cells, because expression of human PRMT2 in this strain was reported in the literature (Lakowski and Frankel, 2009). Furthermore, His₆-RnPRMT1-1 was expressed at a larger scale in *E. coli* BL21(DE3) cells and 2 \times YT medium, as described in the literature (Osborne *et al.*, 2007).

After cell harvest, 1 g of each cell pellet was lysed and purified. Figure 3-10 shows the SDS-PAGE analysis of the pre- and post-induction samples (A) and the soluble fraction samples and re-suspended pellet samples (B) for the tested constructs. No protein expression was detected for GST-HsPRMT2-2, or the codon-optimized version of full-length GST-tagged HsPRMT2 (GST-HsPRMT2-9) both of which were expressed in *E. coli* DH5 α (Figure 3-10 A, lane 1-4).

The other constructs did show some degree of protein expression after induction, but only His₆-RnPRMT1-1 which was expressed in *E. coli* BL21(DE3) and 2 \times YT medium showed a protein band in the soluble fraction (Figure 3-10 B, lane 5). GST-RnPRMT1-1 expressed in BL21(DE3) cells and GST-HsPRMT2-9 expressed in Rosetta(DE3)pLysS cells could not be expressed in a soluble form (compare Figure 3-10 B, lane 7-10).

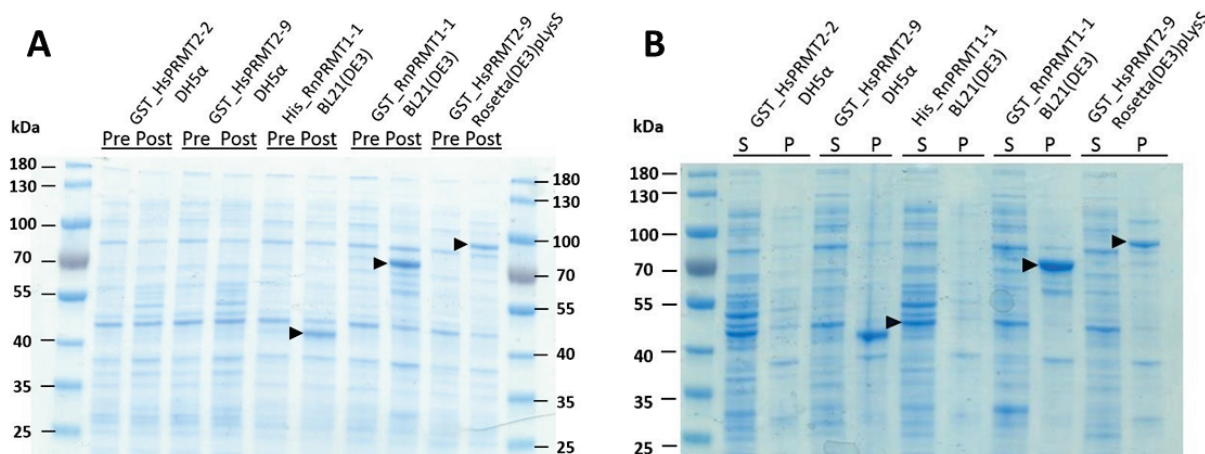


Figure 3-10 SDS-PAGE Analysis of Large-Scale Expression Test of different GST- and His-tagged Constructs in different *E. coli* strains.

A) Pre- and post-induction samples for each tested protein construct were analysed by SDS-PAGE and subsequent InstantBlue staining. Protein expression was not detectable in DH5 α cells (lane 1-4). The arrows indicate protein bands at the expected size in the post-induction samples. **B)** Soluble fraction samples (S) and re-suspended pellet samples (P) were analysed. The arrows indicate protein bands at the expected size. Most of the expressed proteins can be found in the insoluble fraction. The exception is His-RnPRMT1-1 (lane 5). Protein ladder: PAGE Ruler Pre-stained (ThermoFisher). Expected MWs are GST- HsPRMT2-2, 73.0 kDa; GST- HsPRMT2-9, 75.2 kDa; His- RnPRMT1-1, 46.3 kDa; GST- RnPRMT1-1, 65.7 kDa.

To confirm the protein identities, samples were also western blotted using an anti-GST antibody (Figure 3-11). For the constructs which were expressed in *E. coli* DH5 α , no GST-tagged protein was detected, but for GST-RnPRMT1-1 expressed in *E. coli* BL21(DE3) in 2 \times YT medium and also for GST-HsPRMT1-9 expressed in *E. coli* Rosetta(DE3)pLysS and LB medium, bands of the correct molecular weight were visible (Figure 3-11, lane 7 and 8). Therefore, and in order to prepare larger amounts of protein for further analysis, the remainder of the pellets of His₆-RnPRMT1-1, GST-RnPRMT1-1 and GST-HsPRMT2-9 were purified and analysed.

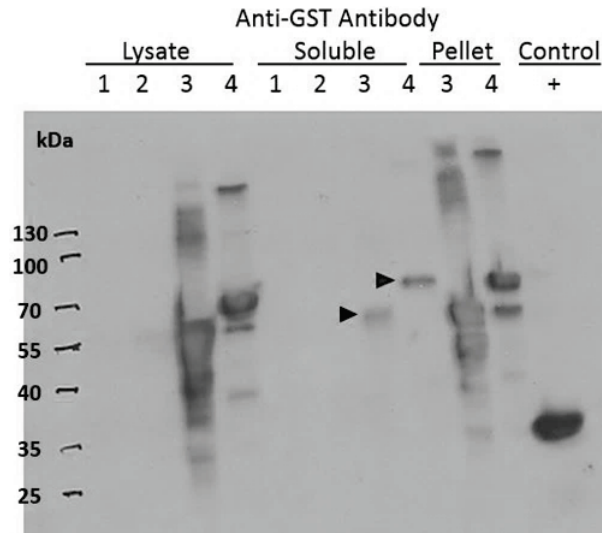


Figure 3-11 Western Blot with anti-GST-Antibody to Detect PRMTs.

1: GST-HsPRMT2-2 in DH5 α ; 2: GST-HsPRMT2-9 in DH5 α ; 3: GST-RnPRMT1-1 in BL21(DE3); 4: GST-RnPRMT1-1 in Rosetta(DE3)pLysS. The arrows indicate protein bands at the expected size in the soluble fraction samples No. 3 and 4. IRF4 was applied as a positive control (Last lane). Expected MWs are GST-HsPRMT2-2, 73.0 kDa; GST-HsPRMT2-9, 75.2 kDa; GST-RnPRMT1-1, 65.7 kDa. Protein ladder: PAGE Ruler Pre-stained (ThermoFisher).

3.4.1 Large Scale Purification of His₆-RnPRMT1-1

After lysis and sonication of a cell pellet from 0.9 L *E. coli* BL21(DE3) culture, the protein was purified using a pre-equilibrated Ni²⁺-NTA-column (Figure 3-12 A). The bound protein was eluted using binding buffer supplemented with 500 mM imidazole. RnPRMT1-containing elution fractions (E1, E2 and E3) were pooled. The His₆-tag was removed by incubation with 3C-protease and tagged and cleaved proteins were separated by subtractive Ni²⁺-IMAC (Figure 3-12 B). The band shift after removal of the His₆-tag is clearly visible (compare pre and post sample) and the size of the protein band is in good agreement with the expected molecular weight of 39.73 kDa. Most RnPRMT1-1 is in the flow-through fraction and wash fractions 1-3. Those fractions were combined, and the protein further purified using size exclusion chromatography (SEC).

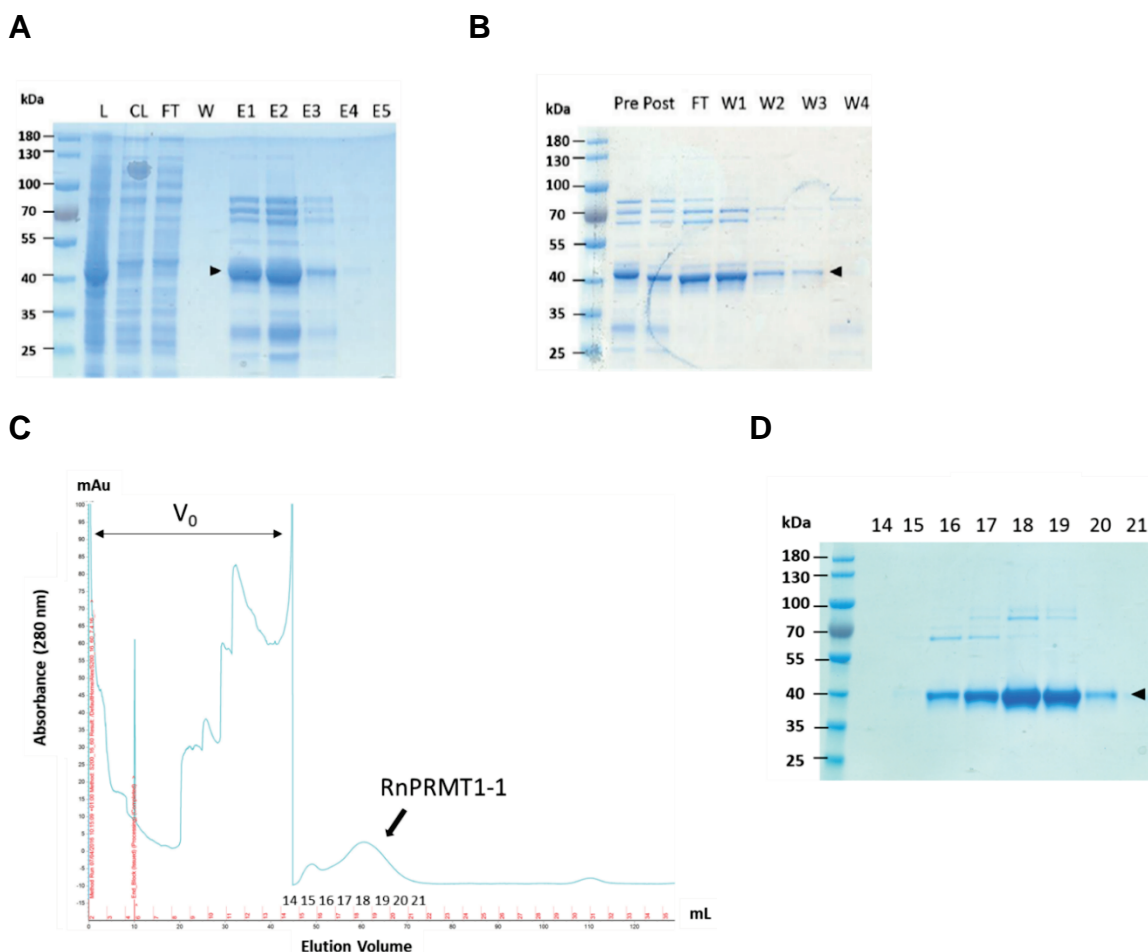


Figure 3-12 RnPRMT1-1 Purification.

A) Affinity chromatography: 10 μ L aliquots of different fractions were analysed by 10 % SDS-PAGE gels and stained with instant blue stain. The arrows indicate the location of the RnPRMT1-1 protein band which has an expected molecular weight of 39.73 kDa. **B)** Elution fractions E1-E3 were pooled (Pre) and the tag removed via 3C-protease treatment (Post). Tagged and untagged protein were separated via affinity chromatography. Most of the protein was in the flow-through and wash fraction 1-3. The fractions were pooled and further purified. FT: Flow through; W: Wash; E1-E5: Eluted fraction; L: Lysate; CL: Cleared Lysate. Protein ladder: PAGE Ruler Pre-stained (ThermoFisher). **C)** Chromatography profile: Cleaved RnPRMT1-1 was analysed by SEC using a HiLoad 16/60 Superdex 200 column pre-equilibrated with SEC buffer (600 mM NaCl, 10 mM imidazole, 40 mM HEPES pH 7.5, 1 mM TCEP, 5 % (v/v) glycerol, buffer) at a flow rate of 0.3 mL/min. The absorbance at 280 nm (blue line) was monitored. The large spikes at the beginning of the run, are most likely caused by the fact that the column was pre-equilibrated at room temperature and then moved to 4 $^{\circ}$ C. **D)** SDS-PAGE analysis of SEC fractions: The protein in 1 mL of each fraction was concentrated via StrataClean Resin and 10 μ L aliquots of each fraction sample were analysed by 10 % SDS-PAGE and stained with Instant Blue stain. The PRMT-containing fractions 18 and 19 were pooled and concentrated to 80 μ L at 1.1 mg/mL. The arrows indicate the RnPRMT1-1 protein band with an expected molecular weight of 39.73 kDa.

Prior to loading the sample was filtered and unfortunately, one third of the total protein was lost during this step. Large protein aggregates of His₆-RnPRMT1-1 have been reported in the literature and the most likely explanation for this loss is that such aggregates were retained on the filter membrane (Lakowski and Frankel, 2009). The

remaining protein was analysed by SEC using a HiLoad 16/60 Superdex 200 column (Figure 3-12 D). A peak corresponding to RnPRMT1 eluted at a volume of 54-68 mL/fraction 16-20. SDS-PAGE analysis (Figure 3-12 C) subsequently showed that the fractions corresponding to the void volume did not contain any RnPRMT1 suggesting that there were no higher order oligomers present at this stage of the purification process.

The void volume of the HiLoad 16/60 Superdex 200 column is 45 mL, the elution volume of the major peak containing RnPRMT1-1, was 60.6 mL. Using the online calibration tool (<http://proteincrystallography.org/protein-purification/superdex-200-calibration.php>) with the calibration curve V_e/V_o vs log MW an estimated molecular weight for RnPRMT1-1 of ~170 kDa was calculated. The predicted molecular mass for RnPRMT1-1 is 39.7 kDa. Comparing these values would suggest that the protein elutes as a tetramer. This is in good agreement with the literature, where it was reported that the protein can exist minimally as a dimer, and mostly as higher oligomers (Zhang and Cheng, 2003). The protein containing fractions (18-19) were pooled and concentrated to 1.1 mg/mL and further analysed by mass spectrometry.

The intact mass of the purified RnPRMT1-1 protein was confirmed by Jonathon Renshaw at AstraZeneca (Alderley Park, UK) using a Quadrupole Time-of-Flight (Q-TOF) LC/MS instrument (Agilent Technologies). The final measured mass was 39,736.44 Da, compared to the ProtParam tool calculated mass for the protein of 39,730.59 Da (Gasteiger *et al.*, 2005). The measured mass agrees with the predicted mass within the accuracy of the mass spectrometry experiment.

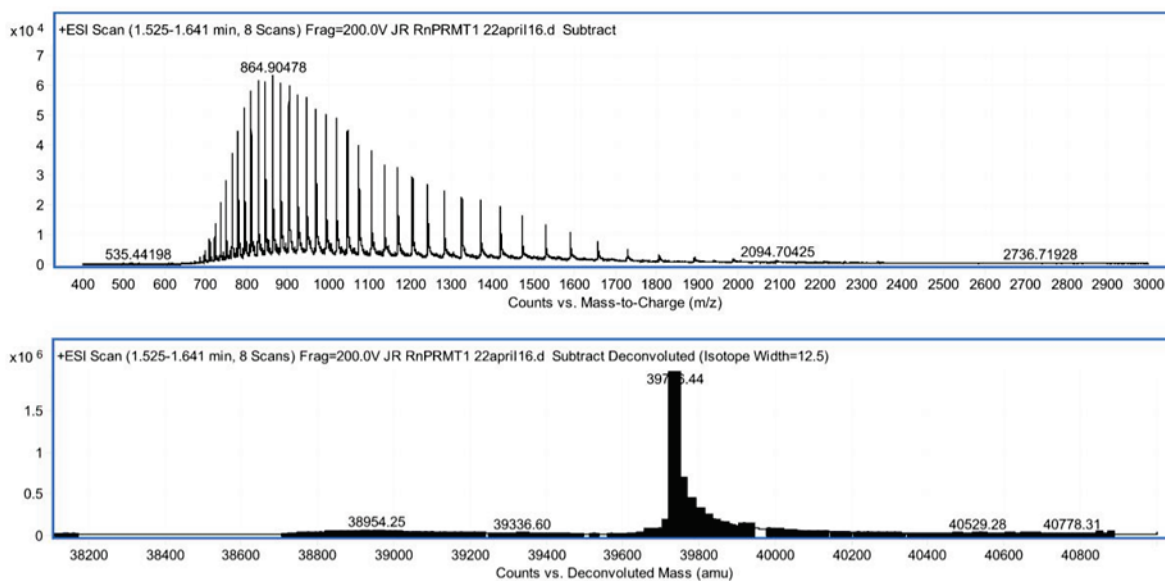


Figure 3-13 Mass Determination of Purified RnPRMT1-1 Sample by Q-TOF LC/MS.

The graph at the bottom shows the deconvoluted mass spectra in which all multiply charged species are recalculated into a single-charged form and clustered according to m/z value and peak width. The final measured intact mass of RnPRMT1-1 (predicted mass of 39,730.59 Da) is 39,736.44 Da.

3.4.2 Large Scale Purification of GST-RnPRMT1-1

In the next step GST-RnPRMT1-1 expressed in *E. coli* BL21(DE3) was purified in order to analyse if a GST-tag compared to a His₆-tag influences protein expression yields. Figure 3-14 shows the SDS-PAGE analysis of the gravity affinity purification of GST-RnPRMT1-1. Elution fractions E1-E5 show two bands close to 70 kDa. The expected size of the tagged protein is 68.2 kDa. To analyse if the lower band is the target protein the elution fractions were selected, pooled and subsequently a small fraction was subjected to protease cleavage of the GST-tag and analysed again by SDS-PAGE (Figure 3-14 B). From this analysis it becomes clear that the upper band is shifting to a lower molecular weight after tag removal and therefore it is very likely that this is the band corresponding to RnPRMT1-1. Compared to the other lower bands (most likely chaperone containing) the intensity of the band is very low. The total protein concentration estimated by SDS-PAGE was less than 1 mg RnPRMT1 which is very low compared to the His₆-tagged RnPRMT1-1, and therefore no further purification steps were conducted.

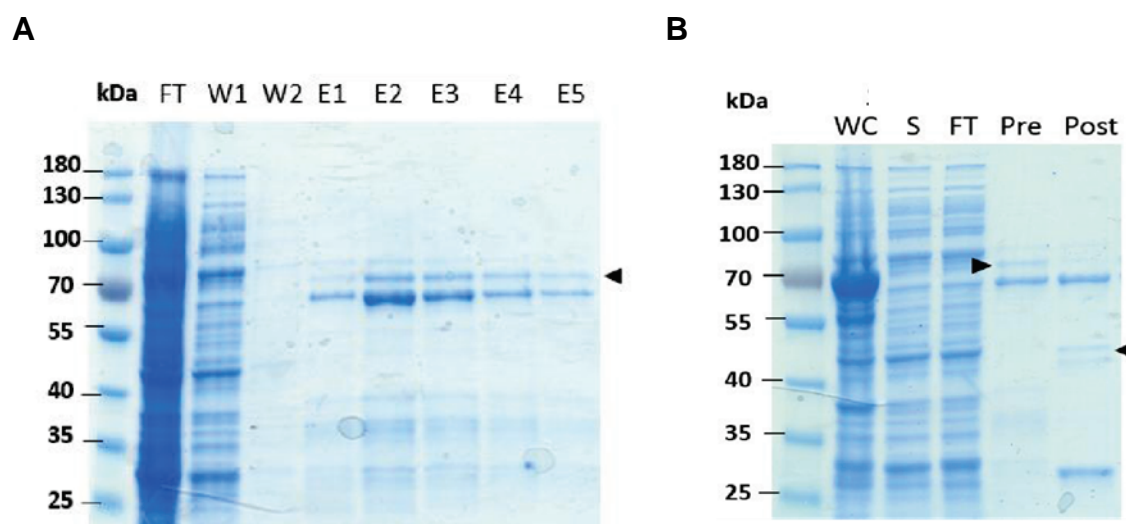


Figure 3-14 SDS-PAGE Analysis of the GST-RnPRMT1-1 Purification.

A) Affinity chromatography: 10 μ L aliquots of different fractions were analysed by 10 % SDS. The arrow indicated the band corresponding to GST-RnPRMT1-1. **B)** The Elution fractions E1-E5 were combined (Pre) and an aliquot was cleaved with 3C-protease overnight and run on the gel (Post). The arrows indicate the protein bands belonging to GST-RnPRMT1-1 (65.7 kDa) and RnPRMT1-1 (39.7 kDa) after tag removal. The bands at 70 kDa are most likely chaperonins. FT: Flow through; W: Wash; E1-E5: Eluted fraction; WC: Whole Cell; S: Soluble. The arrows indicate the protein bands belonging to GST-RnPRMT1-1 and RnPRMT1-1 after tag removal. Protein ladder: PAGERuler Pre-stained (ThermoFisher).

3.4.3 Analysing the Oligomeric State of PRMT1 Using Analytical SEC

Previously it has been reported that PRMT1 can exist as homo-dimer but also as oligomers with sizes reaching from 300 kDa up to 1000 kDa (Zhang and Cheng, 2003). The oligomeric state is the more active state of the enzyme compared to the homo-dimer (Feng *et al.*, 2011). It was also shown that the oligomerisation state of PRMT1 depends not only on protein concentration and temperature but also on the presence or absence of the cofactor SAM or product SAH. In the concentration range below 0.5 μ M, an increase in PRMT1 concentration increases the extent of oligomerisation and positively affects enzyme activity (Feng *et al.*, 2011). On the other hand the additional pre-incubation of the protein with SAM or SAH reduced the size of the PRMT1 oligomers from 500 kDa to 320 kDa (Zhang and Cheng, 2003). In order to analyse the oligomerisation state of our purified RnPRMT1-1 protein, 10 μ M PRMT1 samples were pre-incubated for 30 min on ice with either 25 μ M SAM or SAH, and then analysed by analytical Superdex 75 10/300 GL column. In the first step a calibration curve was prepared to allow the calculation of the molecular weight from the elution volume of the tested samples (Section 2.17.7, Appendix G). Analysis of the different PRMT1 samples

is shown in Figure 3-15. In contrast to literature reports, the addition of substrate or product had only a small effect on the PRMT1 oligomerisation state and a shift in peak onto a 200 kDa peak can be observed. Even without the addition of SAM or SAH, the predominant oligomerisation size of the protein is between 130 and 380 kDa.

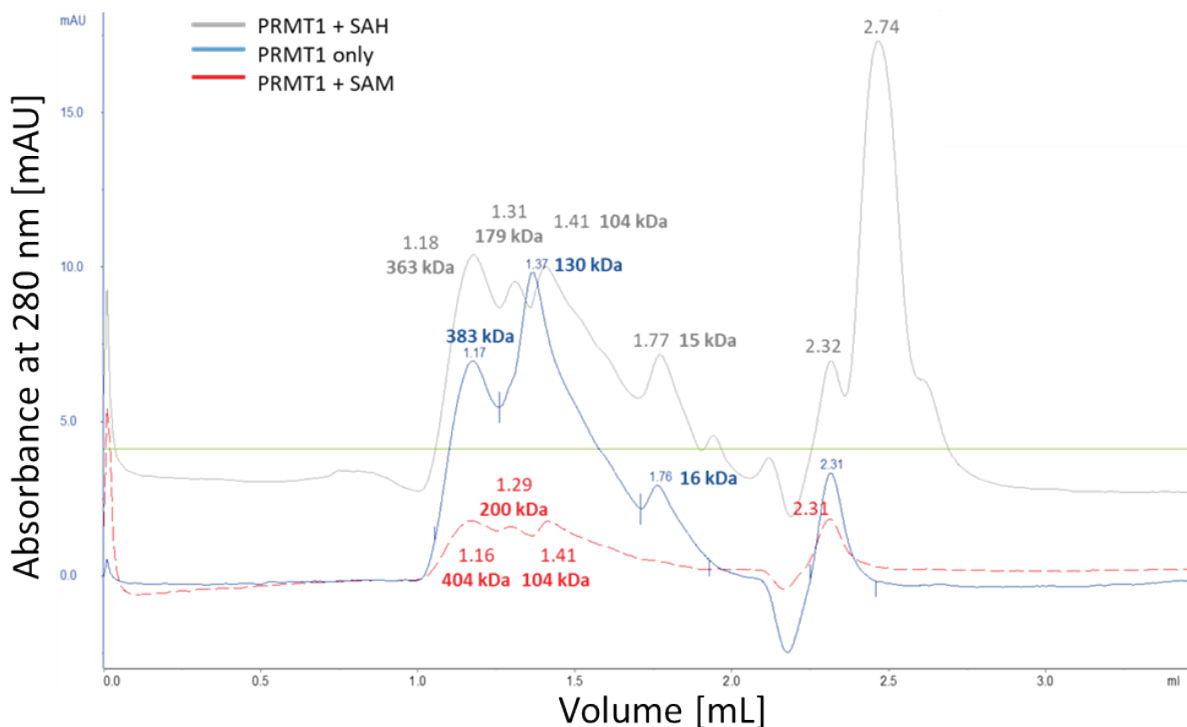


Figure 3-15 Oligomerisation State Analysis of RnPRMT1-1 upon Addition of Cofactor.

The oligomerisation state of RnPRMT1 was analysed using analytical SEC at room temperature. Different PRMT1 samples without cofactor (blue trace) and incubated with SAH (grey trace) or SAM (red trace) were run on a Superdex 75 10/300 GL column at a flow rate of 0.4 mL/min and room temperature. The molecular weight is shown above the corresponding chromatogram peaks in kDa.

3.4.4 Large Scale Purification of GST-HsPRMT2-9

In the next step, the remaining GST-HsPRMT2-9 pellet was purified because it showed some soluble expression in the large-scale expression tests (Figure 3-10 B). Cleared lysate from 0.7 L *E. coli* Rosetta(DE3)pLysS culture was incubated with 50 % (v/v) GSH-Sepharose slurry overnight at 4 °C and applied onto a gravity column. After elution, the PRMT2-containing fractions were identified using SDS-PAGE (Figure 3-16 A). Elution fractions E1-E5 were combined, and the GST-tag was removed by treatment with 3C-protease overnight at 4 °C.

SDS-PAGE analysis of the cleaved and tagged protein samples (Figure 3-16 B) showed that the band corresponding to GST-HsPRMT2-9 (75.2 kDa) is the faint band at ~90 kDa and not the strong band at ~70 kDa, which is most likely heat shock protein 70 (Hsp70). The cleaved protein was concentrated to 0.5 mL and loaded onto 10/300 Superdex 200 GL column. The chromatogram can be seen in Figure 3-16 D. PRMT-containing fractions were identified by SDS-PAGE (Figure 3-16 B). In most PRMT2 containing fractions, a band at ~70 kDa which most likely corresponds to the Hsp70 is visible. This indicates that HsPRMT2 is associated with chaperone proteins and might have a partially unfolded structure. Moreover, HsPRMT2-containing fractions eluted in two different peaks, at 10.6 mL which corresponds to a molecular weight of ~505 kDa, the other eluted at 14 mL corresponding to 105 kDa, which might be indicative of two potential oligomeric states of HsPRMT2. The calibration curve for the 10/300 Superdex 200 column can be found in Appendix F. The HsPRMT2-9 concentration was very low, and as a result the whole fraction was used for SDS-PAGE analysis.

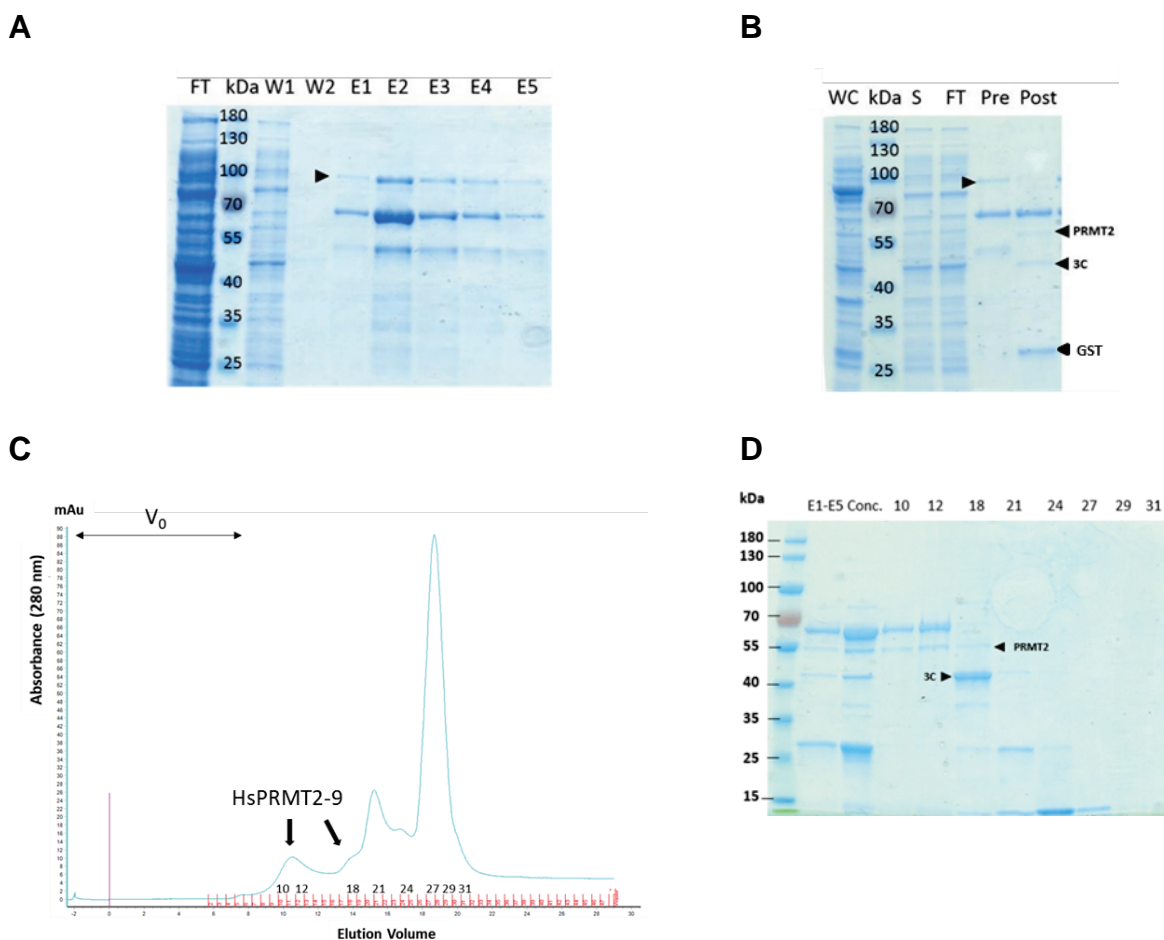


Figure 3-16 SDS-PAGE Analysis of GST-HsPRMT2-9 Purification.

A) Affinity chromatography: 10 μ L aliquots of different fractions were analysed by 10 % SDS. The arrow indicated the band corresponding to GST-HsPRMT2-9. The Elution fractions E1-E5 were combined (Pre) and an aliquot was cleaved with 3C-protease overnight and **B)** run on the gel (Post). The arrows indicate the GST-HsPRMT2-9, HsPRMT2-9 after tag removal, and 3C-protease. The bands at 70 kDa are most likely chaperones. FT: Flow through; W: Wash; E1-E5: Eluted fraction; WC: Whole Cell; S: Soluble. **C)** Chromatography profile: HsPRMT2-9 was analysed on a 10/300 Superdex 200 column pre-equilibrated with SEC buffer (300 mM NaCl, 40 mM HEPES pH 7.5, 20 mM imidazole, 1 mM TCEP, 5 % (v/v) glycerol) with a flow rate of 0.3 mL/min. The absorbance at 280 nm (blue line) was monitored. The PRMT-containing fractions were identified by SDS-PAGE. **D)** SDS-PAGE analysis of SEC fractions: Each fraction (0.5 mL) was concentrated via StrataClean Resin and 10 μ L aliquots of each fraction sample were analysed by 10 % SDS-PAGE. The upper arrow indicates the band which most likely corresponds to the 49.2 kDa target protein PRMT2. The strongly stained band below it is most likely 3C-protease, the band close to 25 kDa, GST. Protein ladder: PAGE Ruler Pre-stained (ThermoFisher).

3.5 Changing the Expression System from *E. coli* cells to the MultiBac™ System

Recombinant protein expression in *E. coli* cells did not yield sufficient amounts of soluble protein despite testing different protein constructs across a variety of strains and expression conditions. Therefore, the expression system was changed from *E. coli* cells to insect cell expression using the Bac-to-Bac Baculovirus Expression System (Invitrogen Life Technologies).

Baculoviruses are insect viruses that carry a large double stranded DNA molecule that can be edited to encode additional foreign genes. The gene encoding the polyhedron protein is expressed at high levels during the late stage of the viral replication cycle but is not needed if the virus is replicated in cultured insect cells. Thus, it can be replaced with the target gene, which will then be under the control of the polyhedrin promoter. The Bac-to-Bac Expression system generates recombinant baculoviruses with the help of the *E. coli* strain DH10Bac (Ciccarone et al., 1997, Luckow et al., 1993). A donor plasmid (pFastBac1 vector), in which the gene of interest is cloned via site-specific transposition, is transfected into the *E. coli* host strain DH10Bac that carries a bacmid baculovirus shuttle vector. Bacterial colonies which carry the recombinant bacmid are identified via antibiotic selection and white-blue screening. The isolated bacmid DNA is then transfected into *Sf9* or *Sf21* insect cells.

The principle of the expression system is shown in Figure 3-17. The genes of interest for this project were cloned into a pFastBac1 vector, in which the genes are under the control of the strong *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV PH) promoter. The vector carries a gentamycin resistance gene, the bacterial target site-specific transposon Tn7 and a Simian Virus 40 (SV40) polyadenylation signal which allows the formation of Mini-Tn7 transposons. The pFastBac1 plasmid is transformed into DH10Bac cells that contain a bacmid carrying a target site for attachment of the bacterial transposon (mini-attTn7), a Kanamycin resistance gene, and a gene encoding the LacZ α peptide from a pUC-based cloning vector. The LacZ α peptide together with the LacZ Ω peptide, which is encoded by the bacterial chromosome, forms the functional enzyme β -Galactosidase (Schneemann and Young, 2003).

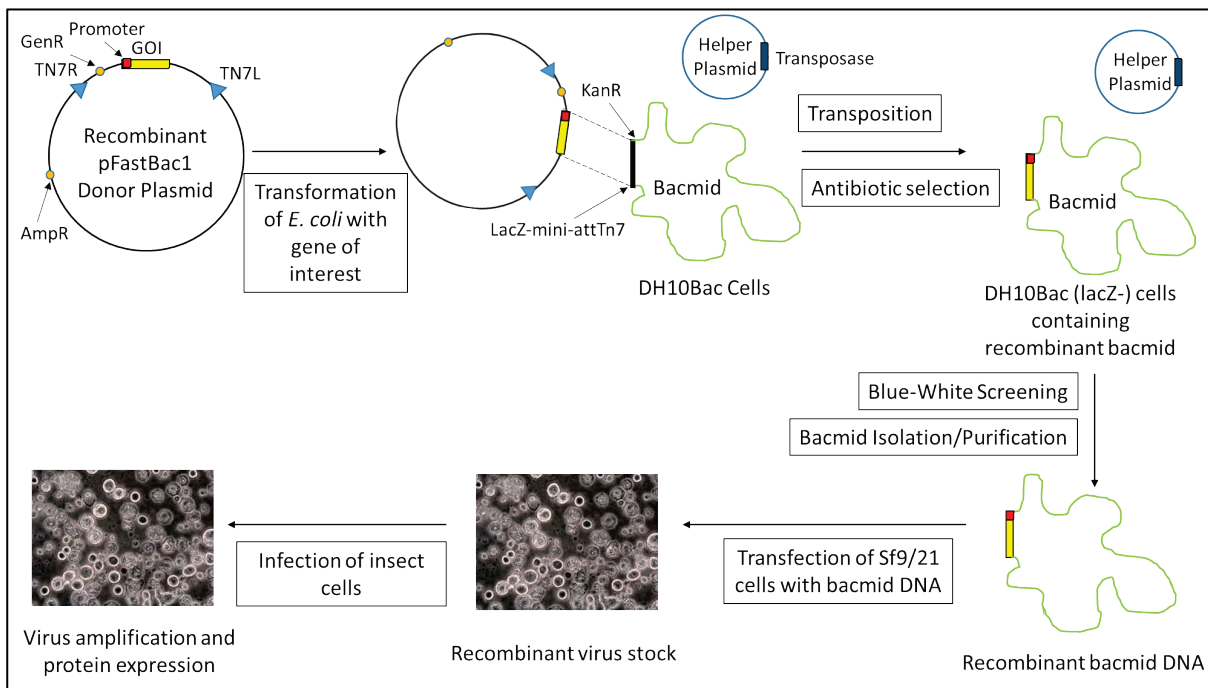


Figure 3-17 Principle of the Bac-to-Bac Baculovirus Expression System.

The bacterial cells contain a helper plasmid, that carries a tetracycline resistance gene and encodes the transposase. After transformation, a recombinant bacmid is generated with the help of the transposase which enables the transposition between the Mini-Tn7 elements of the pFastBac1 vector and the target site on the bacmid. The insertion disrupts the expression of the LacZ α peptide allowing identification of cells containing the recombinant bacmid via blue-white screening (Section 2.9.2). Next, the bacmid DNA is isolated and transfected into insect cells. A recombinant baculovirus is generated, amplified and used to infect insect cells for recombinant protein expression (Chapter 2.9.4 and 2.9.5). In addition to the GST- and His₆-tags, a N-terminal His₁₀-Maltose-binding protein (MBP) tag was chosen because it was shown that GST and MBP can increase protein solubility (Esposito and Chatterjee, 2006).

3.5.1 Small Scale Expression Tests in Sf9 Cells

Small-Scale expression tests in *Sf9* cells were performed as described in Section 2.10. Tested constructs included His₁₀-MBP-HsPRMT2-6 (82.8/39.7 kDa), His₁₀-MBP-HsPRMT2-8 (80.6/37.5 kDa), His₁₀-MBP-MmPRMT2-1 (93.7/50.6 kDa), and His₆-MmPRMT2-1 (55.8/ 50.6 kDa). Each construct was tested in duplicate, and samples were analysed for soluble expression via spin column purification (Section 2.13). After

elution from the affinity resin, the samples were analysed by SDS-PAGE before and after incubation with 3C-protease. The SDS-PAGE analysis can be seen in Figure 3-18. The black arrows indicate bands of the expected molecular weight that shifted after tag removal. In particular, for His₁₀-MBP-HsPRMT2-6 and His₆-MmPRMT2-1 a strong band was visible. They were both expressed at larger scale. Unfortunately, no expression was seen for MmPRMT2-1, but HsPRMT2-6 expression looked more promising, and the purification is described in Section 2.16.7.

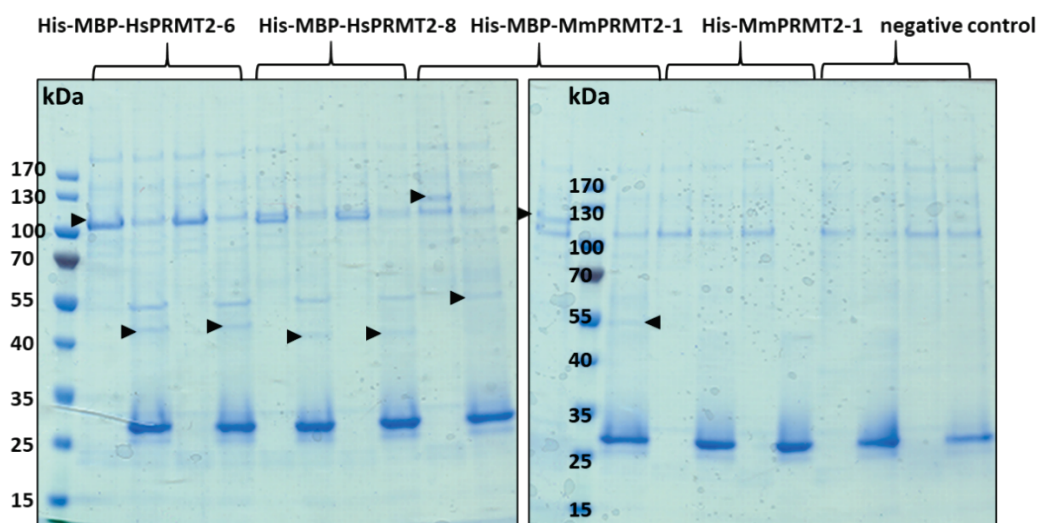


Figure 3-18 Small-Scale Expression Test Analysis in Sf9 Cells.

Small scale expression tests of different viral stocks were performed in Sf9 cells. Cells were grown for 48 h at 27 °C. After purification, one untreated and one elution sample incubated with 3C protease were analysed by SDS-PAGE. Expected MWs are: His₁₀-MBP-HsPRMT2-6: 82.8/ 39.7 kDa; His₁₀-MBP-HsPRMT2-8: 80.6/ 37.5 kDa, His₁₀-MBP-MmPRMT2-1: 93.7/ 50.6 kDa; His₆-MmPRMT2-1: 55.8/ 50.6 kDa. Protein ladder: PAGERuler Pre-stained (ThermoFisher).

Additionally, His-HsPRMT2-1 and His-HsCARM1-1 were directly tested at a larger scale (300 mL Sf9 cells) as described in Section 2.15. The cleared lysate was purified using Ni²⁺-IMAC (Figure 3-19 Gel A 1). In both cases, bands of the expected molecular weight are visible. Elution fractions E1-E3 were combined, dialysed into low imidazole buffer and the tag removed with His-3C-protease overnight. A small shift of the bands corresponding to PRMT2-1 and CARM1-1 is visible in the post sample after tag removal (Gel B). In the final step, the samples were applied to Ni²⁺-IMAC column to separate cleaved and tagged proteins. Protein containing fractions (FT, W1 and W2) were pooled and concentrated.

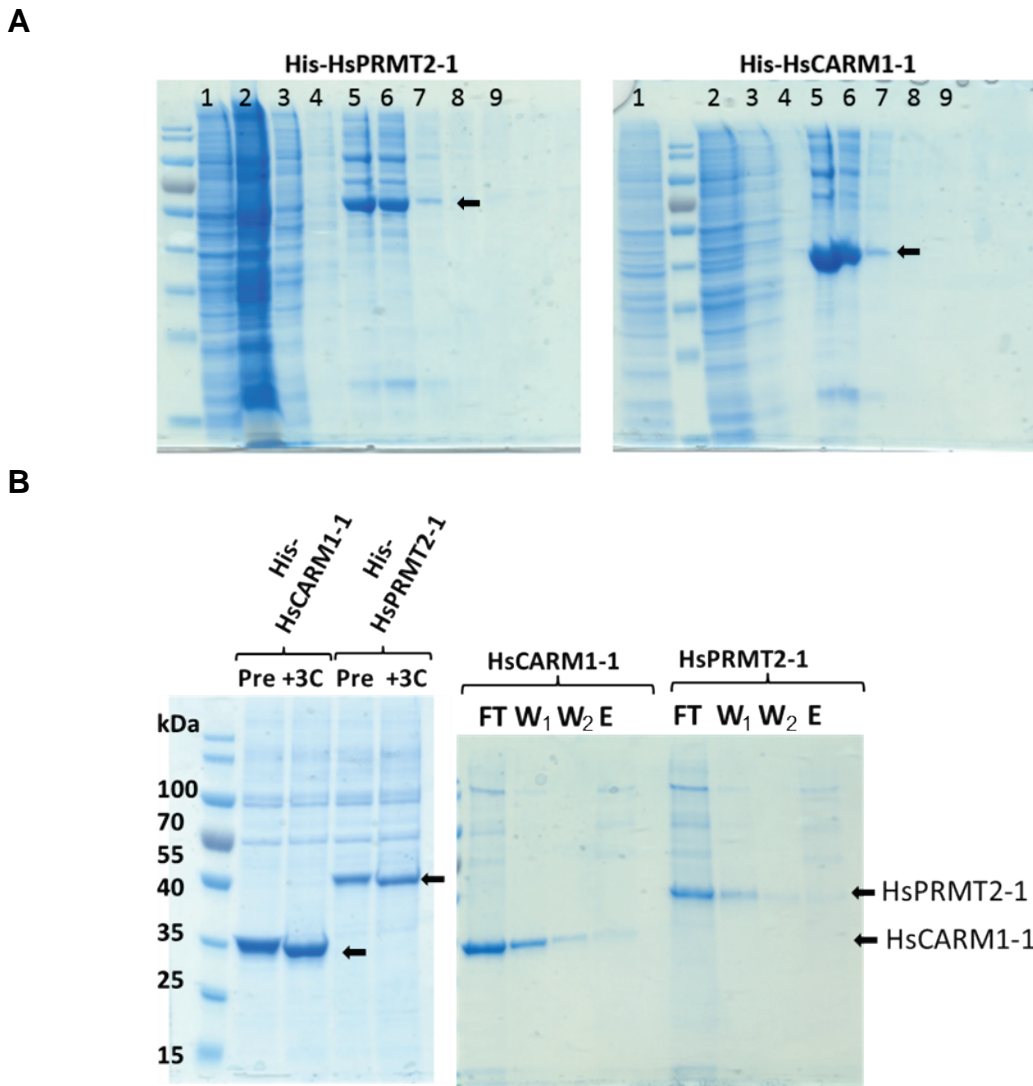


Figure 3-19 Large Scale Expression Test in Sf9 Cells.

Large scale expression tests of His-HsPRMT2-1 and His-CARM1-1 were performed in Sf9 cells. Cells were grown for 48 h at 27 °C. **A)** SDS-PAGE Analysis of Ni²⁺-IMAC. Fractions: 1: Whole cell; 2: Soluble fraction; 3: Flow-through, 4: Wash; 5-9: Elution 1-5. **B)** Gel 1: Pre and Post 3C- protease tag removal. Gel 2: Subtractive Ni²⁺-IMAC Analysis. Flow-through and Wash fractions containing the protein were combined and concentrated. Expected MWs are: His₆-HsPRMT2-1: 49.2/43.3 kDa; His₆-HsCARM1-1: 46.4/ 39.8 kDa. Protein ladder: PAGERuler Pre-stained (ThermoFisher).

The final yield was 0.50 mg HsPRMT2-1 and 0.46 mg HsCARM1-1 from 300 mL cell culture. Intact mass for both proteins was confirmed by MS (Appendix I). Analysis was performed by Dr Jonathan Renshaw (AstraZeneca, UK). Based on the positive results, it was decided to express His₆-HsPRMT2-1 at larger scale as described in Section 2.16.8.

3.5.2 Small Scale Expression Tests in Sf21 Cells

A small-scale expression test with 11 different viral stocks was performed in Sf21 cells as described in Section 2.11. Tested constructs included N-terminal His₁₀-MBP-, His₆- and GST-tagged RnPRMT1-1, HsPRMT2-1 and HsCARM1-1 and His₁₀-MBP- and His₆-tagged HsPRMT2-2. Each construct was tested in duplicate, and samples were analysed for soluble expression via PhyNexus automated purification (Section 2.14). After elution from the affinity resin, the samples were analysed by SDS-PAGE (Figure 3-20).

Protein bands with the correct molecular weight were only seen in the case of His₆-CARM1 (Figure 3-20, A lane 4-6, 46.4 kDa) and His₁₀-MBP-tagged CARM1 (Figure 3-20, B lane 6-8, 82.9 kDa) as indicated by a red arrow. No protein expression was detected for His₆-RnPRMT1-1 (Figure 3-20, A lane 7-9) or His₁₀-MBP-RnPRMT1-1 (Figure 3-20, A lane 1-3). Furthermore, none of the GST-tagged proteins were detectable by SDS-PAGE (Figure 3-20, D). There was a faint band at the expected molecular weight of His₆-HsPRMT2-2 (Figure 3-20, B lane 1 and 2). Based upon these results, it was decided to express His₆-HsPRMT2-2 and His₁₀-MBP-CARM1 at a large scale. Unfortunately, no His₆-HsPRMT2-2 expression was detected. The results of the CARM1 purification will be discussed in the next section.

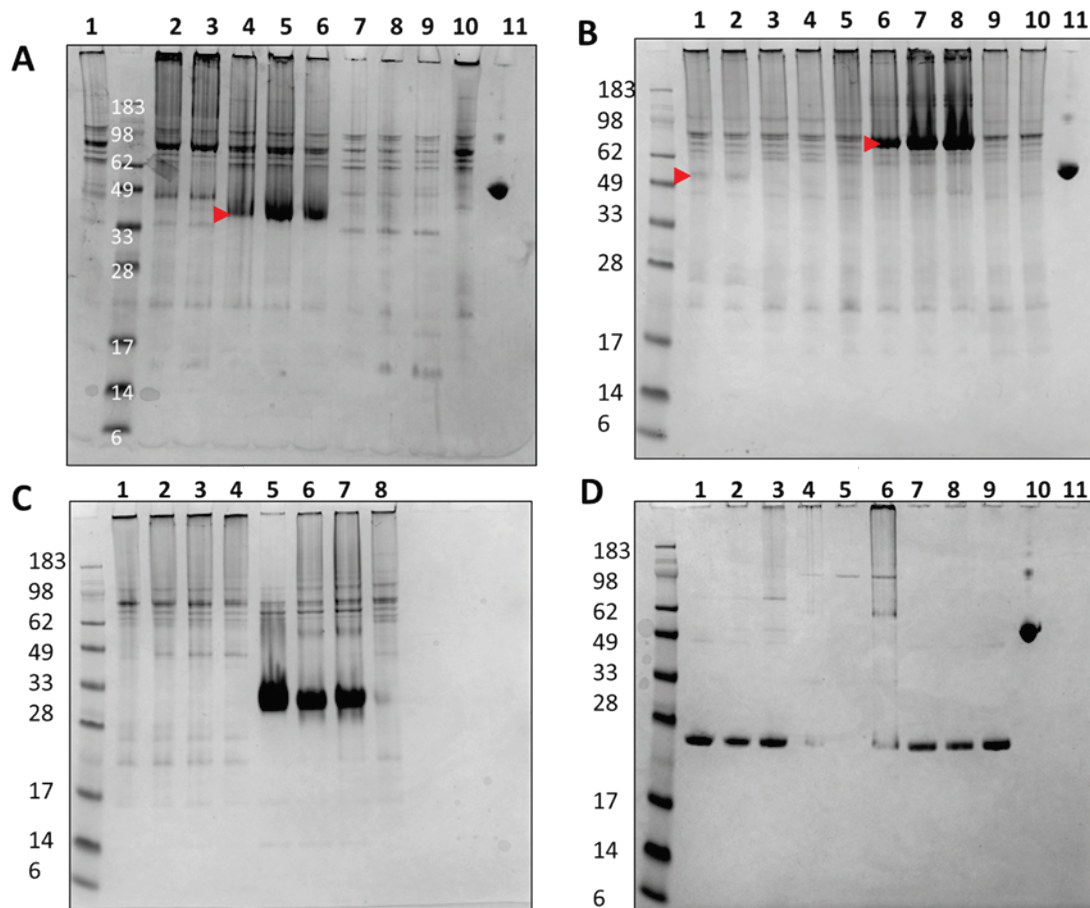


Figure 3-20 SDS-PAGE Analysis of Protein Expression Tests in Sf21 Cells.

Small scale expression tests of different viral stocks (V1) were performed in Sf21 cells. Cells were grown for 48 h at 27 °C and 140 rpm. After purification, the elution samples (15 μ L) were analysed by SDS-PAGE. Gel **A**) Lane 1-3: His₁₀-MBP-RnPRMT1-1 (82.8 kDa) 100/200/400 μ L; Lane 4-6: His₆-HsCARM1-1 (46.4 kDa) 100/200/400 μ L V1; Lane 7-9: His₆-RnPRMT1-1 (46.3 kDa) 100/200/400 μ L; Lane 10: His₆-HsPRMT2-2 (53.6 kDa) 100 μ L; Lane 12: BSA (1 μ g). Gel **B**) Lane 1-2 His₁₀-HsPRMT2-2 (53.6 kDa) 200/ 400 μ L; Lane 3-5: His₁₀-MBP-HsPRMT2-1 (92.3 kDa) 100/200/400 μ L; Lane 6-8: His₁₀-MBP-HsCARM1-1 (82.9 kDa) 100/200/400 μ L; Lane 9/10: His₆-HsPRMT2-1 (55.8 kDa) 100/200 μ L, Lane 11: BSA (1 μ g). Gel **C**) Lane 1: His₆-HsPRMT2-1 (55.8 kDa), 400 μ L; Lane 2-4: His₁₀-MBP-HsPRMT2-2 (90.1 kDa) 100/200/400 μ L; Lane 5-7: positive control (31 kDa); Lane 8: negative control (no virus). Gel **D**) Lane 1-3: GST-HsPRMT2-1 (75.2 kDa) 100/200/400 μ L; Lane 4-6: GST-RnPRMT1-1 (65.7 kDa) 100/200/400 μ L; Lane 7-9: GST-HsCARM1-1 (65.8 kDa) 100/200/400 μ L; Lane 10: BSA (1 μ g); Lane 11: negative control.

3.5.3 Large Scale Purification of His₁₀-MBP-HsCARM1

The His₁₀-MBP-tagged CARM1 expression construct yielded the highest amounts of soluble protein. Fresh V2 virus was generated as described (Section 2.9) and 1 L of Sf21 cells were infected. A 3 mL sample was taken prior to harvest and analysed for soluble expression via PhyNexus automated purification (Section 2.14). SDS-PAGE analysis of the sample (Figure 3-21) demonstrated significant protein expression at the expected mass of His₁₀-MBP-CARM1 (82.9 kDa).

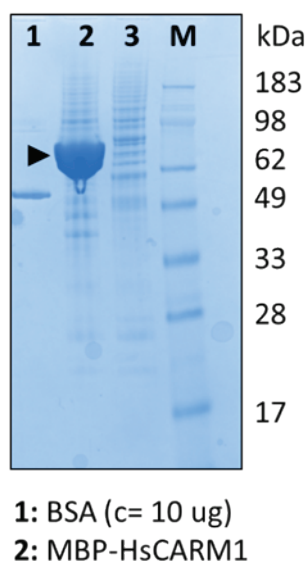


Figure 3-21 Small Scale Expression Test of Large Scale MBP-CARM1-1.

After elution of His₁₀-MBP-CARM1 from the nickel resin, the eluted protein was analysed by SDS-PAGE (Lane 2). BSA was used as a positive control. The gel was stained using InstantBlue. Lane 1, BSA (c 10 µg); lane 2, MBP-HsCARM1-1.

An exemplary His₁₀-MBP-CARM1 purification can be seen in Figure 3-22. First, the cleared cell lysate was incubated with TALON resin, followed by purification via immobilised nickel-affinity chromatography and imidazole elution. After His₁₀-MBP-CARM1 elution, the His₁₀-MBP tag was removed with 3C-protease treatment and the protein dialysed into low imidazole buffer. Subtractive nickel IMAC was used to separate the tags and any excess uncleaved protein from the cleaved CARM1. However, the HisTrap column was not able to remove the His₁₀-MBP impurities (Figure 3-22 A). Later analysis of the intact mass of the co-eluting protein via mass spectrometry confirmed that it is MBP from which the His₆-tag had been removed. MBP was finally successfully separated from CARM1 by size exclusion chromatography (Figure 3-22 A). The purified CARM1 was >95 % pure and stable as judged by SDS-PAGE. The amount

of CARM1 were sufficient for crystallisation trials (8.45 mg/500 mL of Sf21 cell culture) and fragment-screening assays.

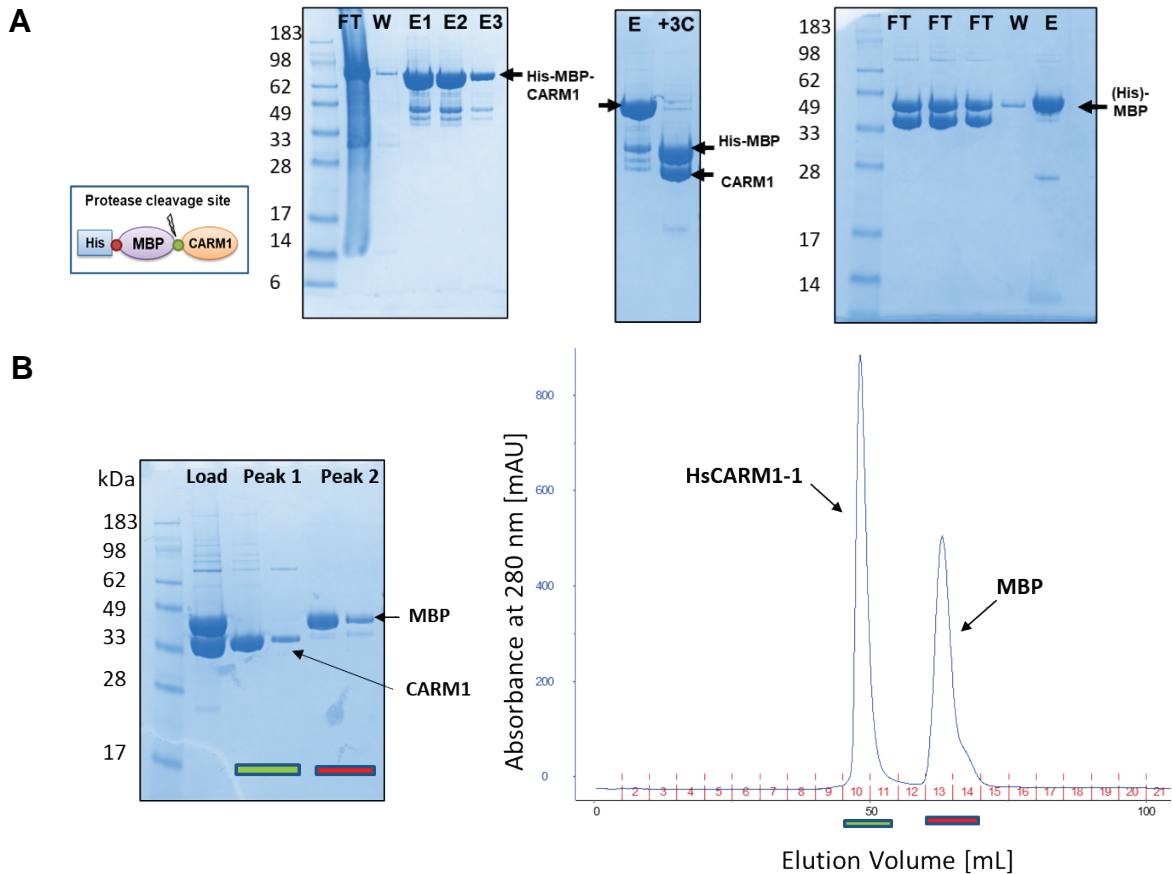


Figure 3-22 Example Purification of His₁₀-MBP-tagged Human CARM1 from Sf21 Cells.

After incubation of the cleared lysate with TALON resin and IMAC purification, the fusion protein was cleaved and dialysed into low imidazole buffer. Separation of the cleaved protein from uncleaved protein and His₁₀-MBP tags was attempted by subtractive Nickel-IMAC. However, another SEC step was needed for separation of CARM1 and MBP. The absorbance at 280 nm (blue line) was monitored. Purity was assessed after every step by SDS-PAGE analysis. The gels were stained with Instant Blue Stain. Protein ladder: PAGERuler Pre-stained (ThermoFisher).

3.5.4 Large Scale Purification of His₆-HsPRMT2-1

His₆-HsPRMT2-1 (51.2 kDa) was expressed in 1.2 L of *Sf9* cells and lysed via sonication (Section 2.16.1) and purified as described in Section 2.16.8. In the first step, the cleared lysate was applied onto a Talon column and flow-through, wash, and elution fractions were collected and analysed for expression. The SDS-PAGE analysis can be seen in Figure 3-23 A. A strong band is visible in the first elution fraction which is in the right molecular weight range near 55 kDa. The elution fractions were combined, and the His-tag was removed with His-3C-protease cleavage and the protein dialysed into a low imidazole buffer. Two bands are visible, one at 100 kDa and one near 55 kDa. The later corresponds to HsPRMT2-1 (49.2 kDa) and a shift after tag removal is clearly visible (Gel 2, Figure 3-23 A). The His-tags and the protease were removed by a subtractive Ni²⁺-IMAC step (Gel 3, Figure 3-23 A). Bands corresponding to HsPRMT2-1 are clearly visible in the flow-through and wash fraction. Both were combined, concentrated, and applied to a Superdex 200 ag 10/300 column. The chromatogram and SDS-PAGE gel are shown in Figure 3-23 B. PRMT2 eluted as a symmetrical peak. Beside the bands corresponding to HsPRMT2-1, a few high-molecular weight bands are visible. Fractions A10-B1 were pooled (orange bar), concentrated to 5.5 mg/mL. The final yield was 0.83 mg/L cell culture.

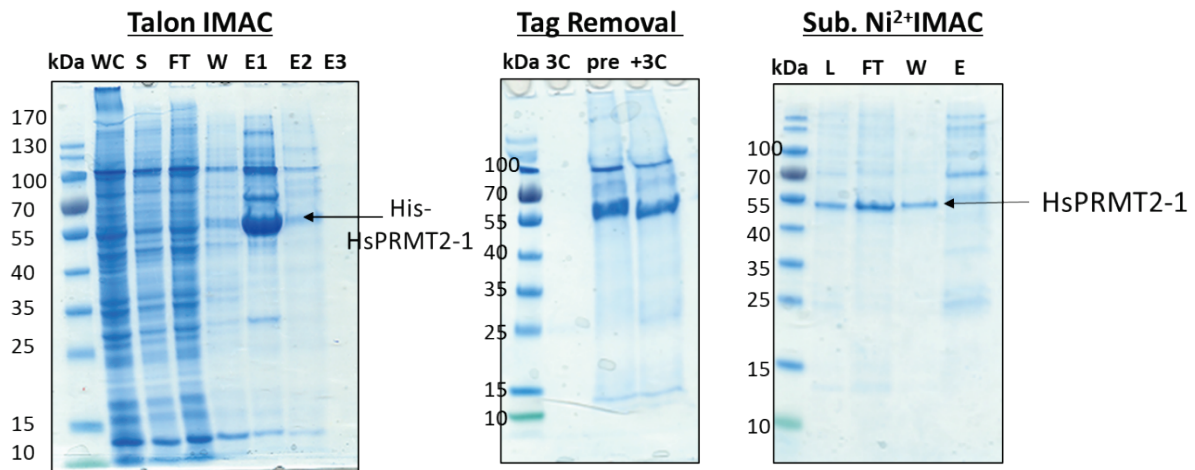
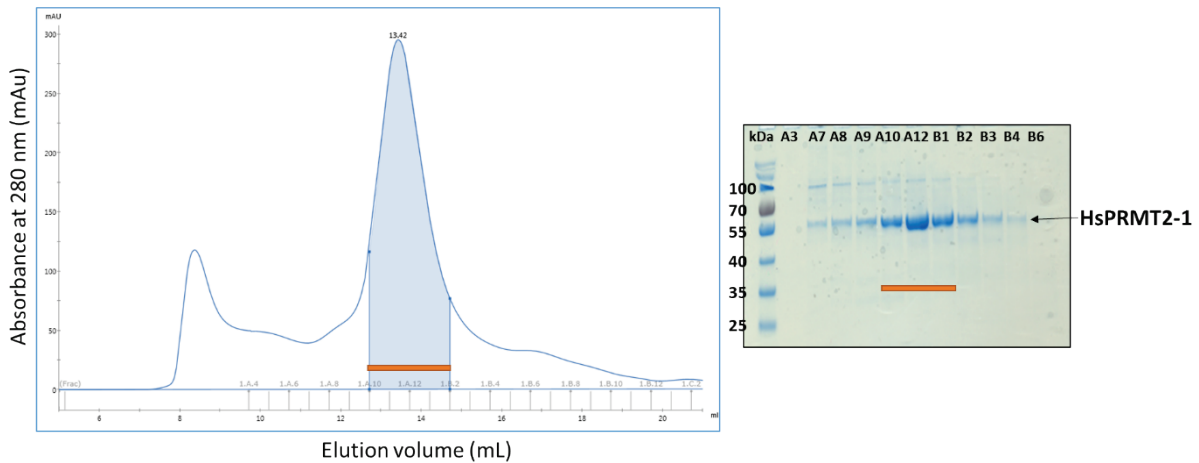
A**B**

Figure 3-23 Example Purification of His₆-tagged Human HsPRMT2-1 from Sf9 Cells.

A) After incubation of the cleared lysate with TALON resin and IMAC purification, the His-tag was removed via 3C-protease treatment and the protein dialysed into low imidazole buffer. Separation of the cleaved from uncleaved protein and His₆-tags was conducted by subtractive Nickel-IMAC. **B)** SEC: The sample was loaded onto a Superdex 200 ag 10/300 column. The chromatogram and SDS-PAGE gel are shown. The absorbance at 280 nm (blue line) was monitored. HsPRMT2-1 eluted as a single peak, fractions A10-B1 (orange bar) were pooled and concentrated. Purity was assessed after every step by SDS-PAGE analysis. Protein ladder: PAGERuler Pre-stained (ThermoFisher). The gels were stained with Instant Blue Stain.

3.5.5 Large Scale Purification of His₁₀-MBP-HsPRMT2-6

His₁₀-MBP-HsPRMT2-6 was expressed in 2.5 L of Sf9 cells as described in Section 2.16.7. After cell lysis, the cleared lysate was applied onto a 5 mL Talon column, and flow-through, wash and elution fractions were collected. The SDS-PAGE analysis of the affinity purification is shown in Figure 3-24 A. The elution fractions were combined, and the tag was cleaved off using His₆-3C-protease while dialysing the protein in a low imidazole buffer. Next, subtractive Talon IMAC was conducted to remove the tag and 3C-protease. The SDS gel is shown in Figure 3-24 A. After tag removal two bands are visible in the post cleavage sample, the upper band is most likely the His₁₀-MBP-tag which has a molecular weight of 43.1 kDa and the lower band HsPRMT2-6 (39.7 kDa). It had been observed during the CARM1 purification that the His₁₀-MBP-tag was not able to bind to Talon resin. The flow-through and wash were concentrated to 2 mL and loaded onto a Superdex 200 16/60 column. The SDS-PAGE analysis is shown in Figure 3-24 B. Unfortunately, it was not possible to separate the protein from the MBP-tag and both bands can be seen in most of the elution fractions (blue bar). Fractions containing HsPRMT2-6 were pooled, diluted with buffer to a salt concentration of 50 mM, and loaded onto a Mono Q 5/50 GL column. The IEX chromatogram is shown in Figure 3-24 C. HsPRMT2-6 eluted as a symmetric single peak, fraction 36 (orange bar) was concentrated to 3.6 mg/mL, aliquoted and stored at -80 °C.

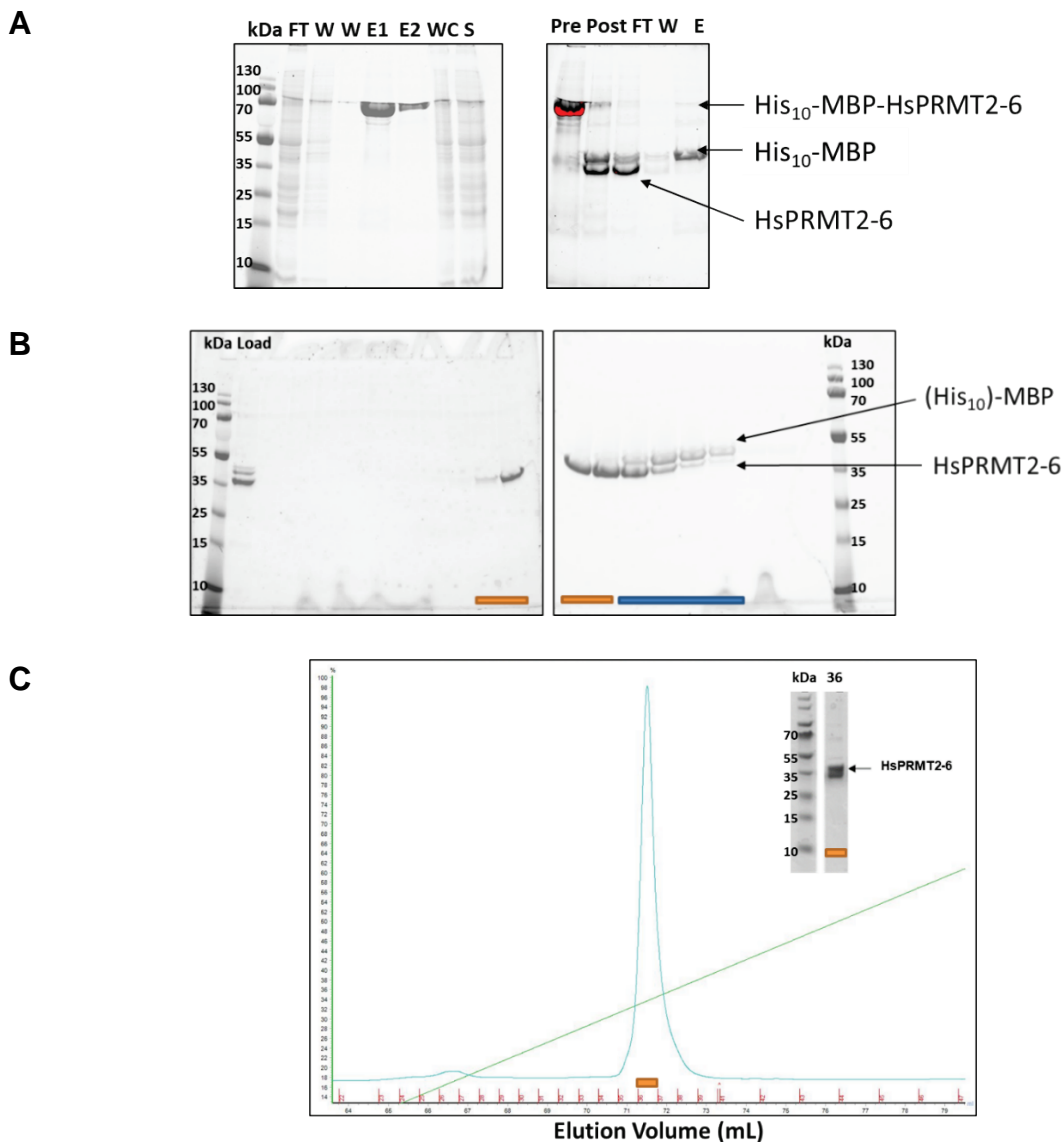


Figure 3-24 Example Purification of His₁₀-MBP-tagged HsPRMT2-6 from Sf9 Cells.

A) After TALON IMAC purification, elution fractions E1-2 were combined, and the fusion protein was cleaved with His-3C-protease and dialysed into low imidazole buffer. Separation of the cleaved protein (39.7 kDa) from uncleaved protein (82.8 kDa) and His₁₀-MBP tags (43.1 kDa) was attempted by subtractive TALON-IMAC, but MBP was not separated from the sample

B) FT and Wash were loaded onto a Superdex 200 16/60 column and elution fractions analysed by SDS-PAGE. However, another purification step was needed for separation of HsPRMT2-6 and MBP.

C) IEX: The sample was loaded onto a Mono Q 5/50 GL column, the protein eluted as a single peak and Fraction 36 (orange bar) was concentrated to 3.6 mg/mL. The absorbance at 280 nm (blue line) was monitored. Purity was assessed after every step by SDS-PAGE analysis. The gels were stained with Instant Blue Stain. Protein ladder: PAG-ERuler Pre-stained (ThermoFisher).

3.6 Limited Proteolysis to optimise HsPRMT2 Construct Design

Limited proteolysis can be a useful tool to identify protein sequences that show high backbone flexibility or are highly disordered regions (Fontana *et al.*, 2004). Proteolysis mainly occurs at flexible loops or disordered regions (Fontana *et al.*, 1986). Limited proteolysis with human PRMT2-1 was performed to identify possible domain boundaries, disordered regions, or flexible loops. The results of this analysis were used to design new protein expression constructs in which those regions were removed and that could result in better expression yields due to increased protein stability.

The human PRMT2 protein was incubated with different proteases which cleave at specific peptide bonds to create protein fragments. Proteases used were endoproteinase Asp-N from *Pseudomonas fragi* that cleaves N-terminally to aspartic and glutamic acid residues and chymotrypsin from bovine pancreas that cuts C-terminally to lysine and arginine residues. A different range of protease:PRMT2 ratios, incubation times and temperatures were tested in order to identify stable fragments that cannot be further cleaved because they have been reduced to a compactly folded domain (Figure 3-25).

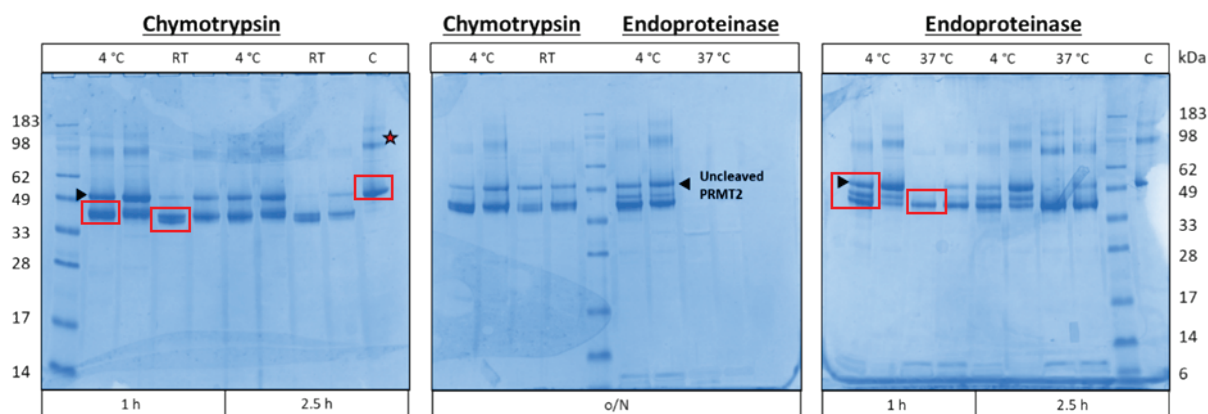


Figure 3-25 Limited Proteolysis and Resulting Proteolytic Fragments of HsPRMT2-1.

SDS-PAGE of human HsPRMT2-1 that was digested into different stable fragments using Chymotrypsin and Endoproteinase. Different temperatures and incubation times were tested. The red boxes indicate fragments that were identified via in-gel digestion, followed by peptide mapping of the cleavage products. Black arrows indicate the uncleaved protein.

The samples were further analysed by mass spectrometry to identify the cleavage sites. Following chymotrypsin digestion, a PRMT2 fragment comprising residues 44-433 was produced after 1 h incubation at 4 °C (Gel 1, Lane 2). If incubated at RT

an even shorter fragment of residues 64-433 was visible (Gel 1, Lane 4). Three different fragments were produced following digestion with endoproteinase for 1 h. At 4 °C two different fragments, spanning residues 42-418 (Gel 3, lower band, Lane 1) and 42-433 (Gel 3, upper band, Lane 1) were identified. At 37 °C, a shorter fragment encompassing residues 75-418 (Gel 3, Lane 3) was produced. The upper band that was seen in the control sample (indicated by a red star) was identified as full-length HsPRMT2. The results indicate that proteolysis mostly occurred within the N-terminal region where the SH3 domain (residues 30-89) is located. Additional cleavage occurred within the C-terminal sequence. However, no internal cleavage sites could be identified.

This analysis indicates that the core of the protein is most likely properly folded and structural ordered. The smallest expression construct that was already tested for protein expression is even shorter than the smallest identified fragment and comprises residues 88-414. Adding some sequence to the shortest expression construct could have been tested to see whether this could increase stability and thus enable crystallisation. However, this was not done as the first full-length MmPRMT2 crystal structure was published and it was decided to focus on this specific construct (Cura *et al.*, 2017).

3.7 Buffer Optimization for HsPRMT2 and RnPRMT1 using Thermal Melt Assays

Buffer screens were performed using thermal denaturation assays. Different buffers, pHs, and ionic strength were screened to find the optimal buffer compositions for RnPRMT1-1 and HsPRMT2-1. The assay principle can be seen in Figure 3-26.

Optimising buffer conditions can potentially increase yields due to higher protein stability. In total, 45 different buffer and salt combinations were tested plus the control buffer. The block layout and buffer composition can be seen in Appendix C.

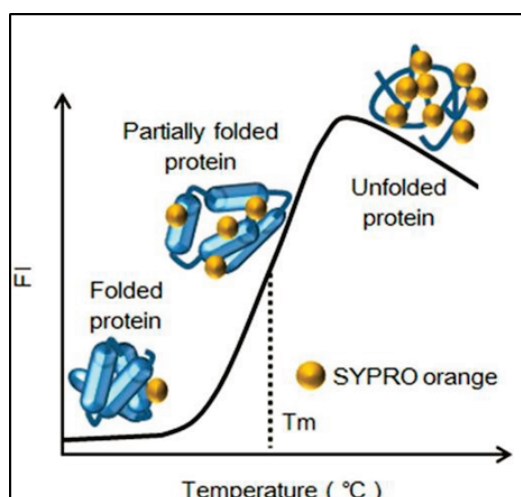


Figure 3-26 Principle of Thermal Denaturation Assay.

A thermal denaturation assay can be used to measure the thermal stability of a protein under different conditions. The fluorescence of the protein-binding dye SYPRO Orange is quenched in aqueous environments. However, upon binding to hydrophobic protein core regions which are exposed when the protein unfolds, it undergoes a significant increase in quantum yield (λ_{ex} 470 nm / λ_{em} 570 nm). The SYPRO Orange fluorescence emission can be monitored while increasing the temperature to obtain the melting temperature of the protein (T_m).

The T_m and thus stability of RnPRMT1-1 was greatly improved in MOPS buffer, whereas different NaCl concentrations had only little effect on its thermostability. An exemplar melting curve is shown in Figure 3-27. The upper panel shows the RnPRMT1 melting curves in MOPS, pH7.2 buffer with either 0, 100 mM, or 300 mM NaCl. The lower panel shows the melting curves of the protein in its original buffer. Switching to MOPS buffer pH 7.2 increased the T_m by 18 °C to 70 °C. Changing the buffer to PIPES, pH 6.8, 300 mM NaCl increased the melting temperature by 22 °C from 52 to 74 °C.

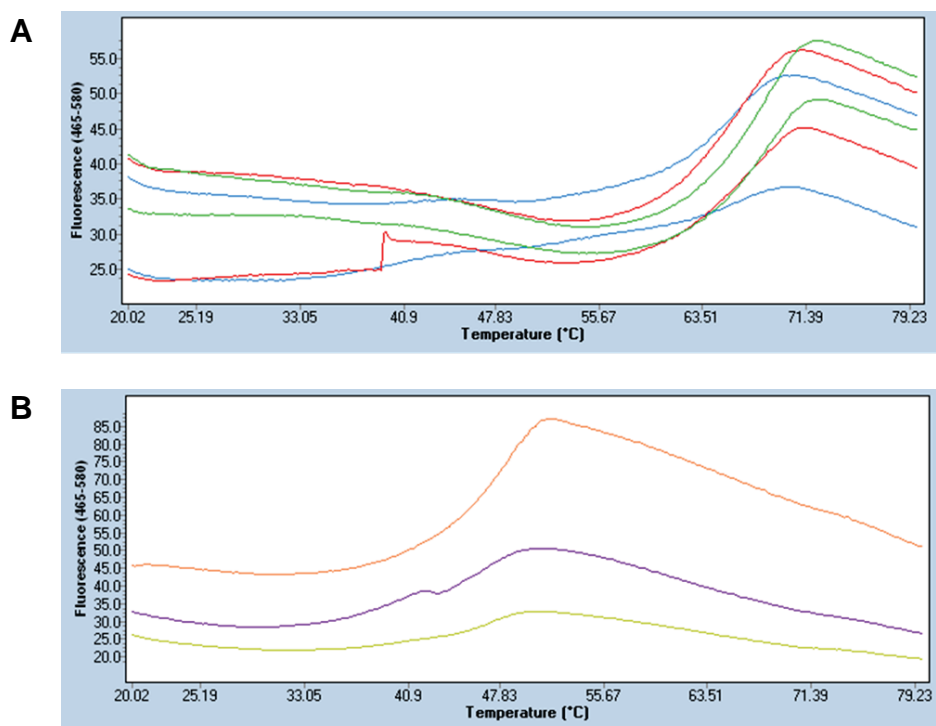


Figure 3-27 Exemplary Thermal Denaturation Assay of RnPRMT1-1.

The raw fluorescence of RnPRMT1 in different buffers was measured as a function of temperature to identify buffer conditions that increase its stability. A) Denaturation curve using MOPS pH 7.2 buffer with 0 (blue), 100 mM (red) and 300 mM (green) NaCl. B) RnPRMT1 was tested at different concentrations (10, 15, 20 μ M) in the original RnPRMT1 purification buffer (40 mM HEPES, pH 8.0, 600 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol).

Table 3-3 summarizes the buffers that stabilised RnPRMT1-1 and increased the melting temperature. The results show that the protein is most stable in a pH range from 6.1- 7.2. The original buffer had a pH of 8.0, so it would be useful to lower the pH in the future. For RnPRMT1, one transition is observed, which indicates a single unfolding event.

Table 3-3 Summary of Buffers that Increased RnPRMT1 Stability as Determined Using a Thermal Denaturation Assay in Comparison with its Original Buffer.

Buffer Composition	T_m¹ (°C)	ΔT_m¹
MOPS, pH 7.2, 0 mM NaCl	70	18
MOPS, pH 7.2, 100 mM NaCl	71	19
MOPS, pH 7.2, 300 mM NaCl	72	20
PIPES, pH 6.8, 0 mM NaCl	71.5	19.5
PIPES, pH 6.8, 100 mM NaCl	73.5	21.5
PIPES, pH 6.8, 300 mM NaCl	74	22
BES, pH 7.0, 0 mM NaCl	71	19
BES, pH 7.0, 100 mM NaCl	73	21
BES, pH 7.0, 300 mM NaCl	65	13
MES, pH 6.1, 0 mM NaCl	69	17
MES, pH 6.1, 100 mM NaCl	71	19
MES, pH 6.1, 300 mM NaCl	72	20

¹The melting temperature (T_m) and the increase in T_m (ΔT_m) in comparison with its original purification buffer is shown.

In the case of HsPRMT2-1, two transition states were visible in most buffer conditions (Figure 3-28). It could be speculated that this is caused by two domains that have different thermal stabilities and unfold at different time points, such as the SH3 and PRMT domains. This observation would be in good agreement with the results from the limited proteolysis of HsPRMT2-1 that showed that the N-terminal SH3 domain is less stable or has greater flexibility than the rest of the protein (Section 3.6). Another possibility would be that the protein exists in different oligomeric states and that for example the monomer and dimer denature at different temperatures or that an impurity is present. In order to compare the effect of different buffers on thermal stability, the second melting temperature was used. In the previously used buffer, the T_m (2) was 52 °C. The melting temperature for mouse PRMT2 that was reported in literature is 53 °C, which is in good agreement with the measured T_m (2) values for human PRMT2 (Cura et al., 2017).

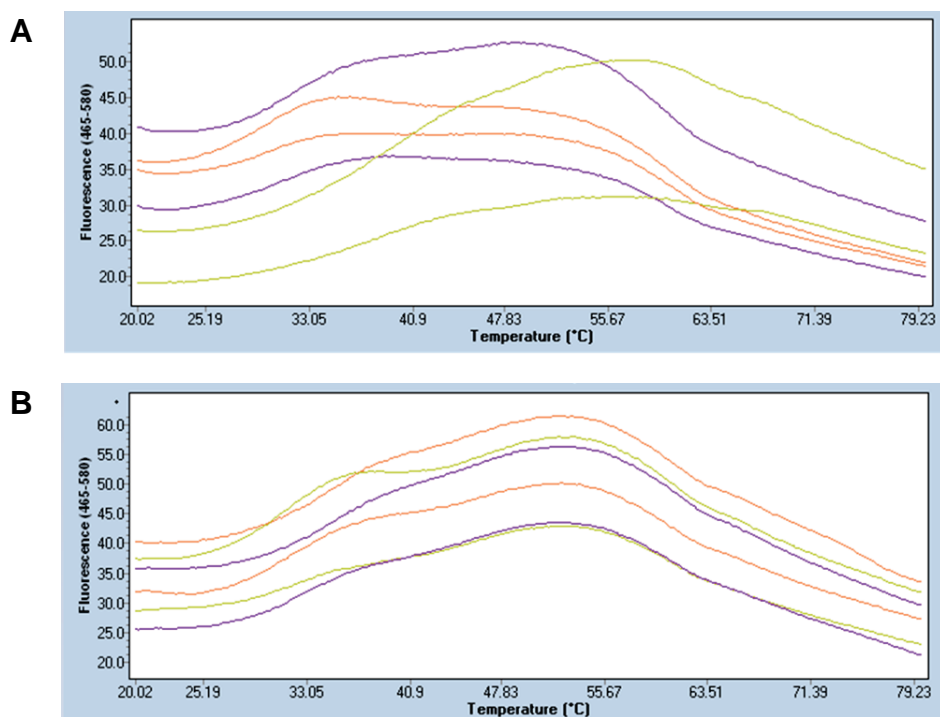


Figure 3-28 Exemplary Thermal Denaturation Assay of HsPRMT2-1.

The raw fluorescence of HsPRMT2-1 in different buffers was measured as a function of temperature to identify buffer conditions that increase its stability. A) Denaturation curve using BES pH 7.0 buffer with 0 mM (red), 100 mM (purple) and 300 mM (green) NaCl. B) HsPRMT2-1 was tested in the original purification buffer (40 mM HEPES pH 8.0, 0.5 mM TCEP and 5 % (v/v) glycerol 600 mM NaCl).

None of the tested buffers had a large effect on the stability of HsPRMT2-1. Only two buffers increased the melting temperature slightly (Table 3-4). The pH range in which HsPRMT2-1 seems to be the most stable spans from pH 7.0- 8.1. The current buffer is in this range with a pH of 8.0. However, the protein did not seem to be very stable in any of the tested buffers. The next step would be to test other buffer compositions with different additives that might stabilize the protein further, for example glycerol (Vagenende *et al.*, 2009). Unfortunately, protein yields did not allow for further buffer screens.

Table 3-4 Summary of Buffers that increased HsPRMT2-1 Stability in Thermal Denaturation Assay.

Buffer Composition	T_m (2) ¹ (°C)	ΔT_m
BES, pH 7.0, 300 mM NaCl	58	6
Tris, pH 8.1, 300 mM NaCl	57	5

¹The melting temperature (T_m) and the increase in T_m (ΔT_m) in comparison with its original purification buffer is shown.

3.8 Purification of GST-MmPRMT2-10

The first high-resolution crystal structure of mouse and zebrafish PRMT2 was published by Cura and colleagues (Cura *et al.*, 2017). The mouse sequence of the crystallised protein (UniProtKB: Q3UKX1) differs from our mouse protein sequence (UniProtKB: Q9R144), they share a sequence identity of 90 %. The sequence alignment can be seen in Figure 3-29. The differences are all located within the C-terminal domain. Additionally, the crystallised construct (PDB: 5FUL) carried a point mutation at the last amino acid residue (R445W).

Analysis of the two different MmPRMT2 full-length gene sequences (GenBank BC122563.1 and AF169620.1) is shown in Appendix E. The sequences have 99.3 % sequence identity, with nine DNA base differences and two gaps, which resulted in the protein sequence differences and different protein length due to different stop codons. At the start of the project, we choose the gene sequence AF169620.1 and not BC122563.1 as it is reviewed by UniProtKB/Swiss-Prot.

Q3UKX1	Q3UKX1_MOUSE	1	MEAPGEGPCSESQVIVPVEEDPVDYGCEMQLLQDGAQLQLQQLQPEEFVAIADYTATDETQ	60
Q9R144	ANM2_MOUSE	1	MEAPGEGPCSESQVIVPVEEDPVDYGCEMQLLQDGAQLQLQQLQPEEFVAIADYTATDETQ	60

Q3UKX1	Q3UKX1_MOUSE	61	LSFLRGEKILILRQTTADWWWGERAGCCGYIPANHLGKQLEEDPEDTWQDEEYFDSYGT	120
Q9R144	ANM2_MOUSE	61	LSFLRGEKILILRQTTADWWWGERAGCCGYIPANHLGKQLEEDPEDTWQDEEYFDSYGT	120

Q3UKX1	Q3UKX1_MOUSE	121	LKLHLEMLADQPRRTTKYHSVILQNKESLKDQVILVDVGCCTGIIISLFCAHHARPKAVYAVE	180
Q9R144	ANM2_MOUSE	121	LKLHLEMLADQPRRTTKYHSVILQNKESLKDQVILVDVGCCTGIIISLFCAHHARPKAVYAVE	180

Q3UKX1	Q3UKX1_MOUSE	181	ASDMAQHTSQLVLQNGFADTIITVFQQKVEDVVLPEKVDVVLVSEWMTCLLFEFMIESILY	240
Q9R144	ANM2_MOUSE	181	ASDMAQHTSQLVLQNGFADTIITVFQQKVEDVVLPEKVDVVLVSEWMTCLLFEFMIESILY	240

Q3UKX1	Q3UKX1_MOUSE	241	ARDTWLKGDGIIWPTTAALHLVPCSAEKDYHDKVLFWDNAYEFNLSALKSLAIKEFFSRP	300
Q9R144	ANM2_MOUSE	241	ARDTWLKGDGIIWPTTAALHLVPCSAEKDYHDKVLFWDNAYEFNLSALKSLAIKEFFSRP	300

Q3UKX1	Q3UKX1_MOUSE	301	KSNHILKPEDCLSEPCTIILQDMRTVQVPDLETMRGELRFDIQKAGTLHGFTAWFSVYFQ	360
Q9R144	ANM2_MOUSE	301	KSNHILKPEDCLSEPCTIILQDMRTVQVPDLETMRGELRFDIQKAGTLHGFTAWFSVYFQ	360

Q3UKX1	Q3UKX1_MOUSE	361	SLEEGQPQVVLSTGPLHPHTHWKQTLFMMDDPVPVHTGDVVTGSSVVLQRNPVWRR-HMSV	419
Q9R144	ANM2_MOUSE	361	SLEEGQPQVVLSTGPLHPHTHWKQTLFMMDDPVPVHTGDVVTGSSVVLQRNPVWRR-HMSV	420
			*****;***** * : : : * : :	
Q3UKX1	Q3UKX1_MOUSE	420	SL-----SWVVSALDPTSQRVGEKVFPINR	445
Q9R144	ANM2_MOUSE	421	SELGCHVRTRSHVSTELTGSFRSGGDS-----	448
			* : : * : * * *	

Figure 3-29 MmPRMT2 Protein Sequence Alignment.

Both Mouse PRMT2 sequences (UniProtKB: Q9R144 and Q3UKX1) were aligned using the Clustal Omega program (Sievers *et al.*, 2011). The overall sequence identity is 90 %. Asterisks show positions with a single, fully conserved amino acid residue, colons high sequence conservation between groups, a period low sequence conservation. The red square highlights the Arg445 residue, which is mutated to a tryptophan in the literature construct that was successfully crystallised (PDB: 5FUL).

In the next step, the MmPRMT2 wild-type sequence and the sequence carrying the point mutation were synthesised from Eurofins Genomics and both were cloned into three different pACEBac expression vectors (described in Section 2.4.2 and 2.4.5) to yield N-terminal His₆-, His₁₀-MBP- and GST-tagged expression constructs (Table 3-5).

Small-scale expression tests were performed in duplicate in *Sf9* cells (Section 2.10). The SDS-PAGE analysis of the elution fractions before and after 3C-protease tag cleavage did not look very promising and only faint bands were visible in some fractions. However, to rule out that the yields on a small-scale experiment were too little to be visible by SDS-PAGE gel, the literature construct MmPRMT2-10 which was purified with a GST-tag was expressed in large-scale (1 L). The large-scale expression of GST-MmPRMT2-10 was successful and will be described in the following section.

Table 3-5 New Mouse PRMT2 Expression Constructs for Insect Cell Expression.

Name	Residue Number	N-terminal Tag	Notes
MmPRMT2-9	1-445	His ₆ -	
MmPRMT2-9	1-445	His ₁₀ -MBP-	
MmPRMT2-9	1-445	GST-	
MmPRMT2-10	1-445	His ₆ -	R445W
MmPRMT2-10	1-445	His ₁₀ -MBP-	R445W
MmPRMT2-10	1-445	GST-	R445W

GST-MmPRMT2-10 was expressed in 3 L of *Sf9* cells, the cells were lysed by sonication and purified as described in Section 2.9 and 2.16.4. The cleared lysate was incubated with GST-resin and applied onto a gravity column, the flow-through, wash, and elution was then collected. The SDS-PAGE analysis of the affinity purification is shown in Figure 3-30. A strong band above 70 kDa is visible in the elution fraction corresponding to GST-MmPRMT2-10 (76.7 kDa).

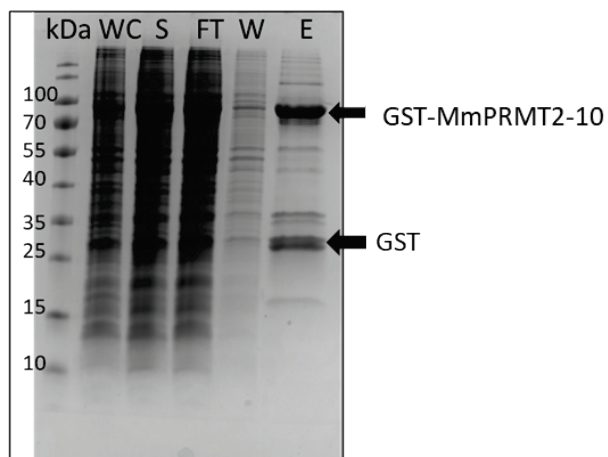


Figure 3-30 SDS-PAGE Analysis of GST-MmPRMT2-10 Expressed in Sf9 cells and Affinity Purified.

The arrows indicate the bands in the elution fractions corresponding to GST-MmPRMT2-10 (MW= 76.7 kDa) and GST (28 kDa). The gels were stained with Instant Blue Stain. Protein ladder: PAGERuler Pre-stained (ThermoFisher).

Next, the GST-tag was removed via His-3C-protease cleavage. The protein was concentrated to 2.2 mg/mL and loaded onto a Superdex 200 16/60 column to separate the His-3C-protease and the GST-tags from the protein. The chromatogram and the SDS-PAGE analysis of the pre- and post-cleavage sample and the elution fractions can be seen in Figure 3-31 A and B. The band shift after tag removal is clearly visible (B, Lane 1 and 2). Unfortunately, it was not possible to separate the GST-tags and the other impurities from the protein MmPRMT2-10 (50.7 kDa) were visible in most of the elution fractions of the four elution peaks but in the first elution fractions (green bar) high-molecular weight impurities are visible and in the last peak it co-elutes with GST (orange bar).

Fractions A2-C3 (green bar) and C4-D11 (orange bar) were pooled. The latter was loaded onto a GST-column to remove the GST-tags. The SDS-PAGE analysis after this purification step can be seen in Figure 3-31 C. The GST-tag was successfully removed from the fractions. The samples were combined, concentrated to 5 mL and diluted with IEX buffer without salt to lower the final NaCl concentration to 50 mM. The sample was then applied onto an IEX column and eluted with a NaCl gradient from 50-500 mM. The chromatogram and corresponding SDS-PAGE gel can be seen in Figure 3-32. A small peak containing MmPRMT2-10 eluted at the beginning at low NaCl concentrations (green bar), the rest eluted in a larger peak at ~83 mM NaCl (orange bar).

MmPRMT2-10 is very clean after the purification step, only small high-molecular impurities near 70 kDa are visible at low salt concentrations. The purest fractions A10-C4 and C5-D7 were pooled and each concentrated to 3.5 mg/mL and used for crystallisation trials. The other fractions A1-A9 were concentrated to 2 mg/mL and frozen.

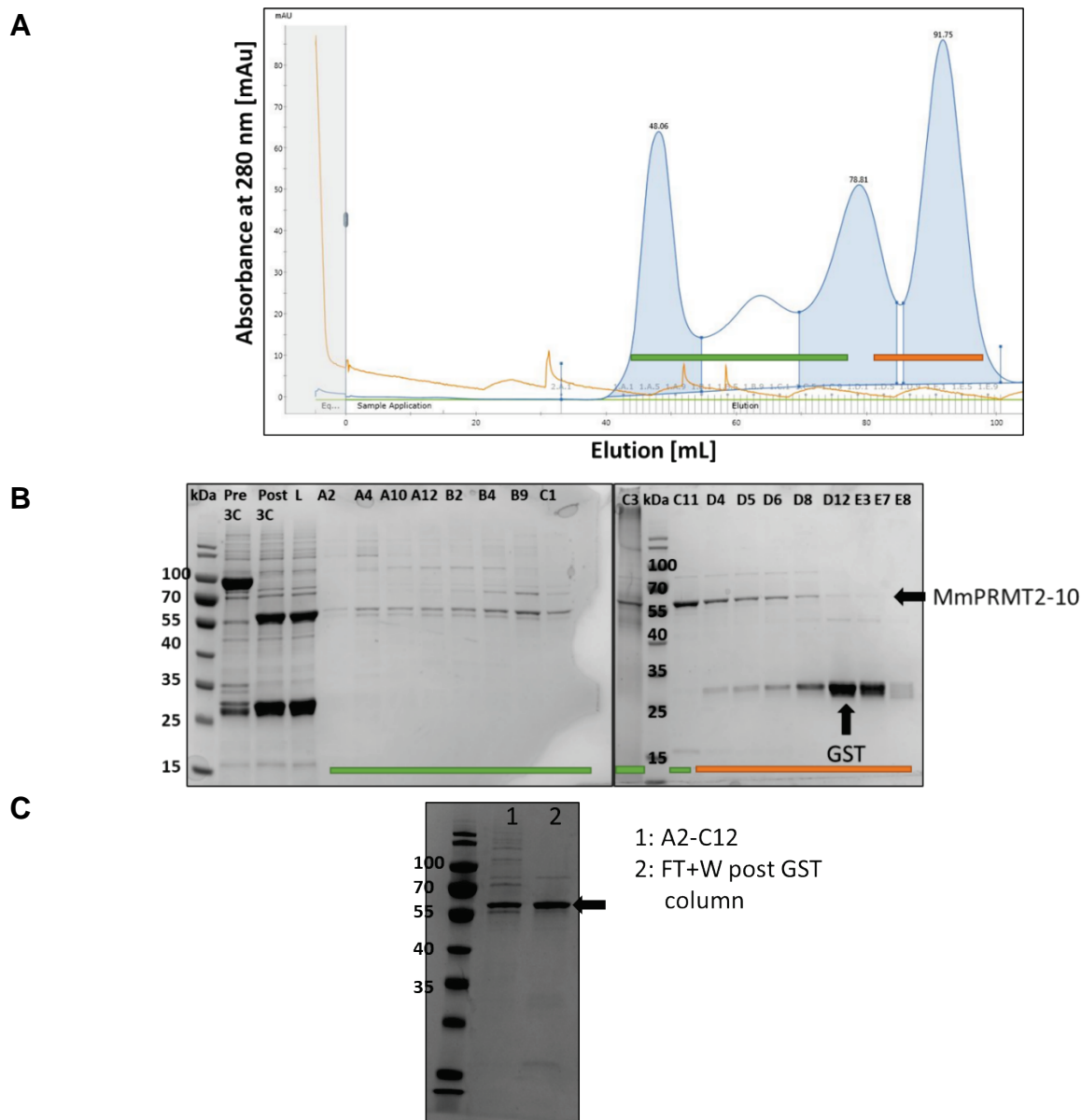


Figure 3-31 SEC and SDS-PAGE Analysis of MmPRMT2-10.

After GST-tag removal, the sample was applied onto a Superdex 200 16/60 column. The chromatogram is shown in **A**, the SDS-PAGE analysis in **B**. PRMT2-10 co-eluted with GST in the last peak (orange bar). **C**) Fractions A2-C12 (green bar) were pooled and concentrated, fraction C4-D11 (orange bar) was purified via subtractive GST-IMAC prior to concentration. The absorbance at 280 nm (blue line) was monitored. HsPRMT2-1 eluted as a single peak, fractions A10-B1 (orange bar) were pooled and concentrated. Purity was assessed after every step by SDS-PAGE analysis. Protein ladder: PAGERuler Pre-stained (ThermoFisher). The gels were stained with Instant Blue Stain.

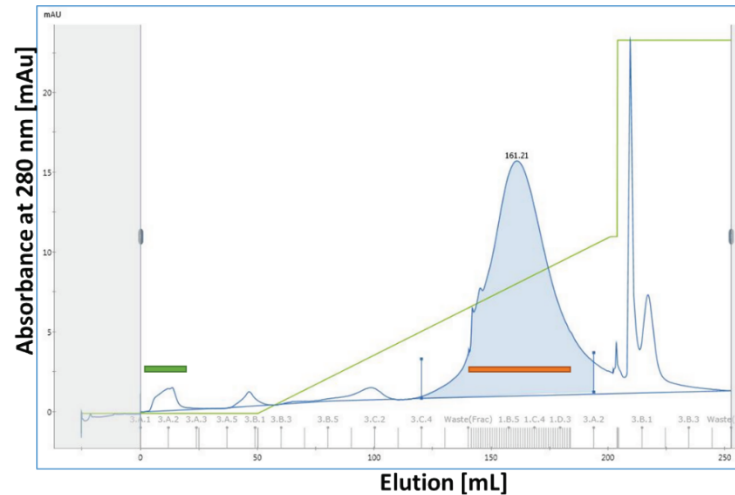
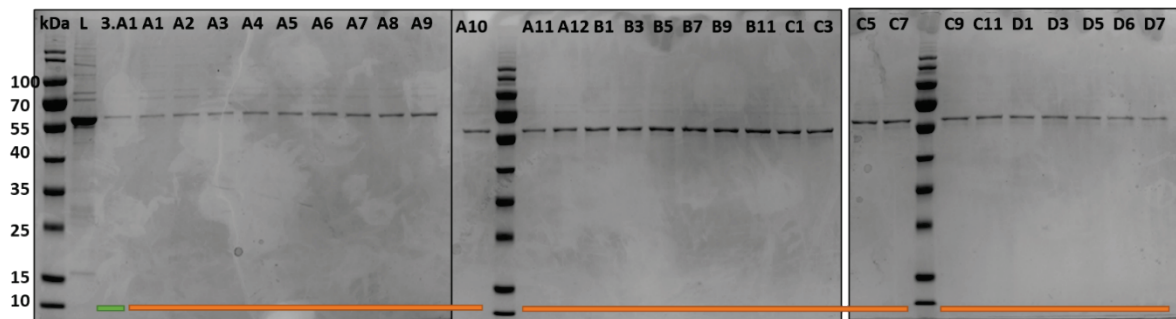
A**B**

Figure 3-32 IEX Chromatography of MmPRMT2-10.

The concentrated MmPRMT2-10 sample was loaded onto a HiTrap Q FF column and eluted with a 0- 500 mM NaCl gradient. The chromatogram and SDS-PAGE analysis are shown. Fractions A10– C4 and C5- D7 were pooled and concentrated to 3.5 mg/mL and used for crystallisation trials. The other fractions 3A.1-A9 were concentrated to 2 mg/mL. Protein ladder: PAGERuler Pre-stained (ThermoFisher). The gels were stained with Instant Blue Stain.

3.9 Summary

At the beginning, a variety of different expression constructs were designed for rat PRMT1, human PRMT2, and human CARM1 for expression in *E. coli* cells. Different *E. coli* strain and expression conditions were screened using small-scale expression tests, but no soluble protein expression was detected for any of the tested constructs. Selected constructs were expressed in a larger scale to rule out that differences in aeration between small-scale and large-scale expression could affect protein production amounts. Rat PRMT1 was successfully expressed and purified. The protein can be used at a later stage for selectivity screening.

Next, the expression system was changed to the Bac-to-Bac Baculovirus Expression System (Berger *et al.*, 2004) with the hope that changing the expression host would help to increase protein solubility and yields.

Again, different CARM1 and PRMT2 constructs were tested in small- and large-scale expression tests in Sf9 and Sf21 cells and the catalytic domain of human CARM1, full-length human PRMT2 and a shorter PRMT2 construct that is missing the N-terminal SH3 domain were successfully expressed and purified. CARM1 purification yields were very high, sufficient for the development of biophysical assays and crystallisation studies. In contrast, PRMT2 yields were very low. Buffer optimisation screens were conducted to identify more suitable buffers for PRMT2 that stabilise the protein during purification. However, none of the buffers increased protein stability for PRMT2. Additionally, limited proteolysis studies were performed with human PRMT2 to identify more stable constructs that could be used for future protein expression and crystallisation trials. Simultaneously, the first high-resolution crystal structures of mouse and zebrafish PRMT2 were published (Cura *et al.*, 2017). Thus, it was decided, to focus on the expression of the crystallised literature construct. Mouse PRMT2 was successfully expressed in insect cells and purified but protein yields were still limited.

In the future, PRMT2 construct with different tags, including C-terminal tags or different insect cell strains (Sf21, High Five™) as expression host or alternative expression systems such as eukaryotic cell-free systems could be used to increase PRMT2 expression yields (Thoring *et al.*, 2019). However, the observed varying and in general low protein yields of PRMT2 do not make it a very promising drug target as its production is very costly and time consuming.

Chapter 4 Enzyme Activity Assay Development and Evaluation

The identification and characterisation of new small molecules that bind the target of interest selectively and with high affinity are very useful tools to assist target validation studies and provide starting points for drug development (Rees, 2016). Traditionally, the interactions of large libraries with thousands to millions of small molecules (<500 Da, up to 30 heavy atoms) were analysed using *in vitro* high-throughput screening (HTS) assays often combined with structural relationship studies and virtual screening to identify so called hits, molecules which influence the protein function upon binding. The aim is to identify drug-like molecules that possess binding affinity of 1 μM or lower and that can be a starting point for lead identification and optimisation (Scapin, 2006, Shoichet, 2004). However, in recent years, fragment-based drug design (FBDD) is gaining more and more importance. In FBDD, screening libraries are much smaller in size (hundreds to thousands) but the molecules also have no drug-like character but instead are smaller (< 300 Da, less than 20 heavy atoms) and thus the possibility of identifying hits is higher (Erlanson *et al.*, 2016). The aim is to identify effective but low affinity binders with binding affinities in the high μM or mM range, which can then be optimised and developed into larger fragments (Mannhold *et al.*, 2015). The advantage of slowly growing the fragment into the chemical space is that it is possible to control important drug properties including solubility and toxicity which can lead to the development of better molecules with better pharmaceutical properties (Hajduk and Greer, 2007, Meanwell, 2015). Due to the lower binding affinities of the fragments, screening assays need less sensitivity and are more robust (Erlanson *et al.*, 2016).

Regardless of which approach is chosen, the biochemical assays used for the identification and quantification of the target enzyme's activity and its modulation by inhibitors are of great importance. An effective assay reports the authentic protein activity while using a system of reduced complexity. The issue with *in vitro* assays is often that the enzymes show reduced or no activity. Possible reasons are for example differences between the chosen protein construct and the wild type protein, missing binding partners that are present in the cell environment, reduced protein stability, and the usage of inappropriate enzyme substrates (Acker and Auld, 2014).

As already discussed, PRMTs belong to the class of transferase enzymes that catalyse the transfer of a methyl-group from a donor to an acceptor molecule. In the

case of PRMTs the methyl-donor is SAM generating the by-product SAH and the acceptor is the substrate arginine residue. A variety of different *in vitro* methylation assay exist for PRMTs, an overview is given in Figure 4-1.

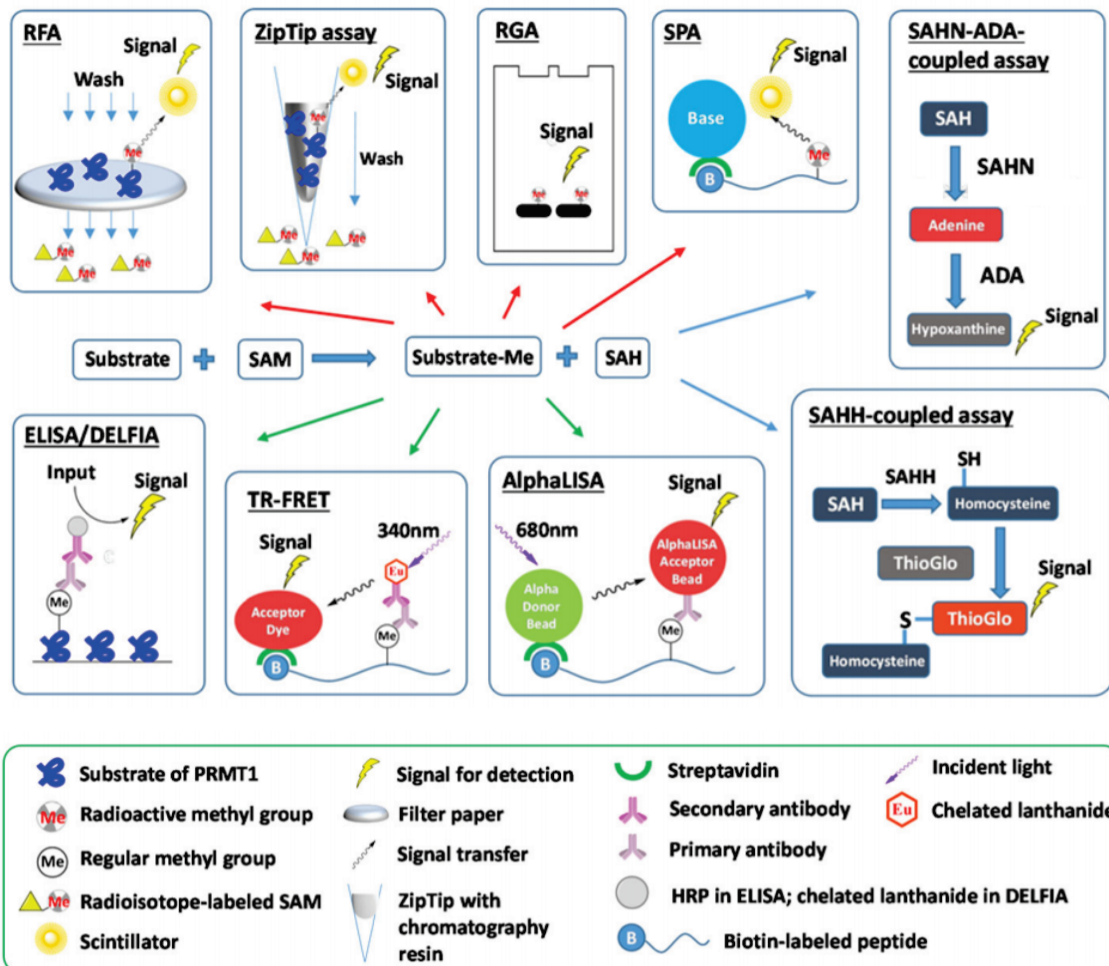


Figure 4-1 Overview of PRMT Activity Assays.

Different assays exist that either measure the amount of methylation on the intact protein, production of the co-product SAH or the amount of methylated product. Radiometric (red arrows), antibody-based (green arrows) and enzyme-coupled (blue arrows) assays are available. Abbreviations: RFA: Radiometric Filter Assay; RGA: Radiometric Gel Assay; SPA: Scintillation Proximity Assay. SAHN: SAH Nucleosidase. ADA: Adenine Deaminase. ELISA: Enzyme-Linked Immuno-Sorbent Assay; HRP: Horseradish Peroxidase; DELFIA: Dissociation-Enhanced Lanthanide Fluorescent Immunoassay; TR-FRET: Time-Resolved Fluorescence Resonance Energy Transfer; SAHH: SAH Hydrolase. Figure reproduced, with permission, from Hu et al. (2016).

Most of the existing PRMT activity assays measure either the formation of the methylated product or the by-product SAH rather than depletion of the substrate due to the slow turnover of the reaction.

Assays include direct assays, where the methylated substrate is detected, such as radiometric assays and antibody-based assays but many other indirect assays exist. At the start of the project, the direct detection of methylation by using radioactively labelled SAM, which allows the quantification of radiolabelled methylated substrate by autoradiography or liquid scintillation counting, was the most sensitive and reliable method to measure *in vitro* PRMT activity (Hu *et al.*, 2016). PRMT2 activity has also been studied using radioactive assays (Cura *et al.*, 2017, Lakowski and Frankel, 2009). Another advantage is that no labelling or antibodies are needed, and the assay can be used for a broad number of PMTs without the need of modifications. However, the method has a number of drawbacks. First, the production of radioactive waste is not environmentally friendly and does not allow differentiation between MMA, ADMA and SDMAs. Moreover, washing steps are needed prior to quantification to remove the radioactively labelled SAM from the reaction. This step can be achieved by using reversed-phase resin-filled pipette tips (ZipTips) (Hevel and Price, 2020), radiometric gel assays (RGA) (Bissinger *et al.*, 2011) or filter-binding assays (RFA) (Alinari *et al.*, 2015).

To overcome this problem, Scintillation Proximity Assays (SPA) were developed in which biotinylated peptides and radiolabelled SAM are used as PRMT substrates. After completion of the reaction, the methylated peptides are immobilised using streptavidin-coated scintillants (plates or beads), the proximity between them causes the scintillation signal, which is not affected by the presence of unreacted SAM (Luo, 2012). The SPA-based assay has been successfully applied for activity measurement of PRMT1 and other methyltransferases and is also HTS compatible (Wu *et al.*, 2012, Rathert *et al.*, 2007).

Furthermore, different antibody-based assays that detect the methylated substrates are available and a variety of antibodies exist that detect the different methylation types. Used formats include Enzyme-Linked Immunosorbent assays (ELISA) (Cheng *et al.*, 2004) and Dissociation-Enhanced Lanthanide Fluorescent Immunoassays (DELFI) (Bissinger *et al.*, 2011), in which the antibody is labelled with HRP or a lanthanide probe, respectively. Moreover, Amplified Luminescence Proximity Homogeneous Assays (Alpha) and Time-Resolved Fluorescence Resonance Energy assays (TR-FRET) have been used (Xie *et al.*, 2014). Most of them are available in an HTS format. Disadvantages include that cross-reactivity or non-specificity of antibodies can

occur. In addition, different antibodies for each methylation site would be needed which is expensive and the format cannot be used for full-length histones.

The reaction by-product SAH can be directly detected via anti-SAH antibodies or indirectly by further conversion to different derivatives using colorimetric, fluorescence, and luminescence based coupled enzyme assays (Luo, 2012). To avoid exposure to radioactivity, an indirect bioluminescence-based assay was selected that measures the production of the by-product SAH. A common drawback of indirect assays is that many different steps are needed involving coupling enzymes which can interfere with the assay and result in false signals. However, in its favour, the assay does not involve radioactivity, it is antibody-free, and it can also be employed to measure the enzyme activity of other members of the methyltransferase family over a broad concentration range. Another advantage is that the assay already existed in a high-throughput format (Hsiao *et al.*, 2016). In the next section, this assay will be explained in more detail.

4.1 The MTase Glo Assay

A luminescent assay to measure methyltransferase activity that indirectly detects SAH formation (Promega MTase Glo™) was used as a first assay to analyse the enzyme activity of the purified PRMTs (Hsiao *et al.*, 2016). The assay principle is shown in Figure 4-2. After completion of the methyltransferase reaction, SAH is enzymatically converted into adenosine, AMP and then ADP, using the three enzymes SAH hydrolase, adenosine kinase (AdK) and polyphosphate-AMP phosphotransferase (PAP). ADP is converted into ATP and then quantified via luminescence using the luciferase/luciferin reaction (Hsiao *et al.*, 2016). The assay has the advantage that it is commercially available, cost-effective, uses only small amounts of protein, and could be used for CARM1, PRMT2 and other potential PRMTs for compound counter screening at a later stage. However, as with any enzyme-coupled assay, a counter assay is needed to detect false-positive hits. Another problem with using SAH detection for PRMT activity measurement is that the cofactor SAM can spontaneously decompose to SAH, which can result in high background signals. For this reason using ultra-pure SAM is very important (Lakowski and Frankel, 2010). The same is the case if the PRMTs show high auto-methylation activity or SAH contaminations are present (Luo, 2012). The decomposition of SAM to SAH is only very slow and thus could only affect

measurements with PRMT that have a low enzyme activity and thus produce only low amounts of SAH (Lakowski and Frankel, 2010).

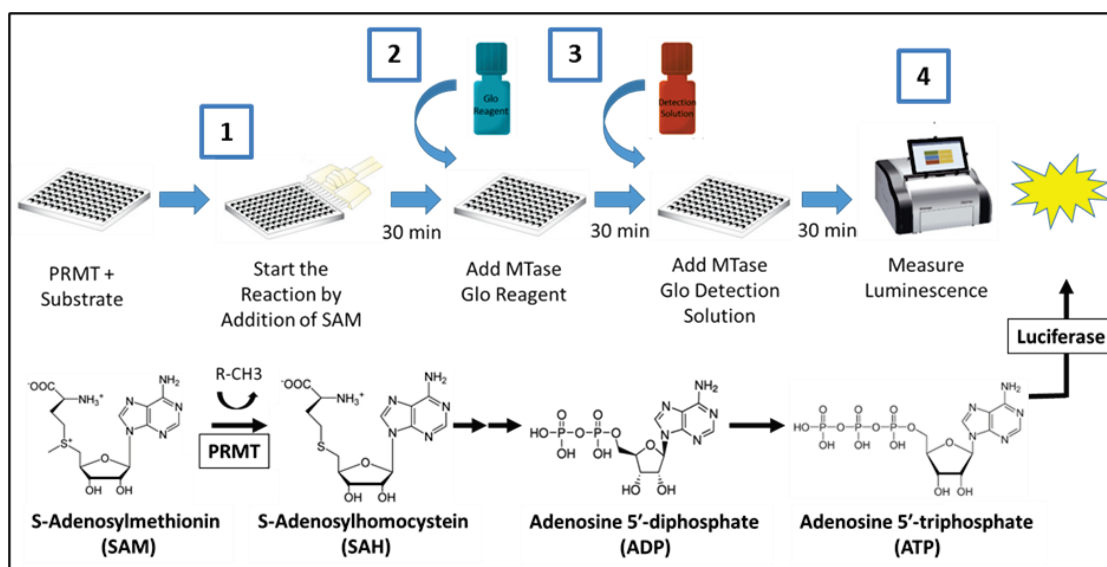


Figure 4-2 Principle of the Bioluminescence-based MTase-Glo Assay.

The assay is used to monitor the activities of PRMTs and their modulation by small molecule inhibitors. 1) PRMTs catalyse the transfer of methyl groups from a donor molecule, S-adenosyl-methionine (SAM), to the terminal guanidino nitrogen of arginine residues, generating S-adenosyl-L-homocysteine (SAH). 2) In the next step the MTase-Glo Reagent is added to convert SAH to ADP. 3) Then the MTase-Glo Detection solution is added to convert ADP to ATP, which can be detected via the luciferase reaction (4). The luminescence is proportional to the amount of generated SAH/enzyme activity.

4.2 Substrates of CARM1 and PRMT2

In addition to the enzymes, other important assay components are cofactors and suitable enzyme substrates with high stability and purity. As already discussed above, all PRMTs use SAM as methyl-donor, which is commercially available at very high purity. As illustrated in Figure 4-3, CARM1 methylates different arginine residues on human histone H3 (Schurter *et al.*, 2001), and PRMT2 was shown to have weak activity on Arg8 of histones H3 and H4 (Lakowski and Frankel, 2009).

Additionally, other cytoplasmic and nuclear non-histone substrates exist for many PRMTs (Lorton and Shechter, 2019). CARM1 for example methylates the poly(A)-binding protein 1 (PABP1) (Lee and Bedford, 2002), the steroid receptor coactivator SRC-3 (Frieze *et al.*, 2008), the histone acetyltransferases p300 and CBP (Xu *et al.*, 2001, Feng *et al.*, 2006) and a variety of splicing factors, for example U1C (Cheng *et al.*, 2007). PABP1 is often used as a CARM1 substrate as it shows high

activity and is easy to produce and commercially available. However, the CARM1 construct that is readily expressed comprises only amino acid residues 135-482 and is missing both the N-terminal PH domain (residues 30-127 in HsCARM1) and the C-terminal domain. The N-terminus of CARM1 was shown to directly interact with a large number of substrates. The unique N- and C-terminal domains of CARM1 are not necessary for methyltransferase activity, but their deletion can influence CARM1 substrate recognition and often reduces or abolishes CARM1 activity. Although the CARM1 catalytic domain alone is able for example to methylate histone H3 (Teyssier *et al.*, 2002), this is not the case for many substrates including BAF155 and PABP1 (Shishkova *et al.*, 2017).

At the start of this project, histone H4 and GST-GAR were the only reported PRMT2 substrates. Thus, it was decided to first use these and the human full-length histone H4 (UniProtKB: P62805) as substrates in the MTase Glo™ Assay. At a later stage, a new substrate RSF-1 was published and produced together with GST-GAR, which will be discussed below. Beside the purified HsPRMT2-1, HsCARM1, and RnPRMT1-1, HsPRMT5(MEP50) was also tested (kindly provided by Dr Paul Stuppel, Monash University, Australia). RnPRMT1 and HsPRMT5 both methylate Arg3 on histone H4 (Zhao *et al.*, 2009, Strahl *et al.*, 2001). HsCARM1 is able to methylate different arginine residues on histone H3 but not on H4 (Figure 4-3), which is why histone H3.1 (UniProtKB: P68431) was used as an assay substrate. For further details refer to Section 1.8.

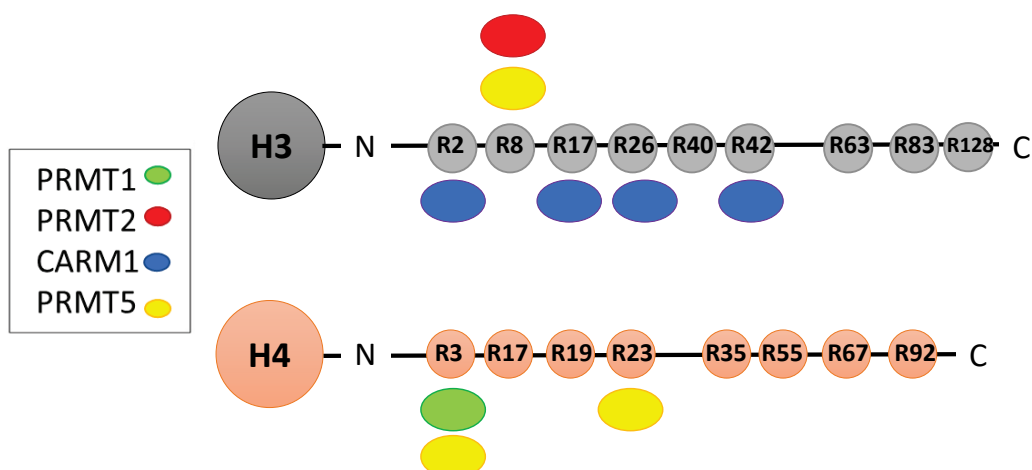


Figure 4-3 PRMT Arginine Methylation Sites on Histone H3 and H4.

Histone arginine methylation positions are shown by coloured ovals for different PRMTs. PRMT1 and PRMT5 methylate Arg3 on histone H4. PRMT1 and PRMT5 methylate Arg3 on histone H4. PRMT5 is also able to methylate Arg8 on histone H3 and Arg23 on histone H4. PRMT2 was shown to have low enzyme activity on histone H4, and Arg8 on histone H3. CARM1 methylates Arg2 (weak activity), Arg17, Arg26 and Arg42 on histone H3. Figure adapted, with permission from Rakow et al. (2020).

To assess RnPRMT1, HsPRMT2 and HsPRMT5(MEP50) activity, assays were performed in a final volume of 4 μ L using 1 μ M SAM and 5 μ M full-length histone H4 (Figure 4-4, A-B, E) and varying enzyme concentrations (HsPRMT2-1: 0- 2.50 μ M; RnPRMT1: 0- 0.32 μ M; HsPRMT5(MEP50): 0- 2.74 μ M) as described in Section 2.18.1. HsCARM1 was assayed using 1 μ M SAM and 1 μ M full-length histone H3 (Figure 4-4 C) and 5 μ M SAM and 5 μ M H3 (Figure 4-4 D) with CARM1 concentrations ranging from 0- 1.57 μ M per reaction using two different CARM1 preparations. A SAH standard curve was prepared for each reaction, as described in Section 2.18.1, an exemplary curve is shown in Figure 4-4 F. Data were plotted using Relative Luminescence Unit (RLU) values and analysed in Excel/GraphPad Prism and are represented as mean \pm standard error. Reactions were performed as technical replicates. The optimal assay enzyme concentration lies in the linear part of the curve, where the enzyme concentration is proportional to enzyme activity and shows a significant luminescent signal (Bisswanger, 2014).

In each assay, the luminescence signal was linearly dependent on the PRMT concentration, suggesting that the activity observed is due to the enzyme activity of the protein. In the case of PRMT1, CARM1 and PRMT5(MEP50) the activity curves are as expected for an enzyme-dependent reaction being first linear and then reaching a plateau due to the depletion of the histone substrate. When comparing the two CARM1

titrations (Figure 4-4 C and D) it becomes evident, that the shape of the two curves are different. The second titration (Figure 4-4 D) was performed with a different CARM1 preparation. The CARM1 activity is higher in the second titration than in the first that used protein which was stored at -80 °C for longer.

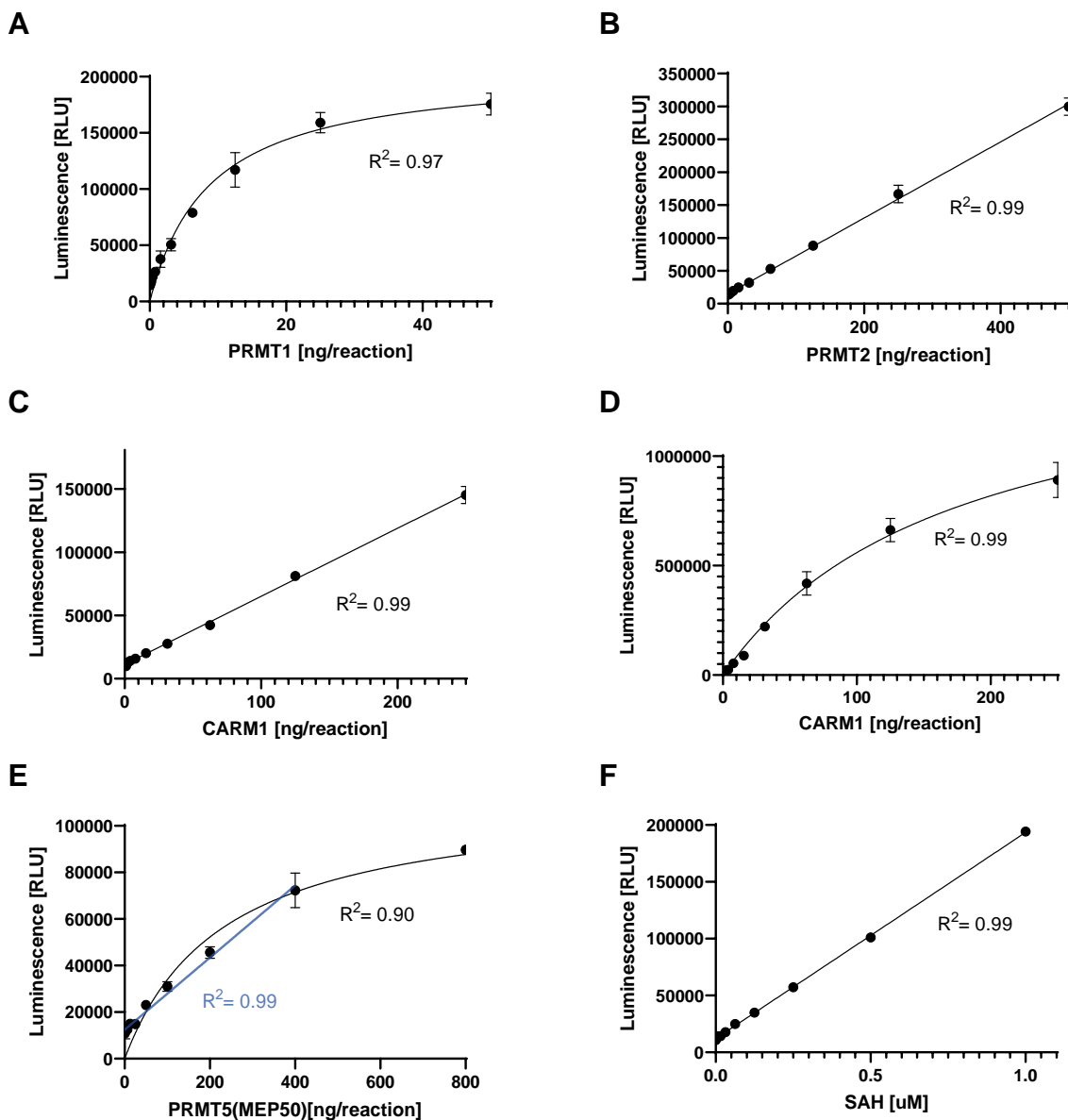


Figure 4-4: Measurement of PRMT Activity using the MTase Glo Assay.

A) PRMT1 titration using 1 μ M SAM and 5 μ M H4; **B)** PRMT2 titration using 1 μ M SAM and 5 μ M H4; **C)** CARM1 titration using 1 μ M SAM and 5 μ M H3; **D)** CARM1 titration using 5 μ M SAM and 5 μ M H3; **E)** PRMT5(MEP50) titration using 1 μ M SAM and 5 μ M H4; **F)** SAH standard curve. Reactions were performed at room temperature for 30 min according to the kit protocol (two-step reaction). Luminescence was measured using a PHERAstar FS (BMG LAB-TECH) plate-reader and data were analysed in GraphPad Prism using relative luminescence units (RLU). Each point represents an average of two experiment. Technical replicates ($n=2$) were performed using the same enzyme preparation at different days.

Next, to find the optimal saturating substrate concentration, which should be between 10-100x the K_m value to ensure 91-99 % occupation of the binding site (Bisswanger, 2014), histone titrations for RnPRMT1 (Figure 4-5) and HsPRMT2-1 (Figure 4-6) were performed as described in Section 2.18.2. PRMT1 is known to methylate Arg3 on Histone H4. To determine a K_m value for this substrate, a histone titration experiment was performed with 10 μ M SAM and 13 nM RnPRMT1 per reaction. PRMT1 showed typical Michaelis–Menten kinetics as a function of H4 concentration with a K_m of $0.11 \pm 0.03 \mu$ M calculated using GraphPad Prism 8.2 software (Prism, 2003). This is much lower than the K_m of $1.45 \pm 0.37 \mu$ M previously reported in the literature (Osborne *et al.*, 2007). However, in the published assay, the SAM concentration was higher (15 μ M), a different assay format (gel based radioactive assay) with different buffer conditions was used and the PRMT1 still carried an N-terminal histidine tag. Thus, the two results are difficult to compare, but a biological repeat should be carried out to confirm the results.

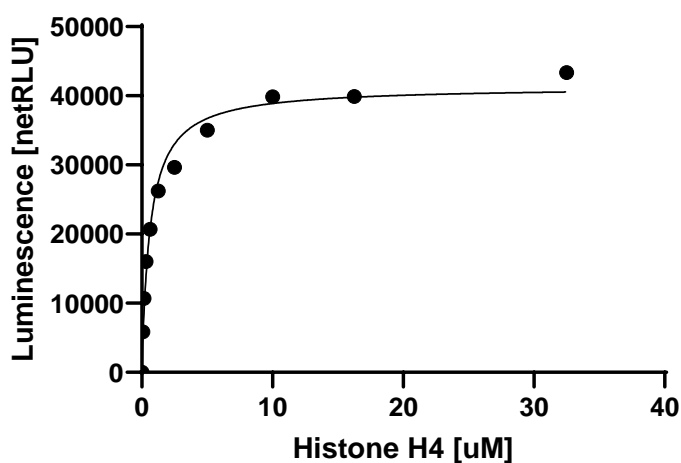


Figure 4-5 Determining the K_m value of H4 for PRMT1.

RnPRMT1 methylation was monitored with varied histone H4 concentrations using the MTase Glo™ Assay. Reactions were performed at room temperature for 30 min with 2 ng/reaction PRMT1 and 10 μ M SAM according to the kit protocol (two-step reaction). Luminescence was measured using a PHERAstar FS (BMG LABTECH) plate-reader and data were analysed in GraphPad Prism using net relative luminescence units (RLU). Each point represents an average of an experiment done in duplicate.

However, when H4 was titrated against PRMT2 in the presence of excess SAM, no curve was observed, suggesting that PRMT2 has no or only very little activity towards H4 (Figure 4-6). A very low PRMT2 activity on H4 has also been reported in other kinetic studies (Lakowski and Frankel, 2009, Cura *et al.*, 2017), and it has been

suggested that the observed methylation activity could be due to auto-methylation (Cura *et al.*, 2017).

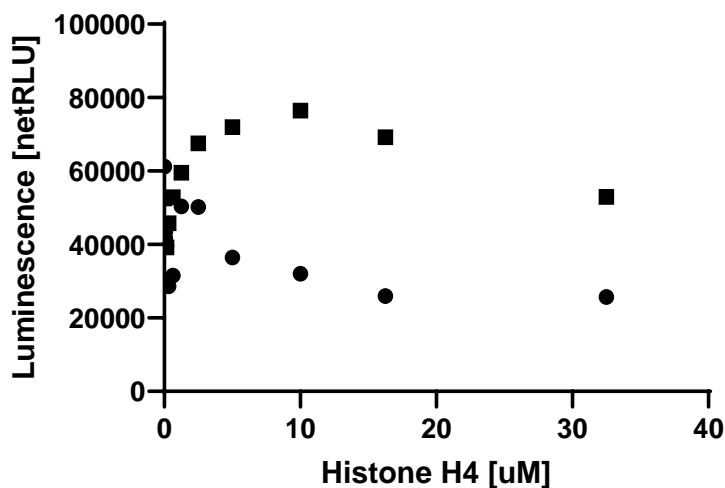


Figure 4-6 Histone H4 Titration for PRMT2.

HsPRMT2-1 catalysed methylation of histone H4 was monitored with increasing H4 concentrations using the MTase Glo™ Assay. Reactions were performed at room temperature for 30 min with 35 ng/reaction PRMT2 and 10 μM SAM according to the kit protocol (two-step reaction). Luminescence was measured using a PHERAstar FS (BMG LABTECH) plate-reader and data were analysed in GraphPad Prism using net relative luminescence units (RLU).

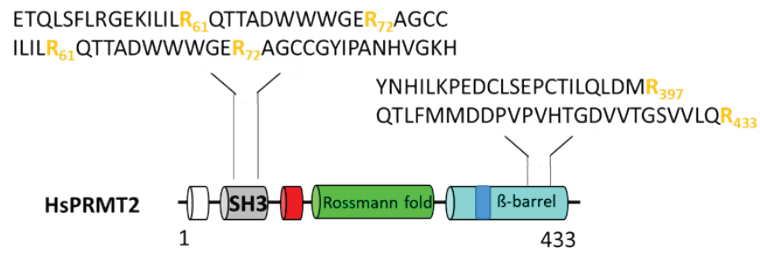
To test whether PRMT2 shows enzyme activity in the absence of histone, a PRMT2 titration was performed in the presence of 10 μM SAM (Figure 4-7 B). A linear curve was obtained but compared with the enzyme activity in the presence of histone H4 (0.22 nM SAH/min/ng HsPRMT2-1), the activity is much lower (0.12 nM SAH/min/ng HsPRMT2-1). The same experiment was performed with the shorter PRMT2 construct HsPRMT2-6 which is missing its N-terminal SH3 domain (Section 3.2). An enzyme titration from 0-3.15 μM was performed in the presence of 10 μM SAM and 0 or 5 μM histone H4, the results are shown in Figure 4-7 C. HsPRMT2-6 also showed activity in the MTase Glo™ assay without any histone substrate present, but it was slightly lower than with histone H4 (0.13 vs 0.17 nM SAH/min/ng HsPRMT2-6). However, *t*-test showed that the activity differences were not statistically significant with *t* = 0.64, *p* > 0.05. Thus, the measured HsPRMT2-6 activity is most likely due to auto-methylation and HsPRMT2-6 has no activity on histone H4 under the tested conditions.

Taken together, these results suggest that both PRMT2 constructs used in these assays have measurable auto-methylation activity. Moreover, HsPRMT2 shows methyltransferase activity also without the SH3 domain, but the activity is lower when compared with the full-length enzyme.

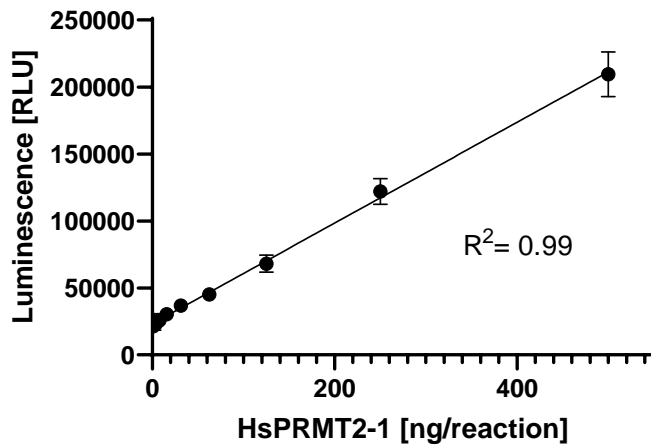
To identify potential auto-methylation sites, 0.2 μ M HsPRMT2-1 or HsPRMT2-6 was incubated with 20 μ M SAM over 2 h and samples were taken at specific time points for analysis by mass spectrometry. Analysis was conducted by Dr Rachel Heap (Newcastle University). In the case of HsPRMT2-1, two potential mono and di-methylation sites at Arg61 and Arg72 were detected. Both residues are located in the N-terminal SH3 domain. Two additional potential methylation sites at Arg397 and Arg433, located in the β -barrel domain were detected in both HsPRMT2-1 and -6 constructs. However, for these sites, methylation probability was only 57-59 %, which was low compared to the sites located in the SH3 domain, which were methylated with 80-90 % probability. In order to improve this data, assay and instrument replicates would be needed, but unfortunately not enough protein was available to conduct those experiments. The location of the potential HsPRMT2 auto-methylation sites identified using mass spectrometry are shown in Figure 4-7 A.

Two other potential PRMT2 substrates have been identified in other studies that showed much higher methylation activity compared to histone H4, namely GST-GAR (Lakowski and Frankel, 2009) and RSF1 (Cura *et al.*, 2017). Thus, it was decided to produce and assay these two substrates.

A



B



C

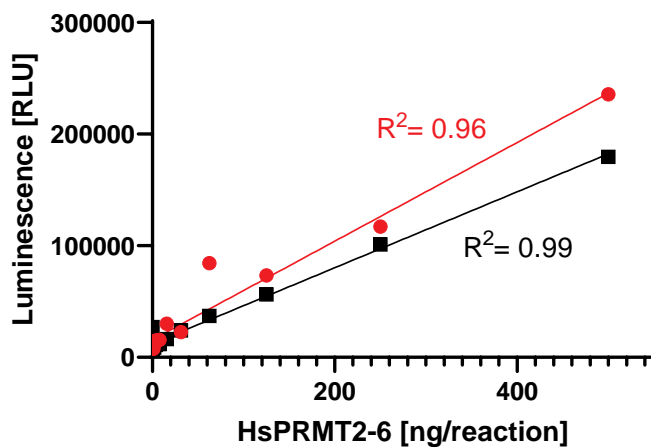


Figure 4-7 PRMT2 Auto-Methylation.

A) Schematic diagram illustrating four potential PRMT2 auto-methylation sites identified using mass spectrometry. **B)** HsPRMT2-1 activity measured in the absence of protein substrate but including 10 μ M SAM. **C)** HsPRMT2-6 activity measured using 10 μ M SAM with and without 5 μ M H4 substrate. Activity was measured using the MTase Glo™ assay format.

4.3 RSF1- a New PRMT2 Substrate?

In the absence of an authentic substrate, the enzyme activity of PRMT2 was questioned, until Cura *et al.* (2017) identified the repressor splicing factor 1 (RSF1) as a potential new substrate that co-purified during recombinant protein expression in the *S. frugiperda* host strain (Cura *et al.*, 2017). The RNA binding protein RSF1, previously known as ROX21 (Brand *et al.*, 1995), has a size of 16 kDa and belongs to the highly conserved serine/arginine (SR) family of splicing regulatory proteins (Brand *et al.*, 1995). Members of the SR family have two domains, one or two RNA-recognition motifs (RRMs) at their N-terminus and a C-terminal arginine- and serine-rich (RS) domain (Long and Caceres, 2009).

RSF1 was shown to co-regulate alternative splicing events with other SR family members (Bradley *et al.*, 2015). This activity is in contrast to earlier research that indicated that it antagonises splice-site recognition by other SR proteins (Labourier *et al.*, 1999). Figure 4-8 shows the RSF1 sequence and a RMM domain model which was predicted using Phyre2 (Kelley *et al.*, 2015). Two of the six methylation sites, Arg7 and Arg24, identified by Cura *et al.* (2017) are highlighted. They are situated in the N-terminal RRM domain (residues 1-77). Arg7 is located in a unique consensus sequence (SxGGxRxY). The remaining methylation sites are located in the disordered RS domain (residues 77-147) (Cura *et al.*, 2017).

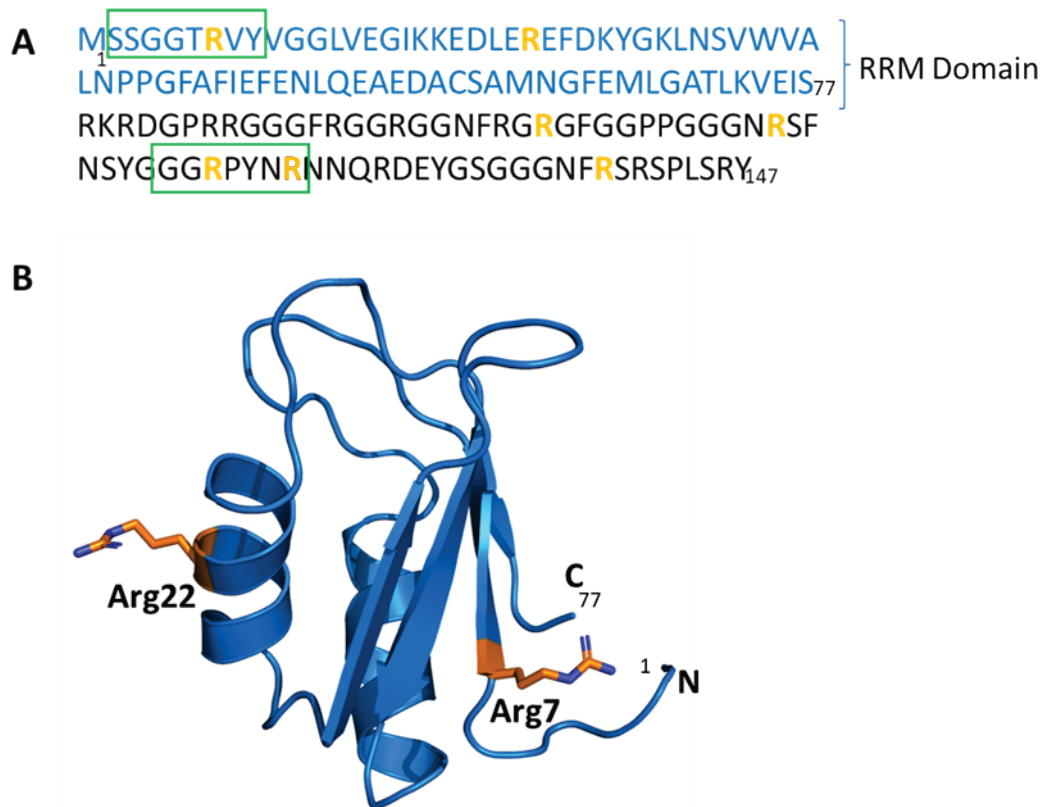


Figure 4-8 *S. frugiperda* RSF1.

A) Amino acid sequence of RSF1 with the six identified *MmPRMT2* arginine methylation sites shown in orange. The RRM domain is highlighted in blue, the disordered RS domain in black. The unique consensus sequence in which two methylation sites are located is boxed in green. **B)** The RRM domain structural model was predicted using Phyre2 (Kelley et al., 2015). The picture was generated using PyMol (DeLano, 2002). Two methylation sites are located in the RRM domain, the rest lies in the disordered RS domain (not shown).

4.4 Expression Test of His₆-RSF1

The *Sf*RSF1 sequence was synthesised by Eurofins Genomics, cloned into a pET3dM-His₆ expression vector and expression tests were performed in different *E. coli* strains as described in Section 2.16.2. His₆-RSF1 was overexpressed in BL21(DE3) and BL21(DE3)Star cells, purified by affinity chromatography, the tag was subsequently removed via 3C-protease cleavage and the fractions dialysed into a low imidazole buffer, followed by subtractive IMAC. The SDS-PAGE analysis of the purification is presented in Figure 4-9.

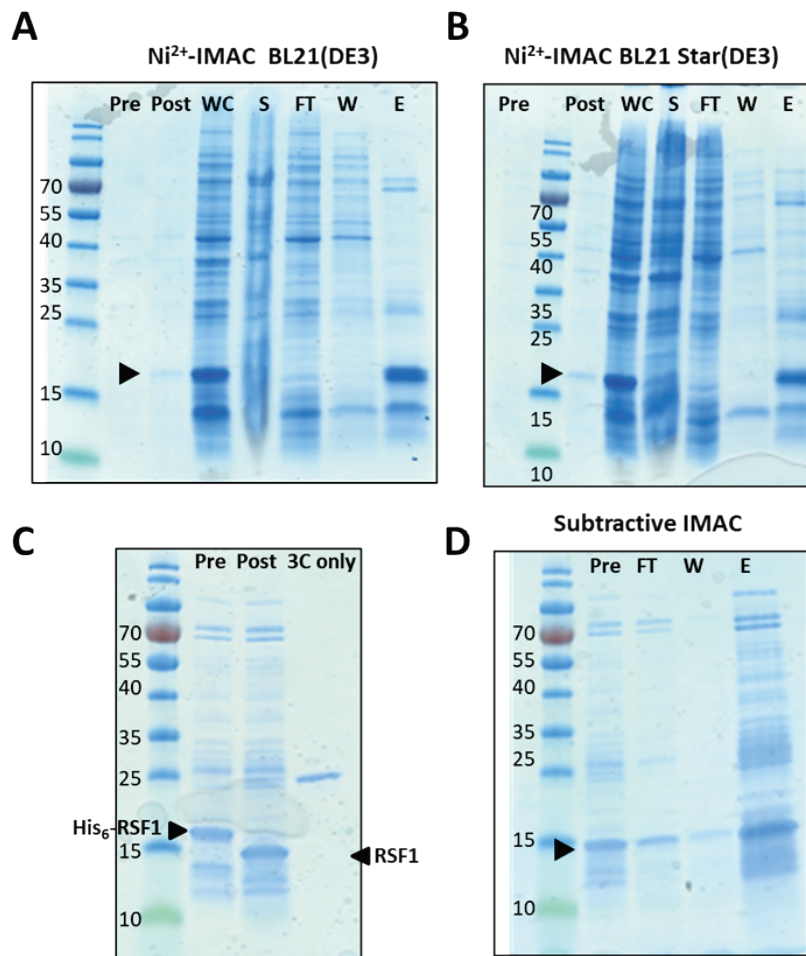


Figure 4-9 SDS-PAGE Analysis of His₆-RSF1 Purification.

His₆-RSF1 was expressed in *E. coli* BL21(DE3) and BL21 Star(DE3) and purified via Ni²⁺ IMAC. The SDS-PAGE analysis is shown in panel **A** and **B**. **C**) The elution fractions were combined (pre) and the His-tag removed via His-3C-protease cleavage (post). **D**) Flow-through and wash were combined, and His-tags removed via subtractive IMAC. The arrows indicate His₆-RSF1 or RSF1 (MW= 16 kDa) after tag removal. FT: Flow-through; W: Wash; E: Eluted fraction; WC: Whole Cell; S: Soluble. Protein ladder: PAGE Ruler Pre-stained (ThermoFisher).

Flow-through and wash were pooled, concentrated to ~ 4 mL and loaded onto an S200 16/60 column. The chromatogram and the SDS-PAGE analysis can be seen in Figure 4-10 A and B.

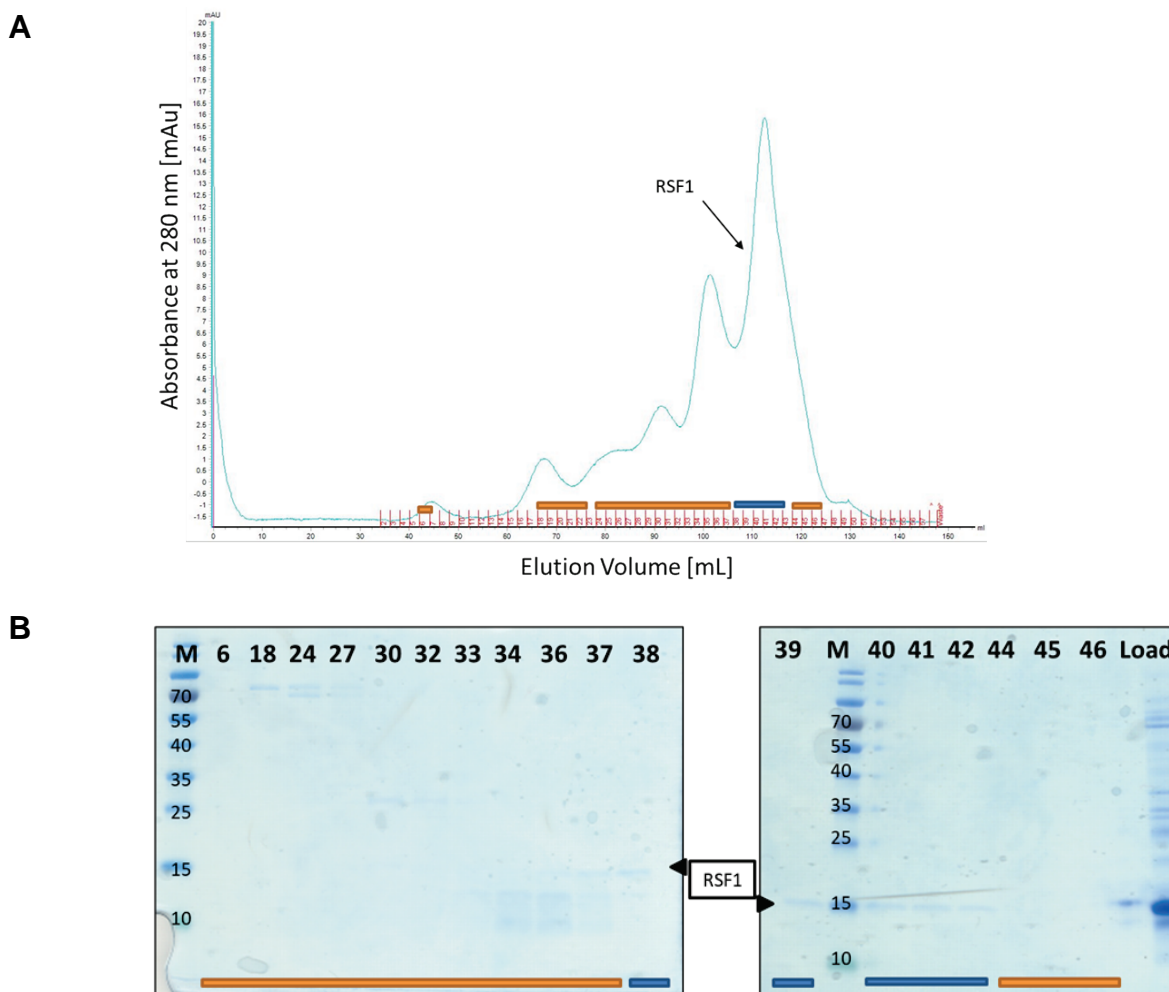


Figure 4-10 Purification of RSF1 by SEC.

A) Chromatography profile: RSF1 was analysed on a Superdex 200 16/60 column pre-equilibrated with SEC buffer (500 mM NaCl, 20 mM Tris pH 8.0, 0.5 mM TCEP) with a flow rate of 0.5 mL/min. The absorbance at 280 nm (blue line) was monitored. The RSF1-containing fractions were identified by SDS-PAGE (38-42; blue bar) (**B**). Protein ladder: PAGE Ruler Pre-stained (ThermoFisher).

Elution fractions containing RSF1 (Fraction 37-42) were identified by SDS-PAGE, the band corresponding to RSF1 is indicated by a black arrow. Fraction 38-42 (blue bar) were pooled and concentrated to 390 μ L at 0.5 mg/mL (final yield 0.2 mg).

The identity of RSF1 which has an expected molecular weight of 16111.80 Da was confirmed by intact mass analysis. The dominant mass species detected was 16169.24 Da. Another species with a mass of 12.35 kDa was also detected, most likely a co-purifying impurity (Appendix H). RSF1 was tested in the MTase Glo™ Assay with 10 μ M SAM and 5 μ M RSF1 and varying HsPRMT2-1 concentrations (Figure 4-11). Although activity could be measured, it was low, and unfortunately there was insufficient material to repeat the assay in parallel with a no substrate control. Thus, it cannot

be said with certainty if the measured PRMT2 activity in the presence of RSF1 is due to auto-methylation of PRMT2 or actual methylation of the new potential substrate.

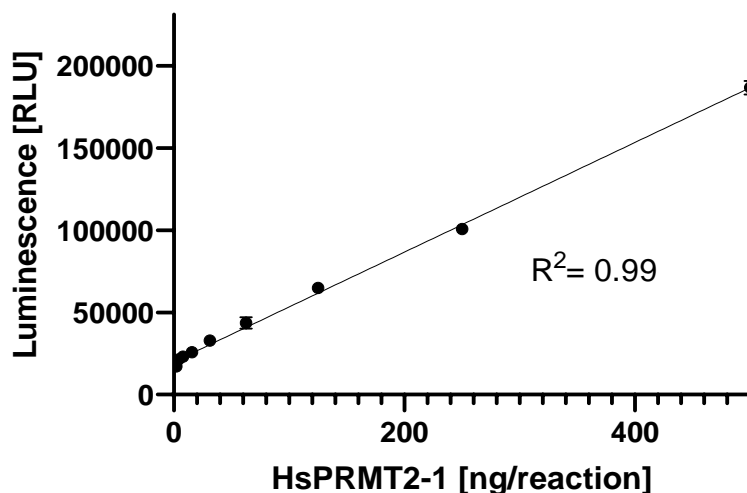


Figure 4-11 RSF1 Activity Assessment using the MTase Glo Assay.

The assay was performed at room temperature for 30 min according to the kit protocol (two-step reaction) using 10 μ M SAM and 5 μ M RSF1. Luminescence was measured with a PHERAstar FS (BMG LABTECH) plate-reader and data were analysed in GraphPad Prism using relative luminescence units (RLU). Each point represents an average of a measurement done in duplicate on the same plate.

4.5 PRMT Substrate GST-GAR

GST-GAR (glycine and arginine-rich) is a 41 kDa fusion protein comprising residues 1-145 of human fibrillarin (UniProtKB: P22087, 321 AA) and a glycine-rich region (AA 8-79) that contains 15 arginine residues. GST-GAR has previously been used as a substrate to assay PRMT1 (Tang *et al.*, 2000), PRMT2 (Lakowski and Frankel, 2009), PRMT3 (Tang *et al.*, 1998) and PRMT6 (Frankel *et al.*, 2002). In contrast, CARM1 is not able to methylate GST-GAR, as its substrates do not contain GAR methylation motifs (Poulard *et al.*, 2016). The aim was to express and purify sufficient amounts of the fusion protein to use it as substrate in activity assays for PRMT1 and 2. The plasmid containing GST-GAR was bought from Addgene (#34697, deposited by Dr Steven Clarke, described in Tang *et al.* (1998) and was supplied in bacteria as an agar stab.

Small-scale expression tests of GST-GAR were performed in four different *E. coli* strains, namely BL21(DE3)plysS, Rosetta(DE3)plysS, BL21Star(DE3) and BL21(DE3)Gold as described in Section 3.3. The expression media tested were AIM,

LB, TB, and 2YT medium. The samples were purified via spin columns using Glutathione Sepharose 4B resin as described in Section 2.13. After elution from the resin, samples were analysed for GST-GAR expression levels by SDS-PAGE. GST-GAR was only detectable when expressed in BL21(DE3)pLysS cells grown in AIM. Figure 4-12 shows the corresponding SDS-PAGE analysis of the elution fractions after purification. A strong band close to the expected molecular weight of 41 kDa is visible in all elution fractions. However, many different bands are visible below the expected molecular weight of GST-GAR reaching from 25-40 kDa, which indicates that GST-GAR is most likely not very stable and is proteolyzed into smaller fragments during expression or purification. This behaviour has been previously reported (Tang *et al.*, 1998). However, it could still be used as a substrate to test enzyme activity.

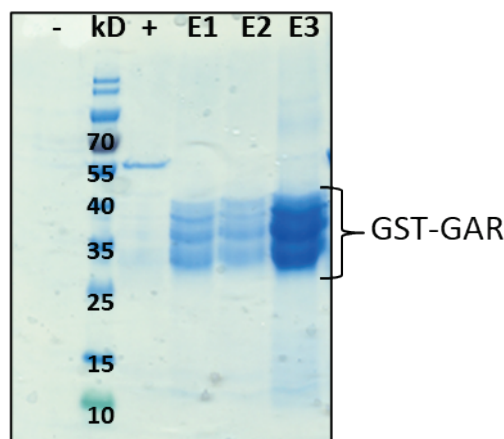


Figure 4-12 GST-GAR Expression Test in BL21(DE3)pLysS.

SDS-PAGE analysis of elution fractions of the Glutathione Sepharose 4B purification. E. coli cells were grown in AIM for 7 h at 37 °C, the temperature was then reduced O/N to 18 °C. After purification 10 µL of the elution fractions were analysed by SDS-PAGE Analysis. GST-GAR has an expected molecular weight of 41 kDa. Protein ladder: PAGE Ruler Pre-stained (ThermoFisher).

Additional small-scale expression tests with GST-GAR were conducted at the Newcastle University Protein and Proteome Analysis Facilities (NUPPA). Four expression strains, BL21-AI, BL21(DE3), BL21(DE3)Codon+, and BL21(DE3)pLysS, were tested at two different temperatures (30, 37 °C). BL21-AI *E. coli* cells are especially designed for the expression of very toxic recombinant proteins. However, no soluble recombinant protein was detectable by SDS-PAGE analysis (Appendix J).

4.6 MALDI-ToF MS-based Methylation Assay

In 2017 a direct, label-free, and sensitive methylation assay was reported for PKMTs (Guitot *et al.*, 2017). This assay uses matrix-assisted laser desorption/ionization-time-of-flight-mass spectrometry (MALDI-TOF-MS) to assess modification of peptide substrates, and it was shown that it can be used for kinetic studies and also inhibitor screening (Guitot *et al.*, 2017, Guitot *et al.*, 2014). Used in combination with a Mosquito Liquid Handling System, the assay can run in HTS format. It has also been formatted to monitor other post-translational modifications including phosphorylation (Heap *et al.*, 2017, Winter *et al.*, 2018) and ubiquitination (Ritorto *et al.*, 2014).

The label-free assay has many advantages compared with the MTase Glo™ Assay. The assay is a direct method that avoids the need for coupled enzymatic assays reducing the risk of assay interference. Moreover, no internal standards are needed, sample volumes are very small, and it is a very fast method (Guitot *et al.*, 2017). Therefore, it was decided to change the PRMT activity assay format to this method, where the appearance of mono- and di-methylated histone peptides is directly monitored. The assay principle is shown in Figure 4-13.

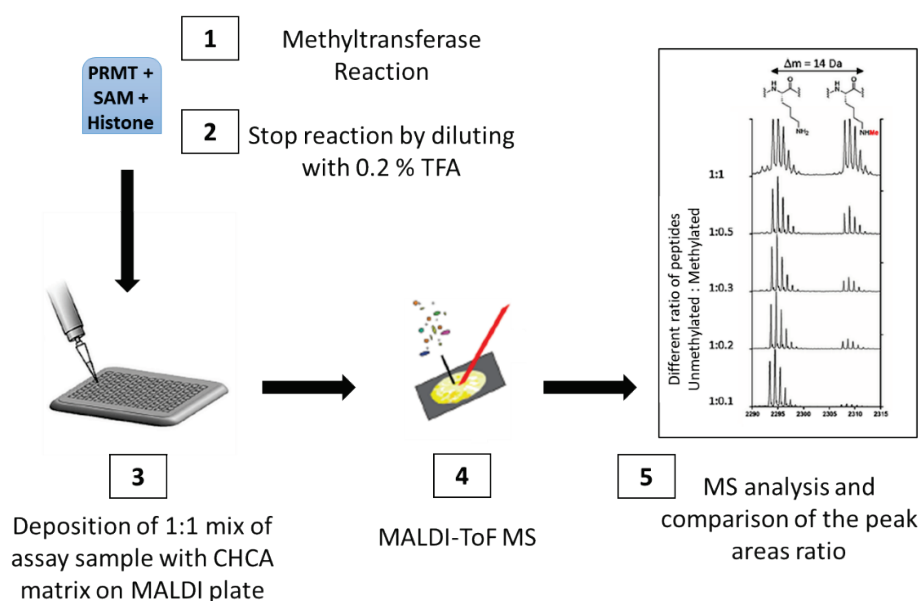


Figure 4-13 Principle of the Direct Label-free PRMT MALDI-ToF Mass Spectrometry Assay.

1) The enzyme is incubated with the histone peptide and SAM. 2) The reaction is then stopped by the addition of 0.2 % TFA. 3) An assay sample is mixed 1:1 with CHCA matrix and spotted on a MALDI AnchorChip using a Mosquito liquid handling robot (TTP Labtech). 4) MS analysis is performed on a RapifleX PharmaPulse MALDI TOF/TOF mass spectrometer using FlexAnalysis 4.0 software (5).

After the methyltransferase reaction is stopped, a 1:1 mix of enzyme assay sample and matrix is deposited onto a MALDI plate and analysed via MALDI-ToF Mass Spectrometry. The peak area of the [M+H]⁺ signals of the unmethylated peptides (A_{unmet}) and the methylated peptides (A_{met}) is recorded and the methylation ratio R is calculated using the following equation (Guitot *et al.*, 2017):

$$R = \frac{A_{met}}{A_{unmet} + A_{met}} \times 100$$

Human histone H3 (residues 22-45) and H4 (residues 2-22) peptides were chosen as substrates for PRMT2 and CARM1. Both were supplied with an N-terminal acetyl and C-terminal amide group and synthesised by Severn Biotech Ltd. (UK). HsPRMT2 has previously been shown to be active on an N-terminal tail of an H4 peptide via UPLC-MS/MS and thin-layer chromatography (TLC) (Lakowski and Frankel, 2009). For the catalytic domain of CARM1, that contains the homo-oligomerization domain, reduced but efficient methylation activity towards histone H3 has been reported (Teyssier *et al.*, 2002). Full-length CARM1 is able to methylate histone H3 at Arg2 (weak activity), Arg17, Arg26 and Arg42 (Schurter *et al.*, 2001, Casadio *et al.*, 2013), which is why the H3(22-45) peptide was chosen as substrate.

4.7 Purity Assessment of H3 and H4 Peptide

In the first step, the purity of the commercially supplied histone peptides was tested to ensure that they have > 95 % purity. Purity analysis was performed by Dr Rachel Heap (Newcastle University) and was determined via MALDI and ESI analysis. From the spectra (Figure 4-14), the purity of the H3 peptide was estimated to be <30 % which did not agree with the supplier's data (Appendix K). Different peptide degradation products or precursors for the peptide synthesis can be seen, for example shorter peptides missing the first (Ala, -71 Da), last (Gly, -57 Da) or both residues were detected. Upon arrival, the peptide was dissolved in sterile water, aliquoted and stored at -80 °C as instructed by the supplier. For analysis one aliquot was thawed on ice before spotting it directly on the MALDI plate. One explanation might be contamination and enzymatic digestions. However, after a new peptide from the same batch was supplied by the company, it was made up fresh and directly analysed but gave the same results. The assay substrate needs to be very pure to be able to detect and especially quantify the

amount of product. As a consequence, the peptide could not be used in this assay format.

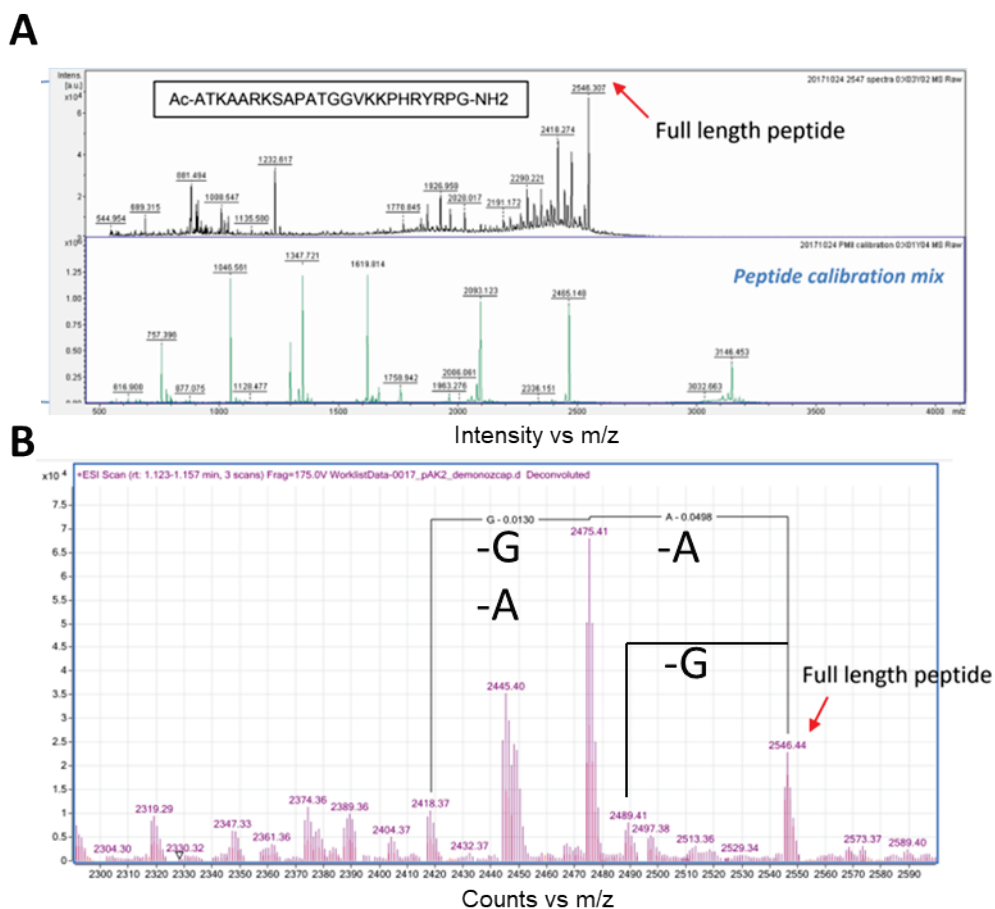


Figure 4-14 Purity Analysis of H3 (22-45) Peptide.

The MALDI (A) and ESI (B) spectra reveal a mass species at 2546.31 Da and 2546.44 Da corresponding to the full-length H3 peptide, which is N-terminally acetylated and C-terminally amidated and has an expected molecular weight of 2545.436 Da. The estimated purity is between 30-40 %.

Despite its low level of purity, the peptide was suitable for assay development while an alternative peptide source was identified. For the enzyme assay, the assay buffer previously used in the MTase Glo™ assay (20 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl₂, 0.1 mg/ml BSA, 1 mM DTT) was chosen but without the addition of DTT and MgCl₂ which are needed in the Glo Assay to ensure the stability and activity of the coupled enzymes. The original assay buffer contains 50 mM NaCl but salt can significantly affect peptide ionisation. Thus, the effect of different salt concentrations (0-40 mM NaCl) on the detection of 500 fmol H3 peptide was analysed (Figure 4-15 A). The results show that the assay is very sensitive, and, at the tested

peptide concentration, peptide intensity was almost not affected by the salt concentration. In addition, a peptide titration with and without 50 mM NaCl in the buffer was performed to obtain the detection limit of the H3 peptide (Figure 4-15 B). The peptide could be detected at a concentration of 128 amol if no NaCl was added to the buffer (blue curve). On the other hand, with 50 mM NaCl in the buffer (red curve), the detection limit was 16 fmol. However, as the salt concentration did not seem to affect the detection at higher peptide concentrations that will be used in the assay, 50 mM NaCl can be used in the assay buffer.

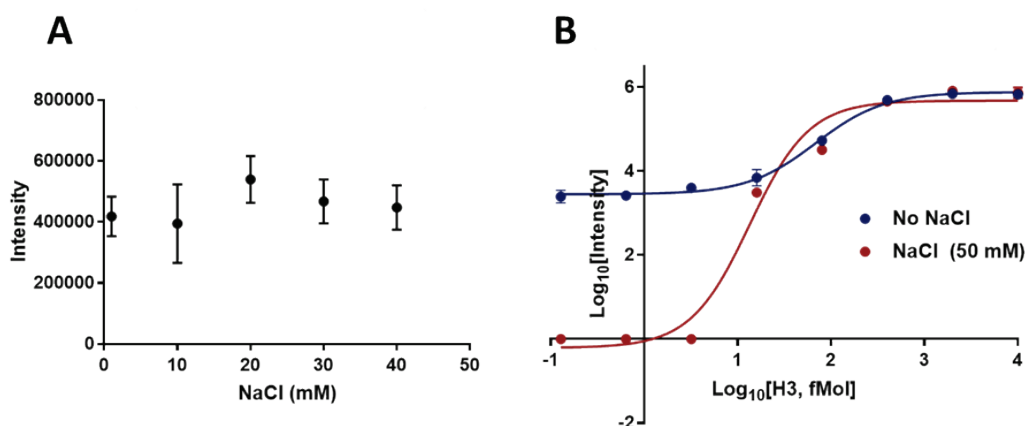


Figure 4-15 Detection of H3(22-45) Peptide by Mass Spectrometry.

Peptide was analysed in the presence/absence of NaCl. **A)** Effect of different NaCl concentrations on ionisation of 500 fmol H3 peptide. **B)** H3 peptide titration using buffer with and without 50 mM NaCl.

Despite the purity issues, the substrate was tested in a methylation assay using CARM1. The final reaction volume was 4 μ L and the reaction contained 800 nM CARM1, 10 μ M SAM and 1 μ M H3 peptide. Assay samples were analysed at two different time points, after 10- and 60-min incubation. The resulting MS spectrum can be seen in Figure 4-16. Methylation would result in a mass shift of 14 Da in the case of mono-methylation and 28 Da in the case of di-methylation. Moreover, the peak corresponding to the unmethylated H3 peptide (2544.592 Da) is expected to decrease with progression of the reaction. Two peaks were visible at the expected mass of the mono-methylated (2558.581 Da) and di-methylated (2572.594 Da) H3 peptides. Their intensity slightly increased over time, indicating that the purified catalytic domain of CARM1 is most likely active. Nonetheless, substrate conversion is low, even at high enzyme concentrations (800 nM). Despite the fact that methylation of the substrate occurred, all precursors of the peptide synthesis are present, and they seem to be methylated as

well. As a consequence, although the peak area ratio of the unmethylated and methylated full-length peptide substrate can be calculated, it does not represent all the modified peptides present and as a result the CARM1 activity cannot be quantitated. Moreover, it could be possible that the enzyme is not very active in the MS assay buffer, as it does not contain any salt, in contrast to the assay buffer used in the MTase Glo™ Assay.

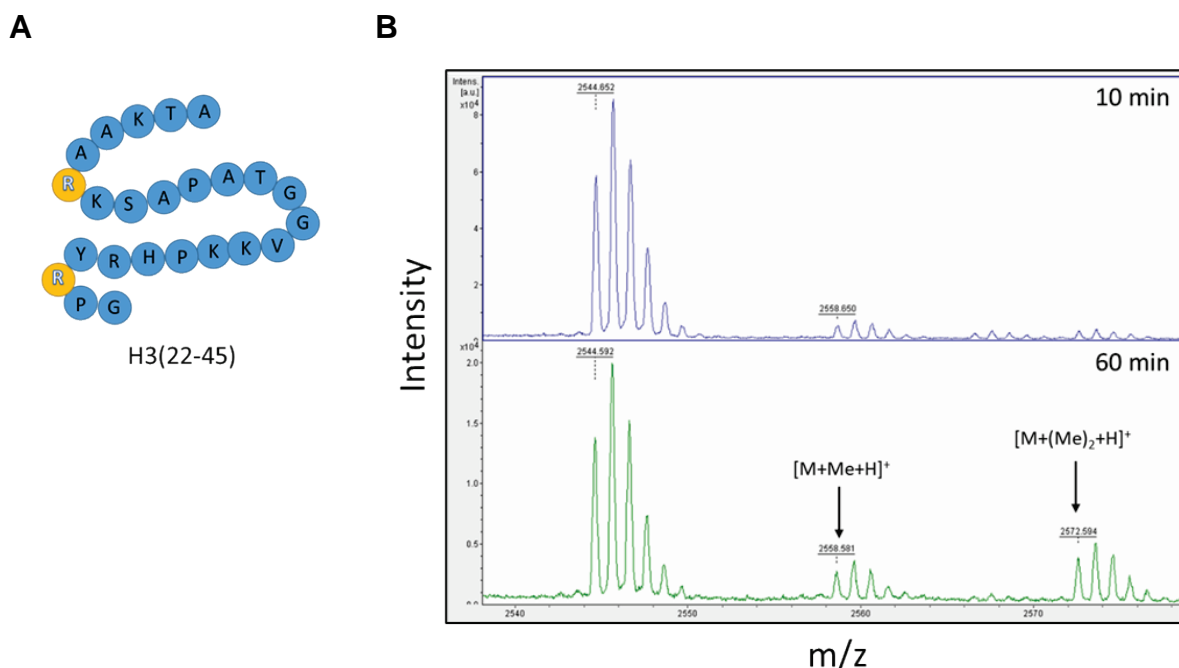


Figure 4-16 MALDI-TOF MS Spectra of CARM1 Activity Assay with H3(22-45).

A) Illustration of H3(22-45) peptide, putative arginine methylation sites are shown in yellow. **B)** 1 μ M of the H3 peptide (m/z 2544.592) was incubated with 800 ng CARM1 and 10 μ M SAM at RT in assay buffer (20 mM Tris pH 8.0, 1mM EDTA, 1 mM DTT, 0.1 mg/mL BSA). Samples were analysed after 10- and 60- min incubation time. Mono- (m/z 2558.581) and di-methylated peptide (m/z 2572.594) were detected.

To analyse the effect of salt in the reaction buffer on CARM1 activity, the assay was performed with and without 50 mM NaCl. The spectra after an incubation time of 1 h can be seen in Figure 4-17. The peptide is methylated in both assay conditions but there is more conversion of substrate if using the assay buffer containing salt.

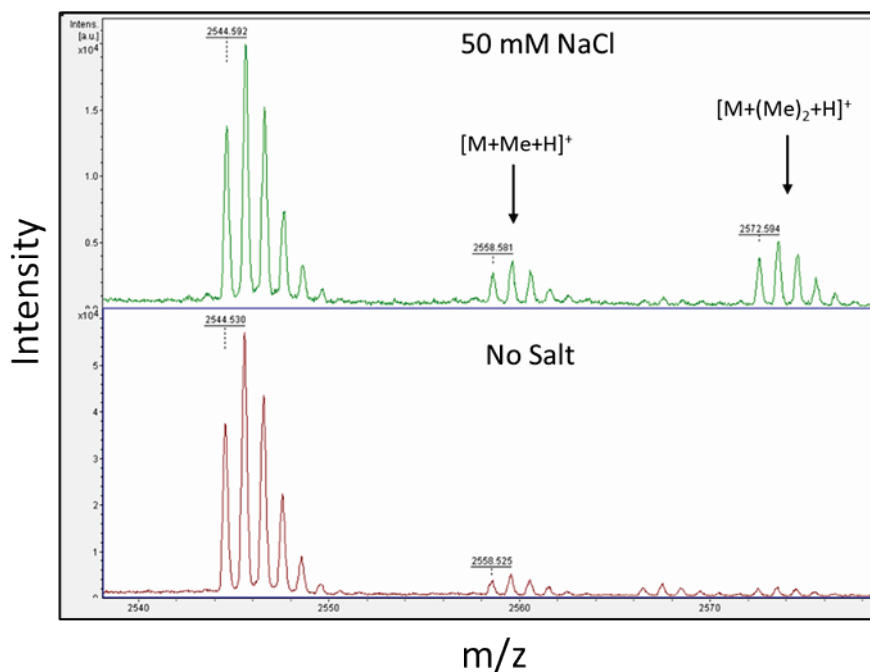


Figure 4-17 Effect of Sodium Chloride on CARM1 Activity.

1 μM of the H3 peptide (m/z 2544.592) was incubated with 800 ng CARM1 and 10 μM SAM at RT in assay buffer (20 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA) with and without 50 mM NaCl. Samples were analysed after 60 min incubation time. Mono- (m/z 2558.581) and di-methylated peptide (m/z 2572.594) were mainly detected if salt was present in the buffer.

The assay was repeated with HsPRMT2-1 using the H4 peptide as the substrate. However, PRMT2 methylation activity could not be detected. Because of the limited amounts of purified PRMT2, it was decided to first focus the assay development using CARM1. A new H3 peptide was ordered which showed a much higher purity (Figure 4-18).

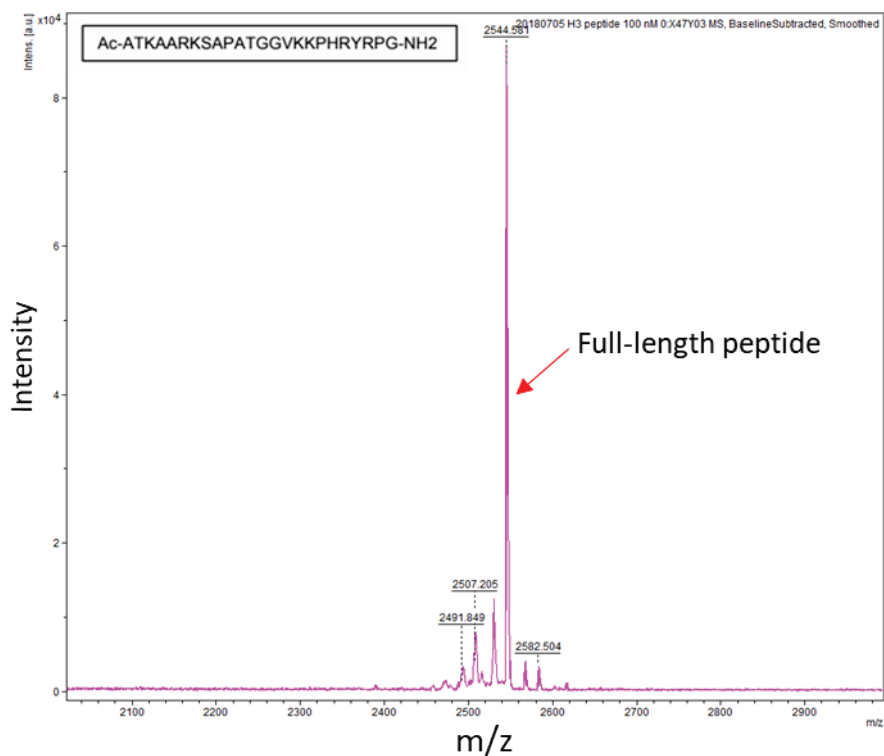


Figure 4-18 Purity Analysis of H3(22-45) Peptide.

The spectrum reveals a mass species at 2544.58 Da corresponding to the full-length H3 peptide, which is N-terminally acetylated and C-terminally amidated and has an expected molecular weight of 2545.436 Da. The estimated purity is >90 %.

The effects of different NaCl concentrations in the assay buffer on H3 peptide ionisation were again analysed (Figure 4-19 A) and in addition a peptide titration in water and buffer with and without 50 mM NaCl was performed (Figure 4-19 B). Only a small amount of the peptide was sodiated, even at high NaCl concentrations and the salt concentration of the assay buffer only slightly affected the H3 peptide peak area/ionisation. Thus, it was decided to use 50 mM NaCl in the assay buffer.

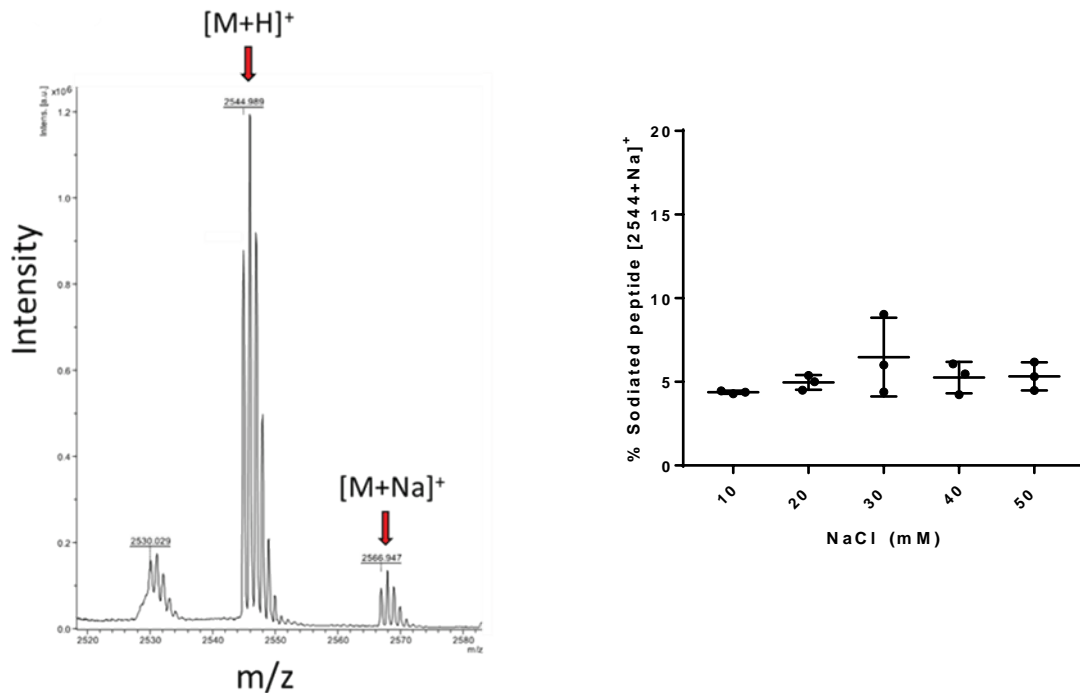
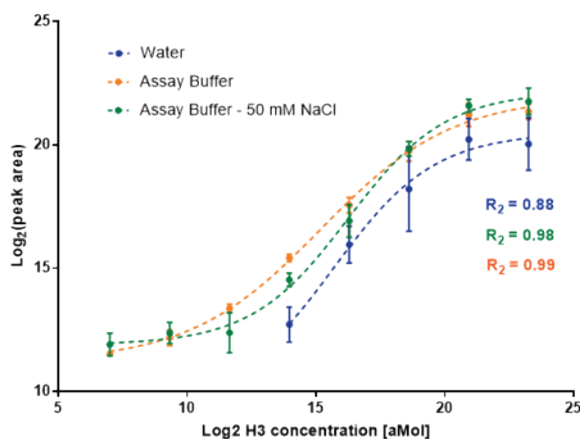
A**B**

Figure 4-19 Effects of Sodium Chloride on H3 Peptide Ionisation.

A) MS spectrum of histone H3(22-45) peptide in buffer containing 50 mM NaCl, indicating sodiated species at m/z 2566.947 for [M +Na]⁺. The percentage of the sodiated species is low, even at high NaCl concentrations. **B)** Effect of different assay buffers on the ionisation of the H3 peptide.

However, despite testing it with the same protein preparation under the previously used assay conditions, no methylated H3 peptide could be detected in the MALDI-ToF assay samples. The assay was repeated with freshly purified enzyme, at two different temperature points (25 °C and 37 °C), and over a time course of 60 min but still no methylation could be detected. The peptide was then tested in the MTase Glo™ assay and compared with full-length histone H3. The results can be seen in Figure 4-20. MmCARM1 is known to auto-methylate at a single site, Arg551 (Arg550

in HsCARM1) (Kuhn *et al.*, 2010), however, the assayed CARM1 construct does not include this residue.

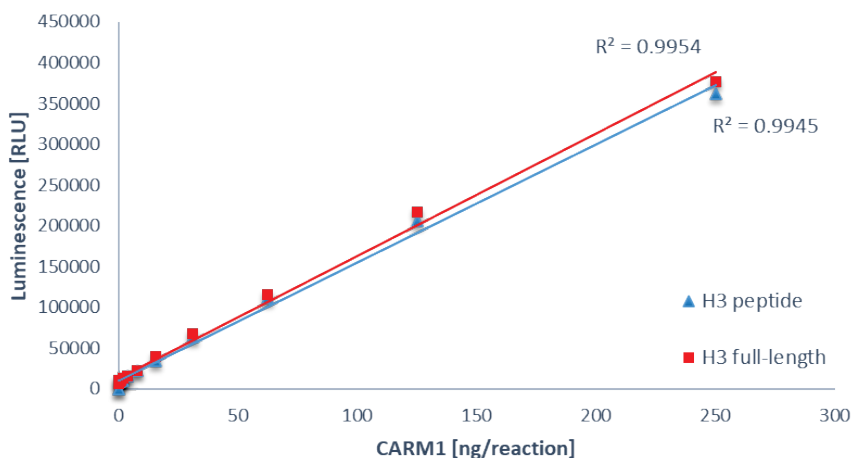


Figure 4-20 Measurement of CARM1 Activity using the MTase Glo Assay.

CARM1 was assayed in the presence of 1 μ M SAM and full-length histone H3 (red) or H3-derived peptide (22-45, blue) as substrate. Luminescence was measured with a PHERAstar FS (BMG LABTECH) plate-reader and data were analysed in Excel using relative luminescence units (RLU). The linear fitting curves are shown in red and blue, respectively.

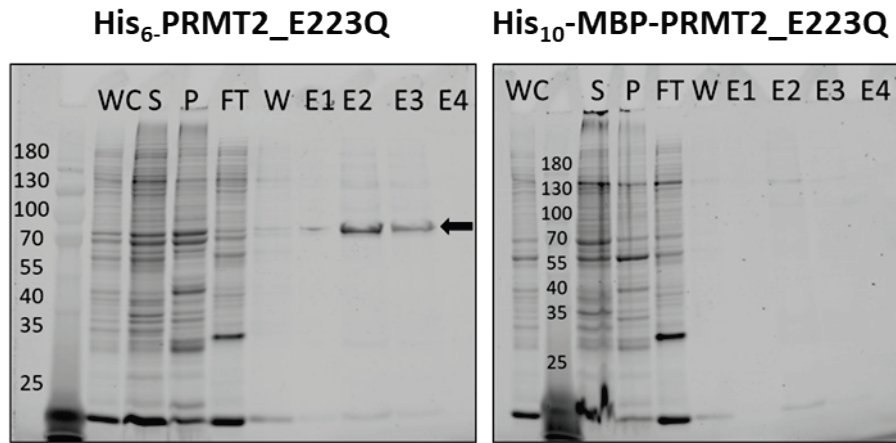
To see whether the CARM1 catalytic domain has activity towards histone H3 Arg17, a commercially available H3(1-22) peptide with an expected molecular weight of 2623.536 Da, carrying a C-terminal amide group, was bought from Anaspec and tested in a methylation assay with HsCARM1. Assay samples were analysed with targeted MS/MS by Dr Claire Jennings. The substrate H3 peptide was identified with a mass of 2624.531 Da and also peaks corresponding to the mono- and di-methylated H3 peptide at 2638.5467 Da and 2652.562 Da were detected. Assay development is currently ongoing.

4.8 Production of a Catalytical Inactive HsPRMT2 Mutant

In order to prepare a catalytical inactive PRMT2 enzyme, the second invariant glutamic acid residue of the double-E loop was mutated to glutamine (E223Q) as described in Sections 2.16.6 and 2.4.7. Mutagenesis of this residue in other PRMTs including PRMT1 leads to complete loss of enzyme activity while maintain SAM binding (Zhang and Cheng, 2003).

An SDS-PAGE analysis of the purification of His₆-HsPRMT2_E223Q and His₁₀-MBP-HsPRMT2_E223Q from 300 mL of *Sf9* insect cell culture can be seen in Figure 4-21. The cells were lysed via freeze-thaw cycles, and the cleared lysate purified with TALON IMAC (panel A). The eluted tagged protein was then cleaved with His-3C-protease, followed by subtractive TALON IMAC (panel B). In the case of His₁₀-MBP-PRMT2_E223Q there was no detectable expression. However, for His₆-tagged protein, protein bands with the expected molecular mass were visible, indicated by a black arrow. The flow through and wash was concentrated to 1 mg/mL resulting in a final yield of 0.32 mg from 300 mL *Sf9* cells. LC-MS analysis was carried out and an intact mass of 49195.15 Da detected (Figure 4-22). This is in good agreement with the expected mass of HsPRMT2_E223Q of 49195.35 Da. Additionally, a species with a mass of 49271.58 Da was detected. The mass difference of 76.2 Da suggests that the protein was post-translational modified and mono-acetylated (+42.0376 Da). Due to ongoing assay development the inactive mutant has not yet been tested in the activity assay.

A



B

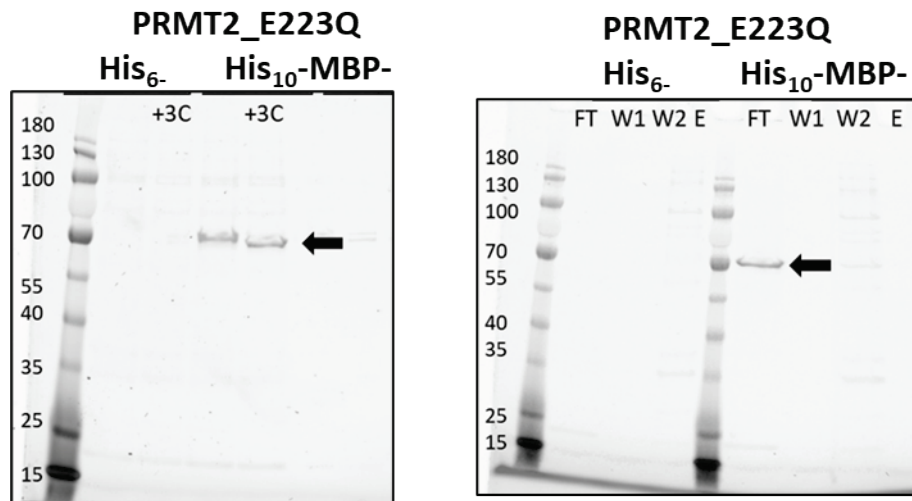


Figure 4-21 SDS-PAGE Analysis of the Purification of HsPRMT2_E223Q.

A) SDS-PAGE analysis of TALON-IMAC purification. The bands corresponding to His₆-PRMT2_E223Q in the elution and wash fractions are highlighted by an arrow. No protein was observed for the His₆-MBP-tagged protein (A, Gel 2). His-tags were removed with His-3C-protease cleavage (**B**, Gel 1). After tag removal, the cleaved protein, highlighted by an arrow, was separated from the tags and uncleaved protein via subtractive TALON-IMAC (**B**, Gel 2). Flow-through and Wash fractions were combined and concentrated.

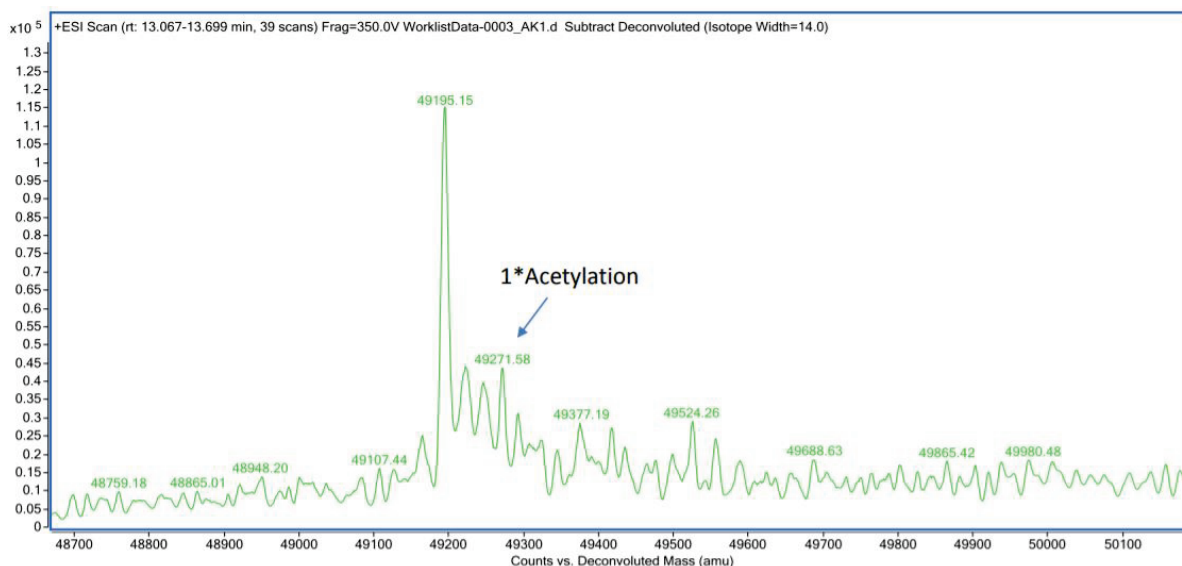


Figure 4-22 Analysis of HsPRMT2_E223Q Sample by Mass Spectrometry.

The deconvoluted spectrum reveals a dominant mass species at 49195.15 Da corresponding to HsPRMT2_E223Q, which has an expected molecular weight of 49195.35 Da.

4.9 Summary

Alteration of the enzyme activity or aberrant expression of PRMTs, proteins that catalyse the methylation of arginine residues on histones and non-histone proteins, has been observed in many human diseases including cancer (Yang and Bedford, 2013). As a consequence, different PRMT family members have emerged as potential drug targets and inhibitors are being developed (Smith *et al.*, 2018). Despite the interest in PRMTs, the knowledge about their biology, regulation, and action is still limited. This is especially the case for the lesser studied family members which include PRMT2. One of the reasons is that selective, cell-permeable chemical probes that can act as tools for target validation and potential start points for drug discovery are missing.

The project's aim was to develop such chemical probes for PRMT2, which are selective against other family members. In a first step, a commercial luminescence based enzymatic activity assay, the MTase Glo™ assay (Hsiao *et al.*, 2016), was used to characterise the activities of the enzymes. For all previously purified enzymes, full-length PRMT2, PRMT1, and the catalytic domain of CARM1, enzyme activity was successfully demonstrated using full-length histones as substrates. Interestingly, a shorter PRMT2 construct, that is missing its N-terminal SH3 domain, also showed enzyme

activity in the MTase Glo™ assay. Previously, it had been suggested that the SH3 domain of PRMT2 is necessary for enzyme activity (Lakowski and Frankel, 2009).

One drawback of the MTase Glo Assay was that high background signals were observed for PRMT2 and CARM1, most likely due to auto-methylation activity. For characterisation of PRMT1 and 2, K_d measurements for the histone substrates were performed, but were only successful for PRMT1. It was not possible to measure any K_d for PRMT2 and histone H4. This might suggest that PRMT2 has no activity on histone H4 and that the measured enzyme activity is due to auto-methylation activity, which has been previously observed in other activity studies (Cura *et al.*, 2017).

Four potential new methylation sites on PRMT2 were identified but experimental repeats were not possible due to limited protein amounts. Recently, it has also been shown that PRMT2 is able to di-methylate Arg8 on histone H3 (Dong *et al.*, 2018). This modification was not known when the enzyme activity assays were performed but it would make sense to assay PRMT2 with the full-length histone H3 or the H3 peptide used for CARM1 and PRMT5 activity assessment in the future. Additionally, two other potential PRMT2 substrates, GST-GAR and a newly identified protein RSF1 were successfully produced and PRMT2 activity was observed using RSF1 as substrate in the MTase Glo™ Assay. Unfortunately, limited protein amounts made further activity quantification and assay development of PRMT2 impossible.

Next, the development of a MALDI-TOF based activity assay for direct detection of arginine methylation on histone peptides was attempted. This technique distinguishes substrate from auto-methylation activity. However, assay development was complicated by purity issues of the histone peptides. Moreover, CARM1 activity was only seen for the first histone H3 peptide but not when using the purer second peptide. To rule out that buffer differences are the reason for the inactivity, the CARM1 should be tested in the MS assay with the same buffer used in the MTase Glo Assay. Additionally, full-length CARM1, which is commercially available, could be tested to see whether the assay set up inhibits enzyme activity or if the CARM1 catalytic domain is not active on the H3 peptides.

Chapter 5 Structural Characterisation of PRMT2

Protein Arginine Methyltransferase 2 (PRMT2) belongs to the Type 1 PRMT family class that is capable of mono-methylating and asymmetrically di-methylating arginine residues (Bedford, 2007). In addition to its catalytic domain, it contains an N-terminal SH3 binding domain (Figure 1-7). The PRMT2 gene HRMT1L1 was first identified by Katsanis *et al.* in 1997 as a homolog of rat PRMT1 when screening the database for expressed sequence tags (EST) (Katsanis *et al.*, 1997). Five distinct isoforms of PRMT2 have since been identified, the full-length protein and four truncated isoforms produced by alternative splicing at the 3' end of the pre-mRNA in exon 7-10 or alternative polyadenylation (Figure 5-1, Baldwin *et al.*, 2014).

PRMT2 has weak methyltransferase activity on histone H4 *in vitro* (Lakowski and Frankel, 2009) and is able to asymmetrically di-methylate Arg8 of histone H3 *in vivo* after recruitment to target promoters by β -catenin (Dong *et al.*, 2018). PRMT2 downregulation positively influences several oncogenic signalling pathways such as PI3K-AKT, MAPK, JAK-STAT, and Wnt/ β -catenin pathways (Dong *et al.*, 2018). PRMT2 expression was shown to be upregulated in glioblastoma. Inactivation of PRMT2 inhibited glioblastoma cell growth and stem cell renewal *in vitro* and also suppressed tumour growth (Dong *et al.*, 2018). The methylation activity of PRMT2 is low compared to PRMT1, but more research is needed to identify more specific *in vivo* PRMT2 substrates (Lakowski and Frankel, 2009).

PRMT2 has been identified as a potentially interesting target for oncology as it can interact with a number of receptors including the androgen receptor (AR) (Meyer *et al.*, 2007), estrogen receptor alpha (ER α), progesterone receptor (PR), peroxisome proliferator-activated receptor γ (PPAR γ) and retinoic acid receptor (RAR α) (Qi *et al.*, 2002) where it acts as transcription co-activator. Moreover, it was shown that PRMT2 inhibits NF- κ B signaling and induces apoptosis (Ganesh *et al.*, 2006). Additionally, PRMT2 is thought to have a regulator role in Wnt/ β -catenin (Blythe *et al.*, 2010) and leptin signaling (Iwasaki *et al.*, 2010). Nuclear loss of PRMT2 has also been linked to tumor grade and overexpression of cyclin D1 in breast cancers (Zhong *et al.*, 2014). The PRMT2-ER α interaction will be described in more detail as an example of its involvement in receptor signaling pathways.

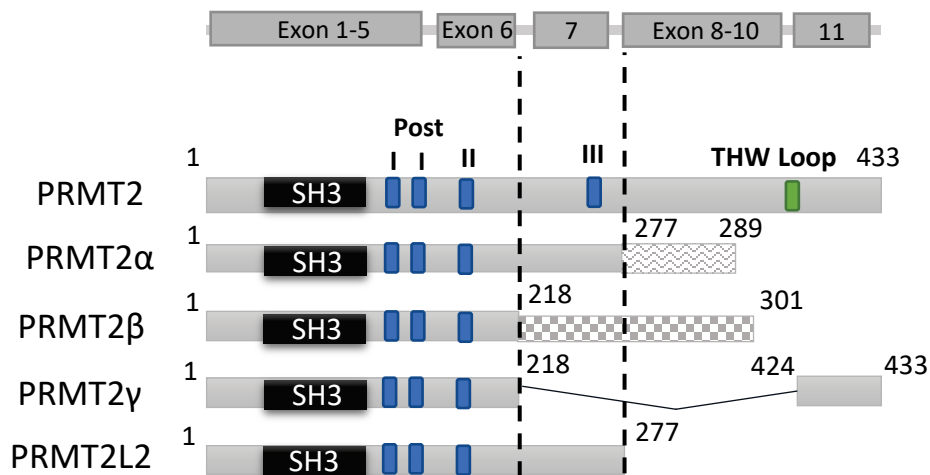


Figure 5-1 Overview of PRMT2 Isoforms.

PRMT2 is made up of 11 exons. *PRMT2L2* is a result of alternative polyadenylation that results in premature termination codon after exon 7. *PRMT2α*, β and γ are produced by alternative splicing at the 3' end. Exon 8 -10 are deleted in *PRMT2α* and a frame shift results in 12 new amino acids at the C-terminus (shaded). *PRMT2β* has a deletion of exon 7-9, resulting in a frame shift and a new specific sequence of 72 amino acid residues (shaded). In the case of *PRMT2γ* an in-frame deletion of exon 7-10 occurred (Baldwin *et al.*, 2014).

Two studies conducted by Zhong *et al.* (2012, 2011) identified additional *PRMT2* isoforms beside the original full-length form identified in 1997 by Katsanis *et al.*, named *PRMT2L2*, *PRMT2α*, *PRMT2β*, and *PRMT2γ*. An overview of the different isoforms is given in Figure 5-1. The isoforms differ in their sequence in exon 7-10. The first alternatively spliced *PRMT2* variant reported was *PRMT2L2* that is produced via alternative polyadenylation which results in an early transcription termination after exon 7 (Zhong *et al.*, 2011). The three truncated isoforms *PRMT2α*, *PRMT2β*, and *PRMT2γ* are a result of exon deletions and were isolated from breast cancer cells (Zhong *et al.*, 2012). Two of the splice variants, *PRMT2α* and β have unique C-terminal sequences as a result of the frame shifts. All four truncated *PRMT2* isoforms lack the THW-loop and *PRMT2α*, *PRMT2β*, and *PRMT2γ* also lack the conserved motif III but all contain the N-terminal SH3 domain. As described earlier (Section 1.4), the THW-loop is involved in cofactor binding and motif III is important for fold stability (Figure 1-9). Thus, loss of these motifs most likely results in the loss of their methyltransferase activity. However, enzyme activity has not been assessed yet. Studies with GFP-tagged *PRMT2* isoforms showed that *PRMT2*, *PRMT2α* and *PRMT2γ* are mainly found in the nucleus, but not in the nucleoli (Baldwin *et al.*, 2014). *PRMT2L2* is predominately localised in the cytoplasm and *PRMT2β* is evenly distributed (Zhong *et al.*, 2011).

Expression levels of each isoform differed across different breast cancer cell lines and were increased in ER and PR-positive breast cancer cells (MCF7, BT474, ZR-75-1) compared to double-negative cells (MDA-MB-231, MDA-MB-453 and SK-BR-3) (Zhong *et al.*, 2011, Zhong *et al.*, 2014). Moreover, mRNA levels of all PRMT2 splice variants were increased in breast tumour tissue compared to normal breast tissue, accompanied by an increase in overall PRMT2 protein expression level (Zhong *et al.*, 2012). Additionally, PRMT2 expression levels were higher in ER-positive tumour samples compared to ER-negative samples. Protein-protein interaction studies by Zhong *et al.* showed that all PRMT2 isoforms were able to bind the ER α receptor via their N-terminal SH3 domain *in vitro* and in MCF7 cells in the presence of estrogen. Moreover, binding to the AR receptor was also observed in HepG2 cells in the presence of dihydrotestosterone (Zhong *et al.*, 2011). The transcription factor Snail is known to repress E-cadherin which results in epithelial to mesenchymal transition (EMT) (Batlle *et al.*, 2000), and upregulation of Snail is often found in ovarian carcinoma (Elloul *et al.*, 2006). It was also shown that Snail can act as a repressor of E-cadherin upon estrogen-induced tumour cell migration and invasion in ovarian cancer cells (Barrallogimeno and Nieto, 2005). To assess whether the PRMT2 variants have effects on ER α target genes, luciferase promoter assays with Snail and E-cadherin were performed. It showed that all isoforms enhance promoter activities of Snail in the presence of estrogen, and decrease promoter activity of E-cadherin in the presence of estrogen and in the case of PRMT2 α and PRMT2 γ also without estrogen (Zhong *et al.*, 2012). The fact that the isoforms not only differ in their subcellular location but also in binding affinity and transcriptional activity of receptors, might indicate that they have different functions.

Knockdown of all four truncated isoform mRNAs increased cell proliferation, cell formation and E2 factor (E2F) expression in ER α -positive breast cancer cells and PRMT2 depletion in a xenograft mouse model increased tumour growth (Zhong *et al.*, 2011). This is in agreement with an earlier study, which showed that full length PRMT2 can repress E2F transcriptional activity in a retinoblastoma gene product (RB)-dependent manner (Yoshimoto *et al.*, 2006). The subsequent study (Zhong *et al.*, 2014) showed that the tumour growth suppression in ER α positive breast cancer cells is regulated by PRMT2 via the E2F/cyclin D1 pathway. PRMT2 downregulates cyclin D1 and decreases the affinity between ER α and its activator protein 1 (AP-1) site through indirect binding to the AP-1 site on the cyclin D1 promoter. Additionally, it was shown that

nuclear loss of PRMT2 correlates with increased cyclin D1 levels and therefore PRMT2 might drive breast tumour cell proliferation. An additional study by the same group (Zhong *et al.*, 2017), confirmed that the splice variant PRMT2 β has anti-tumour effects. PRMT2 β suppressed cell proliferation and colony formation of breast cancer cells and it was able to inhibit Akt/GSK-3 β signalling activity resulting in suppression of cyclin D1 expression.

More research is needed to determine the specific function of each isoform and its involvement in cancer. However, alternative splicing of PRMT2 and nuclear loss might be involved in breast tumorigenesis. The fact that these alternative splice variants are lacking different conserved motifs from the catalytic domain including motif III and the THW-loop and thus are most likely inactive, might suggest an involvement of their SH3-domain which is present in all isoforms (Figure 5-1). Despite the involvement of PRMT2 in many different cellular processes the protein is still not well characterised and not much is known about its substrate recognition.

5.1 Overall Structure of PRMT2

The first crystal structures of mouse and zebrafish PRMT2 were published in 2017 by Cura *et al.* (2017). However, although the full-length proteins were crystallised, subsequent structure determination revealed no electron density to support the presence of the N-terminal sequences that include the SH3 domain. The structures of mouse (PDB: 5FUL) and zebrafish (PDB: 5FUB) PRMT2 include residues 107-445 and 72-408 respectively.

Structure based sequence alignment of PRMT2 with other known family members showed that the catalytic core is highly conserved (Figure 5-2). The core consists of the cofactor-binding domain (green) comprising residues 107-254 (MmPRMT2 numbers) and includes the three invariant motifs. It is followed by the β -barrel domain (residues 255-265 and 299-455, yellow) that includes the THW loop and which is interrupted by the dimerisation arm (residues 266-298, purple). The catalytic core of PRMT2 shares the highest sequence identity with PRMT6 (39 %) but is also very similar to the other Type 1 PRMTs including PRMT8, PRMT1, PRMT3, and CARM1 reaching from 35-37 % (Cura *et al.*, 2017). In contrast to PRMT6, PRMT2 contains an additional N-terminal SH3 domain. In the following paragraphs the SH3 domain of PRMT2 will be discussed after a short general overview of the domain type

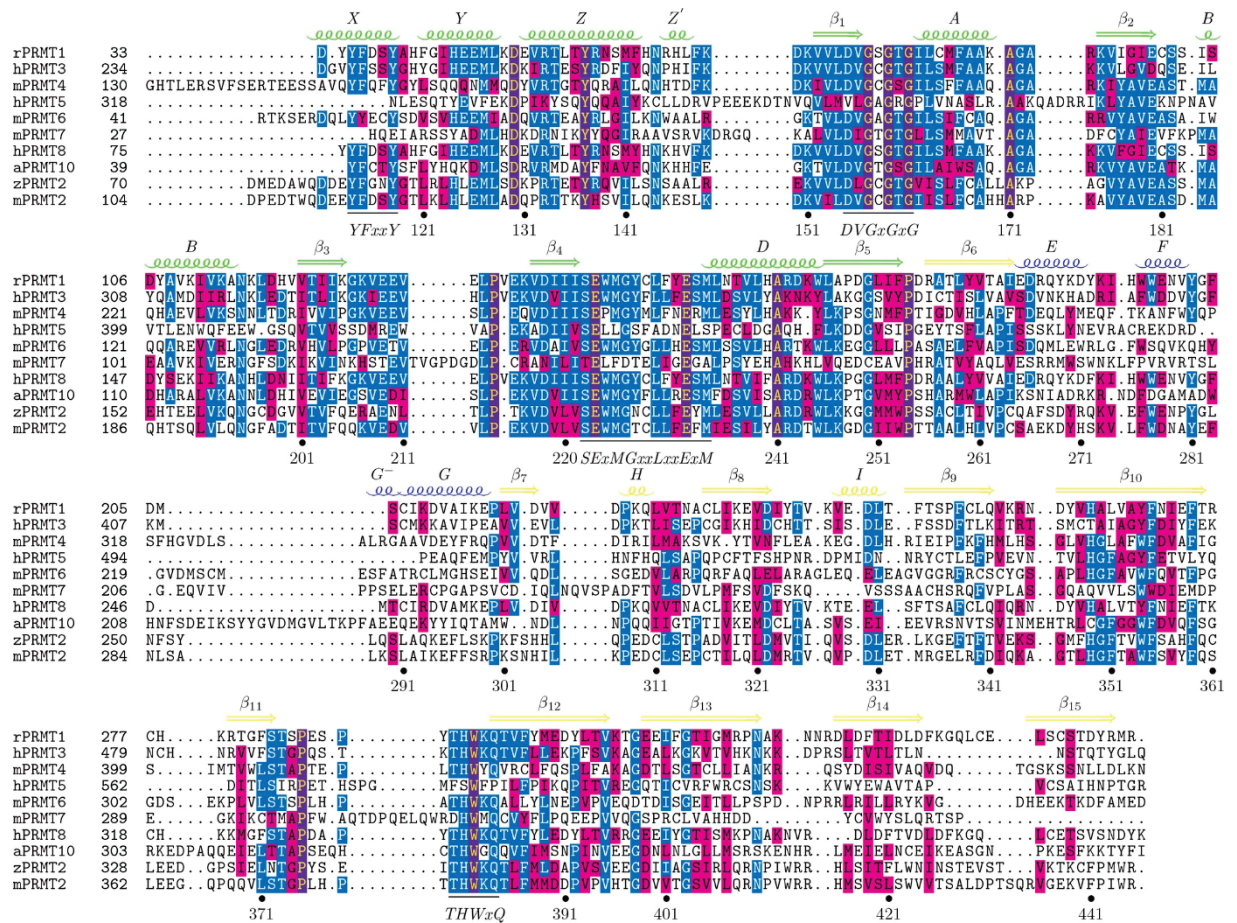


Figure 5-2 Structure-based Sequence Alignment of the Catalytic Core of Different PRMTs.

Sequence alignment was prepared using the crystal structures of RnPRMT1 (PDB: 1R08), HsPRMT3 (PDB: 2FYT), MmCARM1 (PDB: 5K8V), HsPRMT5 (PDB: 4X61), MmPRMT6 (PDB: 6P7I), MmPRMT7 (PDB: 4C4A), HsPRMT8 (PDB: 5DST), AtPRMT10 (PDB: 6PDM), DrPRMT2 (PDB: 5FUB) and MmPRMT2 (PDB: 5FUL). The secondary structure of the cofactor binding domain (green), β-barrel domains (yellow) and the dimerisation arm (purple) is shown above the sequence. MmPRMT2 residue numbers are shown below the sequence. The four conserved motifs are highlighted. Amino acids are coloured according to their similarity: Invariant (violet), similar (blue). Abbreviations: Hs= Homo sapiens; Mm= Mus musculus; Rn= Rattus norvegicus; At= Arabidopsis thaliana. Figure taken, with permission, from Cura et al. (2017).

5.1.1 Overview of SH3 Domains

The first SH3 domain was described in 1988 as a region of high sequence similarity within different signalling proteins including the Src family of tyrosine kinases, the Crk adaptor protein, and phospholipase C-γ (Mayer et al., 1988). The first crystal structure of an SH3 domain, that of human tyrosine-protein kinase Fyn was solved by Noble et al. in 1993 (Noble et al., 1993). Until now, many SH3 domain structures have been identified in signalling proteins and many of their structures have been solved by NMR and X-ray crystallography, as reviewed by Kaneko et al. (2008). SH3 domains are 55-

70 amino acid residues in size and are involved in a variety of cellular processes including signal transduction mechanisms and cell-cell communication (Pawson, 1995, Kurochkina and Guha, 2013). They all share a common fold consisting of five β -strands which form two anti-parallel β -sheets (Figure 5-3). The loops connecting these strands are named the Arginine-Threonine (RT) loop, the neuronal Src (n-Src) insertion point loop and the distal loop (Koch *et al.*, 1991, Kurochkina and Guha, 2013).

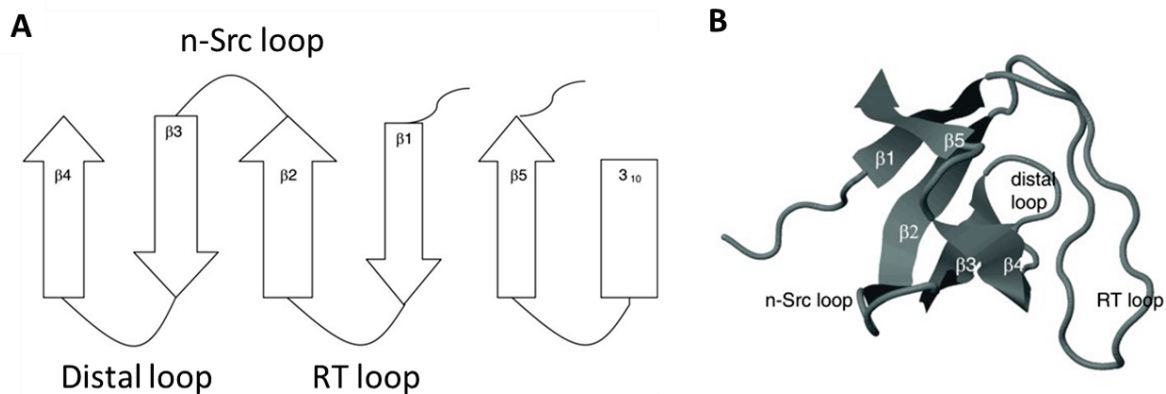


Figure 5-3 Structural Architecture of SH3 Domains.

The topology of secondary structure elements (A) and the tertiary fold (B) of SH3 domains is shown. The SH3 domain consists of a β -sandwich consisting of five β -strands that are connected by three loops and a short 3_{10} helix. Figure adapted, with permission from Kurochkina and Guha (2013).

SH3 domains facilitate protein-protein interactions by recognizing seven to nine amino acid long proline-rich core regions on cellular proteins (Saksela and Permi, 2012). Most SH3 domains bind to proteins harbouring a PxxP motif where P is proline and x represents any amino acid and mainly those that have a left-handed helix with three residues per turn, known as polyproline type II (PPII) conformation (Yu *et al.*, 1994, Saksela and Permi, 2012). It was shown that the PxxP motif binds in distinct hydrophobic binding pockets on the SH3 domain surface, illustrated in Figure 5-4.

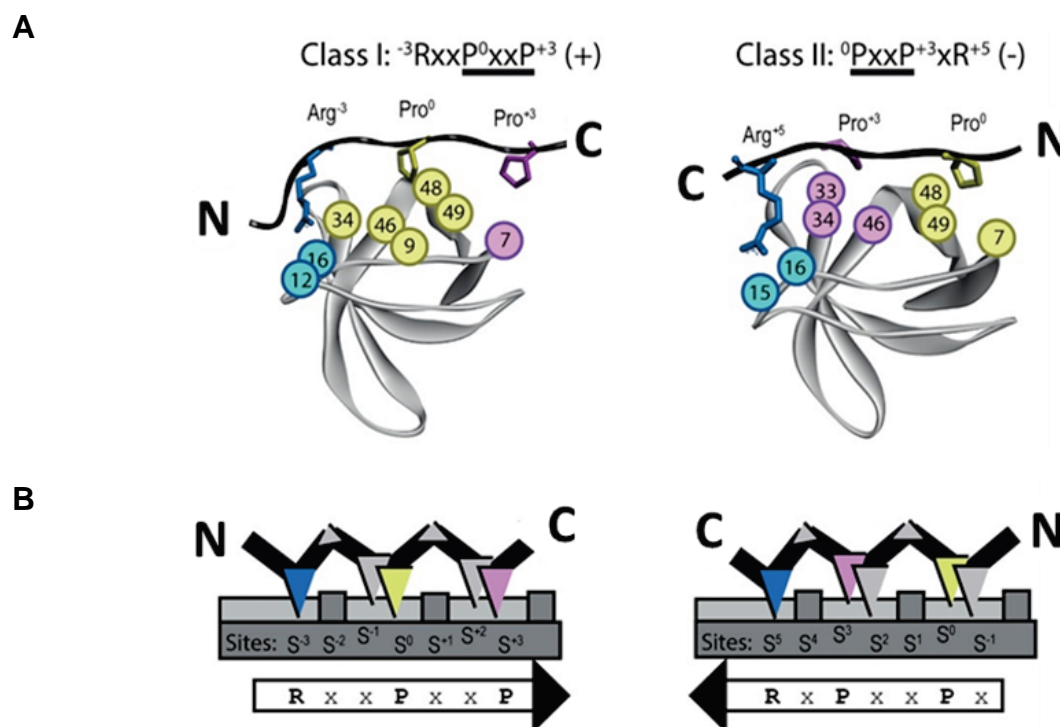


Figure 5-4 SH3 Domain Binding to a Class I or Class II Peptide.

A) SH3 domain crystal structures in complex with Class I (PDB: 2D1X) or Class II ligands (PDB: 3U23). The side chains of the peptides that interact with the hydrophobic SH3 domain binding pockets are shown as coloured sticks. Residues that interact with the peptide are shown as spheres. The relative peptide orientation is shown with a “plus” (+) or “minus” (-) and the amino- (N) and carboxyl-termini (C) are indicated. **B)** Schematic representation of the binding sites. Class I peptides bind in a plus orientation (C ← N), whereas Class II ligands bind in a minus orientation (N → C). The PxxP motif of the consensus sequence is underlined. Figure adapted, with permission, from Teyra *et al.* (2017).

The first specific and closest to the N-terminus located proline residue of the motif is named Pro⁰, and its binding site S⁰. The other specific proline residues/binding sites are named negatively toward the N-terminus and positively towards the C-terminus. Two different orientations of the peptide containing the PxxP helix are possible due to its pseudo-symmetry. They depend on the position of a positively charged residue R (R is arginine or lysine), which interacts with a negatively charged third binding pocket also known as the specificity pocket (Feng *et al.*, 1994). This pocket is formed by residues from the RT loop and n-Src loop (Yu *et al.*, 1994). The target proteins and peptides of SH3 domains can be grouped into three classes according to their consensus motifs or canonical sequences: Class I domains bind ligands that conform to the consensus motif $^{-3}\text{RxxP}^0\text{xxP}^{+}$ in a plus (+) orientation. In this orientation the proline residues interact with two external hydrophobic sites near the RT loop, the S⁰ site is proximal and the S⁺³ site distal to the specificity site. In contrast, Class II SH3 domains

recognise peptides with a $^0\text{PxxP}^3\text{xR}^5$ motif in a minus (-) orientation. The proline residues interact with two internal hydrophobic sites near the n-Src loop and the 3_{10} helix, the S^{+3} is proximal and S^0 site distal to the to the specificity site (Yu *et al.*, 1994, Feng *et al.*, 1994, Lim *et al.*, 1994).

The human genome encodes more than 300 SH3 domains, all with a similar fold and with only mediocre binding affinities ranging from 1-100 μM (Mayer, 2001, Lee *et al.*, 2002, Kärkkäinen *et al.*, 2006). Thus, the question arises how the high target specificity of SH3 domains observed *in vivo* is achieved. It was shown that other non-consensus ligands do exist (Saksela and Permi, 2012). An analysis of more than 100 human SH3 domains showed that more than 50% of them exhibit such non-canonical specificities (Teyra *et al.*, 2017). In some cases, the ligands were still recognised by the SH3 domains via two proline residues but the spacing between the two differed from the canonical sequence (Hoelz *et al.*, 2006). Other SH3-ligands exists that only have one proline residue but additional binding sequences, such as the SH3 domain of GRAP2 that interacts with LCP2 (Lymphocyte Cytosolic Protein 2) via a $3\text{RxxP}^0\text{xxxP}^{4(+)}$ motif (Liu *et al.*, 2003). It was also shown that SH3 domains are able to bind proteins or peptides that lack the PxxP motif (Mongioví *et al.*, 1999, Kang *et al.*, 2000) and can interact with two ligands at the same time (Douangamath *et al.*, 2002).

For example, the SH3 domain of tyrosine kinase FynT binds the SH2 domain of SAP (SLAM-associated protein) only via non-canonical surface interactions (Chan *et al.*, 2003). Moreover, the peroxisomal Pex13p SH3 domain was shown to bind two ligands at the same time, one via the PPII helix, the other one via its α -helix (Douangamath *et al.*, 2002). Additionally, protein dimerisation via the SH3 domains has been observed (Nishida *et al.*, 2001, Levinson *et al.*, 2009). Thus, a number of different factors outside the known PxxP motif influence specificity and affinity of SH3 domains (Teyra *et al.*, 2017).

Protein-protein interactions that are mediated by SH3 domains have also been involved in a variety of diseases, including cancer and neurodegenerative diseases, such as Huntington disease (Gao *et al.*, 2006). For example, different proteins that are implicated in Alzheimer's disease interact through their SH3 domain with the PxxP motifs of the Tau protein which is strongly implicated in the disease (Lee, 2005). One of these proteins is Fyn kinase and the protein-protein interactions is a potential drug target (Lau *et al.*, 2016). Very recently, a peptide inhibitor of the Tau-SH3 interactions

has been reported that reduced amyloid- β toxicity (Rush *et al.*, 2020). Another example for targeting SH3 domain interactions, is the CT-10 regulator of kinase protein II (Crk II), which is involved in a large number of cancers including aggressive lung, breast, and ovarian cancers (Bhatt *et al.*, 2016). Crk II has an N-terminal and C-terminal SH3 domain and the N-terminal domain interacts with proline rich motifs of cAbl kinase positively influencing cancer metastasis (Donaldson *et al.*, 2002, Bhatt *et al.*, 2016). Successful targeting of a SH3 domain with a covalent inhibitor has been shown for the growth factor receptor-bound 2 (Grb2). Binding of the peptide inhibited the protein interactions of Grb2 and the Ras guanine nucleotide exchange factor Sos1 which is a drug target for cancer therapy (Yu *et al.*, 2017).

5.1.2 The N-terminal SH3 Domain of HsPRMT2

PRMT2 can interact via its SH3 domain with proline-rich PRMT2-binding proteins, including two proline-rich sequences in the N-terminal domain of PRMT8 (Sayegh *et al.*, 2007) and the heterogeneous nuclear ribonucleoprotein (hnRNP) E1B-AP5 (Kzhyshkowska *et al.*, 2001). Furthermore, it was shown that PRMT2 can interact and most likely forms oligomers with PRMT1 (Pak *et al.*, 2011). Interaction of PRMT2 with PRMT1 increased the methyltransferase activity of PRMT1. Interestingly, inhibition of methylation by adenosine dialdehyde prevented the interaction of PRMT1 with full-length PRMT2 (Pak *et al.*, 2011). However, PRMT2 mutants lacking the SH3 domain were still able to bind PRMT1, which might indicate that the SH3 domain facilitates PRMT1/PRMT2 interaction in a methylation-dependent manner (Pak *et al.*, 2011). Currently, it is not clear whether the SH3 domain of PRMT2 is necessary for its *in vitro* or *in vivo* enzyme activity, as the literature results are inconsistent. In some enzyme activity studies with PRMT2 full-length and a truncated construct that was missing the SH3 domain, activity was completely abolished upon SH3 deletion (Lakowski and Frankel, 2009), whereas in another study reduced enzyme activity was still measurable (Cura *et al.*, 2017). In the present study using the MTase GloTM assay, PRMT2 activity was also detected without the SH3 domain (Section 4.1) supporting the findings by Cura *et al.* (2017). PRMT2 was also shown to interact with the actin nucleator Cobl via its SH3 domain and methylation of Arg1226 and Arg1234 of Cobl by PRMT2 caused the formation of new actin filaments (Hou *et al.*, 2018). Moreover, the SH3 domain of

PRMT2 interacts with a number of splicing factors including SAM68 and is involved in alternative splicing (Vhuiyan *et al.*, 2017).

So far, no crystal structure has been determined for the SH3 domain of PRMT2. However, an NMR structure has been previously solved (PDB: 1X2P, Figure 5-5). It adopts the typical SH3 domain fold with 5 β -sheets that forms two antiparallel β -sheets: β 1 (Glu9-Ala12); β 2 (Lys30-Arg35); β 3 (Trp41-Glu46) and β 4 (Gly51-Pro55). The β -sheets are connected by the RT (Ile13-Glu29), n-Src (Gln37-Asp40) and distal (Arg47-Cys50) loops. It needs to be noted that β 5 shows a very high degree of flexibility in the NMR structure.

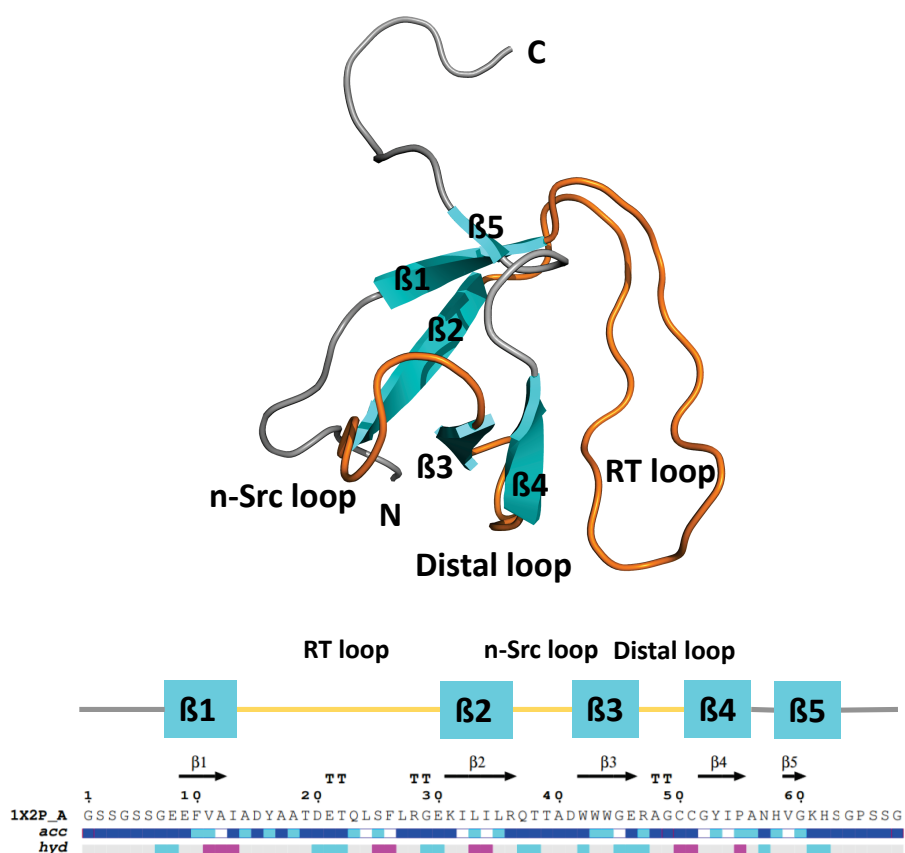


Figure 5-5 NMR Structure of HsPRMT2 SH3 Domain.

ESPrpt output obtained from HsPRMT2 sequences (Gouet *et al.*, 1999). Secondary structure elements are presented on top: β strands are shown as arrows, turns with TT letters. Accessibility of the SH3 domain is rendered by a bar below: blue is accessible, cyan is intermediate, white is buried. Hydrophobicity is rendered by a second bar: pink is hydrophobic, cyan is hydrophilic. The UniProtKB sequence region (orange) that is included in the PDB entry is shown underneath.

A structural overview of the PRMT2 SH3 domain and its binding surface is shown in Figure 5-6. The two xP binding pockets are mainly hydrophobic and are lined by aromatic residues Tyr16, His58, Trp43 and Pro55. The RT-loop is involved in formation of the specificity pocket, which is mostly negatively charged.

From the structural alignment, it became evident that the canonical negatively charged acidic residue located in the specificity pocket of the SH3 domain is replaced by a neutral glutamine residue (Gln/Q23) in PRMT2. Changes at this position have been observed in other SH3 domain containing proteins including Abelson kinase (Abl) (Pisabarro *et al.*, 1998), Insulin receptor tyrosine kinase substrate (IRTKS) (Aitio *et al.*, 2010) and β -PAK-interactive exchange factor (β PIX) (Hoelz *et al.*, 2006). They all bind Class I ligands but their ability to bind RxxPxxP ligands in the specificity pocket was weakened. However, all of them had additional specificity areas to enhance ligand affinity and thus ligand binding. Without further structural studies of the PRMT2 SH3 structure in complex with binding partners it is not clear whether other specificity areas outside of the “traditional pocket” do exist for PRMT2 as well.

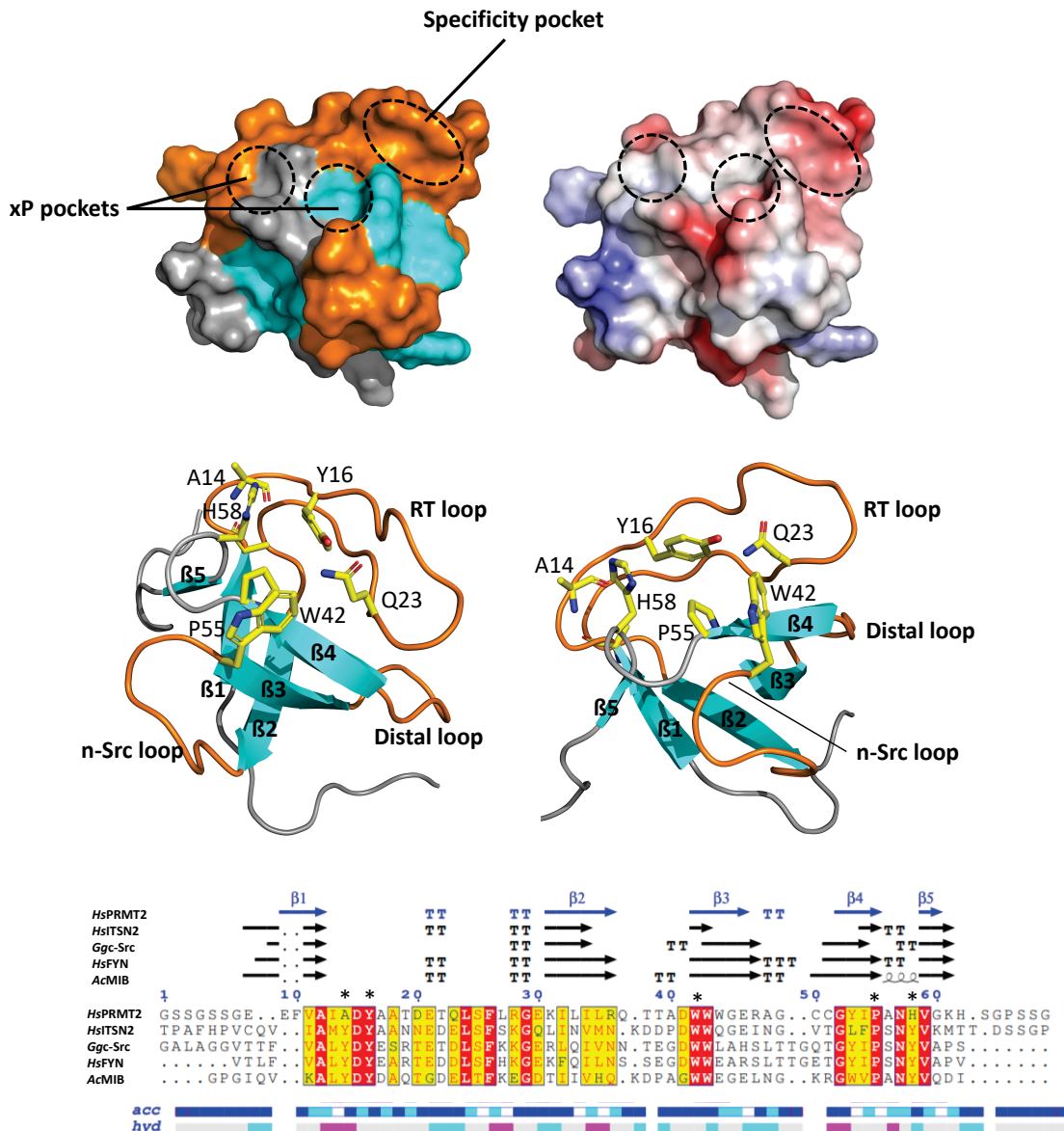


Figure 5-6 Structural Analysis of the HsPRMT2 SH3 Domain and its Binding Surface.

The secondary structural elements are labelled and coloured: β -sheets cyan, loop regions orange. The key residues that form the conserved xP pockets are shown as stick models (A14+Y16 + H58 and W43 + P55). Q23 of the specificity pocket is also shown. A structure-based sequence alignment of HsPRMT2 and some selected SH3 domains is shown underneath. β strands are shown as arrows, turns with TT letters. Accessibility of the SH3 domain is rendered by a bar below: blue is accessible, cyan is intermediate, white is buried. Hydropathy is rendered by a second bar: pink is hydrophobic, cyan is hydrophilic. The residues that are involved in formation of the two xP pockets are highlighted with a star. HsPRMT2 (PDB: 1X2P), HsITSN2: intersectin 2 (PDB: 1UDL), Ggc-Src: Gallus Gallus c-Src (PDB: 1PRL), HsFYN: Fyn tyrosine kinase (PDB: 4ZNX), AcMIB: Acanthamoeba myosin I (PDB: 2DRM).

5.2 Aims

At the beginning of this project the structure of PRMT2 had not been determined. Thus, the aim was to determine the first 3-dimensional structure of PRMT2. Knowledge of the structure can be used to characterise the protein and to study its substrate recognition behaviour and can assist the identification of new substrates and interaction partners. Until now, only one *in vivo* PRMT2 substrate, histone H3, has been identified (Dong *et al.*, 2018). Moreover, a number of interaction partners including PRMT8 have been reported but the interaction mechanism is not known. As already discussed, whether the SH3 domain is required for the enzyme activity of PRMT2 is still questionable. Molecular details of cofactor and potential substrate binding pockets could be used to answer this question.

A high-resolution crystal structure would also benefit studies to develop biochemical probes to specifically target PRMT2. One of the aims of the thesis was to develop selective chemical probes for PRMT2 that can act as tools for target validation and potential start points for drug discovery using fragment-based screening by X-ray crystallography. This approach requires target protein in sufficient amounts and purity to produce crystals that diffract to high resolution. These crystals need to be stable and robust to be used in ligand soaking. They need to diffract to at least ~ 2.5 Å to allow the correct placement of the ligand into the electron density and analysis of the binding mode. After the successful production of human and mouse PRMT2 protein (Chapter 3), the next step was to identify and optimise crystallisation conditions for PRMT2 and produce a good crystal system for crystallographic fragment screening.

Additionally, it would be very helpful to determine a crystal structure of PRMT2 where the SH3 domain is visible. Structural studies of the SH3 domain in complex with different peptides could help to identify and characterise interaction partners of PRMT2. If it would be possible to identify specific binding pockets, they could potentially be used for drug targeting of PRMT2 isoforms where the disease might be caused by protein interactions via their SH3 domain as might be the case for some of the truncated PRMT2 isoforms that are suspected to have no enzyme activity. However, more research is needed to clarify how these isoforms contribute to cancer development and to confirm that they possess no enzyme activity.

5.3 Crystallisation Trials of HsPRMT2 and MmPRMT2

Heterologous expression of human and mouse PRMT2 in insect cells yielded sufficient pure protein to carry out crystallisation trials (Chapter 3.5.1). However, in the case of human PRMT2 these trials were limited: of all tested HsPRMT2 constructs only HsPRMT2-1 and HsPRMT2-6 yielded sufficient amounts of pure protein. For crystallisation trials with HsPRMT2-6, the purified protein was concentrated to 9.6 mg/mL and incubated with 0.2 mM SAH for 3 h at 4 °C. Two Index crystal screens, one at 20 °C, the other at 4 °C were prepared due to the limited amount of purified protein. HsPRMT2-1 was concentrated to 3.8 mg/mL and incubated with 0.2 mM SAH on ice for 1 h. Protein amounts were sufficient to prepare one Index crystal screen at 20 °C.

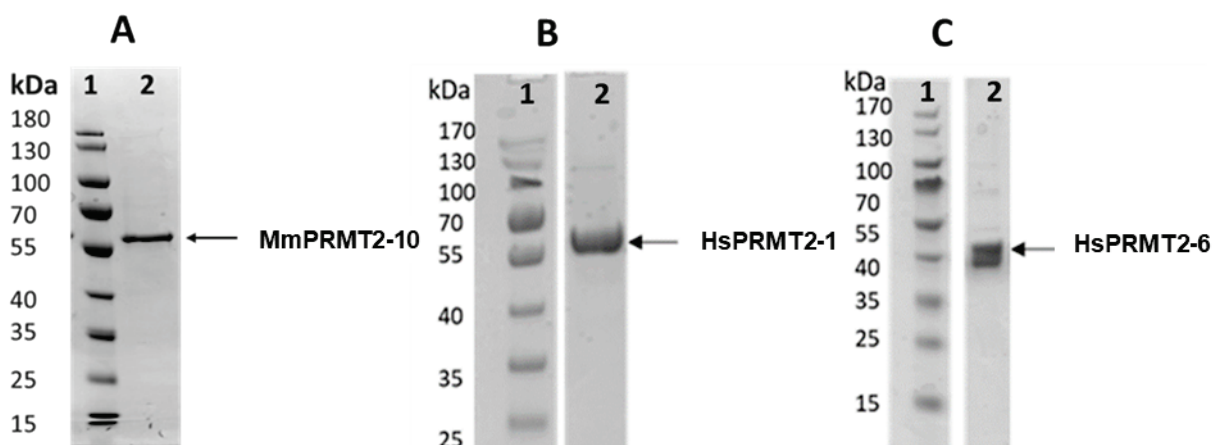


Figure 5-7 HsPRMT2 and MmPRMT2 Crystallisation Trials. Protein used for Crystallisation Experiments.

12 % SDS-PAGE stained with InstantBlue dye. Lane 1, PageRuler protein ladder A) Lane 2, MmPRMT2-10 (MW= 50.69 kDa); B) Lane 2, HsPRMT2-1 (MW= 49.04 kDa C) Lane 2, HsPRMT2-6 (39.8 kDa).

The plates were analysed for crystal growth every few days for six weeks. However, no crystals were obtained. Unfortunately, the following purifications of HsPRMT2 did not yield enough pure protein to perform further crystallisation screens.

Crystallisation trials with MmPRMT2 were more successful. MmPRMT2 expression was also very limited, and from all the tested expression constructs (Chapter 3.3) only MmPRMT2-10 was taken forward into crystallisation trials. MmPRMT2-10 is a full-length construct (1-445) that harbours a point mutation at the C-terminal end (R445W) and has been successfully crystallised in the literature (Cura *et al.*, 2017). The purified protein (Section 3.8) was concentrated to 3.5 mg/mL and for co-crystallisation with

Sinefungin (SNF) it was subsequently incubated with 1 mM SNF for 1 h at 4 °C. Both samples (+/- SNF) were centrifuged at 10 000 g for 10 min at 4 °C before setting up the trays to remove insoluble particles. Different crystallisation screens including INDEX, Pact, PEG and JCSG+ were prepared as described in Section 2.22 and the plates were stored at RT. MmPRMT2-10 with/without SNF did not crystallize in any of the PEG Screens. However, the first crystal hits appeared after a few days in JCSG+ and Pact Screens for MmPRMT2-10 with and without SNF (Table 5-1 and Figure 5-8). In total 12 crystals were harvested and tested at Diamond Light Source, UK.

Table 5-1 Growth and Harvest Conditions for MmPRMT2-10 Crystals used for Structure Determination.

Protein complex	Ligand	Growth conditions	Harvest conditions
MmPRMT2-10	None	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES pH 7.0, 20 % (w/v) PEG 6000	20 % (v/v) PEG400
MmPRMT2-10	SNF	0.2 M Calcium chloride dihydrate, 0.1 M MES pH 6.0, 20 % (w/v) PEG 6000	20 % (v/v) PEG400
MmPRMT2-10	SNF	0.02 M Magnesium chloride, 0.1 M HEPES pH 7.5, 22 % (w/v) PAA 5100	25 % (v/v) EG
MmPRMT2-10	SNF	0.2 M Calcium chloride hexahydrate, 0.1 M HEPES pH 7.0, 20 % (w/v) PEG 6000	20 % (v/v) PEG400

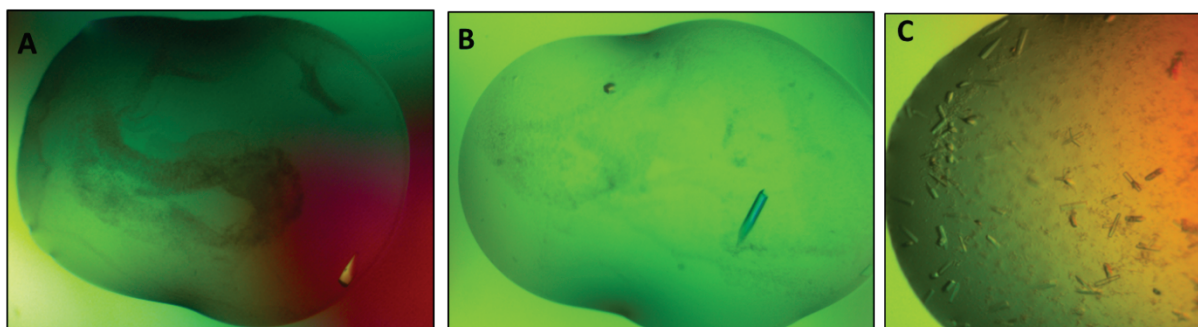


Figure 5-8. MmPRMT2-10 Crystallisation Trials with/without Sinefungin.

A) MmPRMT2-10 no SNF, PACT Screen: C10: 0.2 M $MgCl_2 \times 6H_2O$, 0.1 M HEPES pH 7.0, 20 % PEG 6000, **B)** MmPRMT2-10 with SNF, PACT Screen: B11: 0.2 M $CaCl_2$, 0.1 M MES pH 6.0, 20 % PEG 6000, **C)** MmPRMT2-10 no SNF, JCSG+ Screen: G2: 0.02 M $MgCl_2$, 0.1 M HEPES pH 7.5, 22 % PAA 5100.

Three of the crystals diffracted to a resolution of 2.3-2.7 Å and the crystal structures have been determined. The data was collected at the Diamond Light Source (Didcot, UK) on the I03 beamline by Dr Arnaud Basle (Newcastle University, UK).

5.4 Overall Crystal Structure of MmPRMT2 in Complex with Sinefungin

Diffraction images were processed using Xia2 with 3dii using XDS and XSCALE (Kabsch, 2010). The data was then processed using CCP4i2 (Potterton *et al.*, 2018). The crystal structure of full-length MmPRMT2-10 in the presence of the known PRMT inhibitor, the SAM analogue Sinefungin (SNF), was solved by molecular replacement using Phaser (McCoy *et al.*, 2007) and the published high resolution MmPRMT2 structure (PDB code: 5FUL). Model building was conducted in COOT followed by refinement in REFMAC5 (Murshudov *et al.*, 2011). MmPRMT2 in complex with SNF crystallised in the space group C222₁ (93.37 % probability by Pointless) and the structure was refined to a resolution of 2.3 Å with an R-factor of 18.8 % and an R-free value of 24.6 %. The statistical details for data processing and refinement are provided in Table 5-2.

The full-length protein comprising residues 1-445 was used for the crystallisation experiments, however, the first 106 N-terminal residues including the SH3 domain were not visible in the electron density map. The structure starts at Asp107 and ends at Trp445. MmPRMT2 shows the typical Type 1 PRMT structural features as described in detail in Section 1.3. Briefly, the protein is arranged as a head-to-tail dimer. The SAM-binding domain adopts the typical Rossmann-fold (green, residues 107-254, Figure 5-9). The Rossmann fold is a structural motif, in which two units consisting of alternating β -sheets and α -helices are sandwiched together (Hanukoglu, 2015). In the MmPRMT2 structure it consists of four α -helices and five β -strands. The N-terminal region consists of three α -helices named α -X, α -Y and α -Z, of which α -X and α -Y (residues 107-120 and 121-140, red) fold over the cofactor mimic. The Rossmann fold is connected to a β -barrel domain via a conserved proline residue (Pro254). The β -barrel domain consists of 11 β -strands and comprises residues 255-266 and 300-445 (cyan, Figure 5-9). The inserted dimerisation arm (residues 267-299) that interacts with the Rossmann domain of the other monomer is shown in blue. The dimerisation arm consists of two α -helix segments that are separated by a loop.

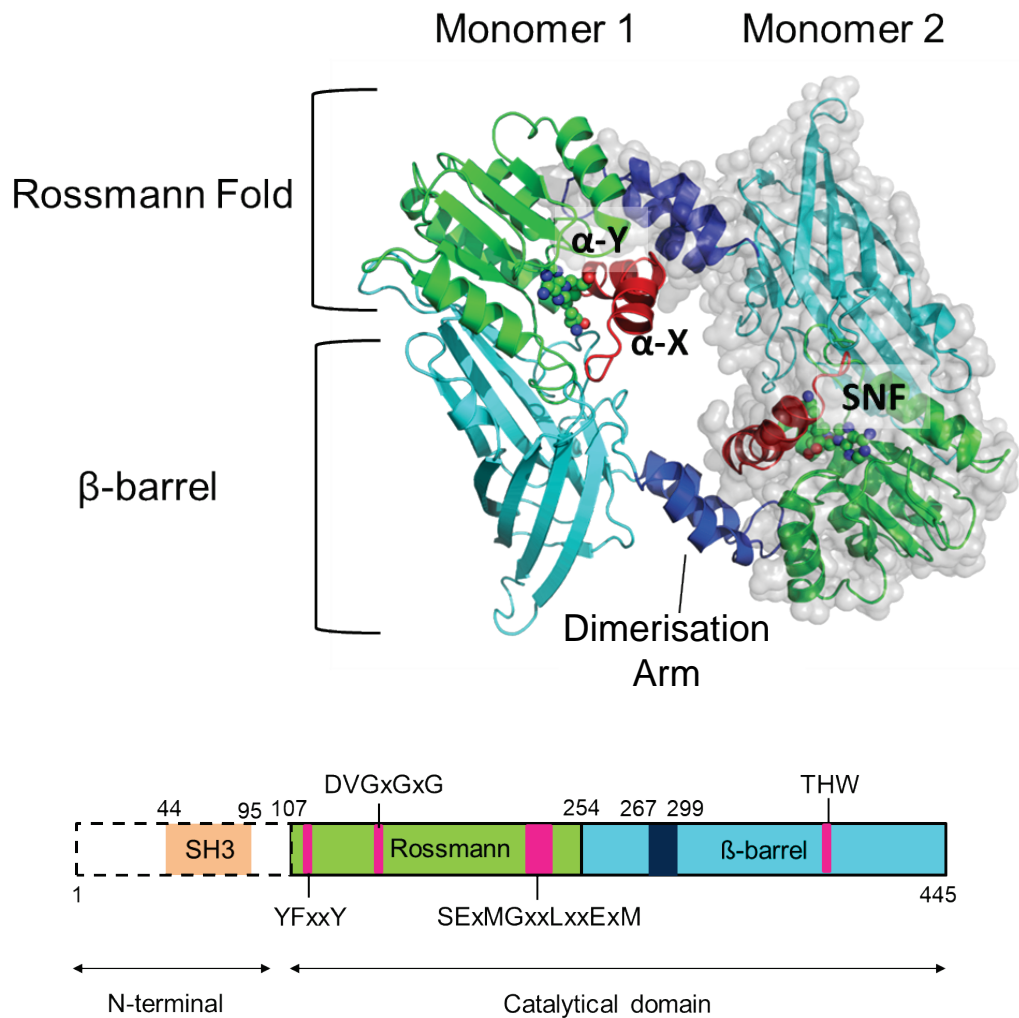


Figure 5-9 Structural Overview of MmPRMT2 in Complex with Sinefungin.

The PRMT2 head-to-tail dimer is shown as cartoon. The SAM binding domain that adapts a Rossman fold is shown in green, its N-terminal α -helix in red. The β -barrel domain (cyan) is interrupted by the dimerisation arm (blue) that interacts with the Rossman fold of the other monomer. One monomer shows the transparent surface. The bound inhibitor Sinefungin is represented as spheres. The figure was generated using Pymol (DeLano, 2002). A schematic illustration of the modular architecture of PRMT2 and the invariant motifs is shown at the bottom. Dashed lines indicate regions that were not visible in the electron density map.

Table 5-2 Crystallographic Data and Refinement Statistics of MmPRMT2-10 in Complex with Sinefungin.

Ligand	Sinefungin
Wavelength (Å)	0.979
Data processing	
Resolution Range (Å)	33.03- 2.30 (2.36-2.30)
Space Group	C222 ₁
Unit cell (Å, °)	66.06, 114.2, 131.9 90, 90, 90
Number of total reflections	161851 (16056)
Number of unique reflections	22602 (2215)
Multiplicity	7.2 (7.2)
Completeness (%)	100
I/σ (I)	8.7 (1.5)
Resolution limit	2.3
R _{meas}	0.145 (1.507)
CC _{1/2}	0.989 (0.731)
Refinement	
R factor	0.188
R free	0.246
Number of non-hydrogen atoms:	
Protein	2722
Ligands	27
Water	45
Validation	
RMS Bonds (Å)	0.0009
RMS Angle (°)	1.64
Ramachandran favoured (%)	97.63
Ramachandran Outliers (%)	0
Average B factor (Å ²)	48.71
Rotamer Outliers (%)	2.3

X-ray diffraction data were collected from a single crystal on beamline I03 at the Diamond Light Source (Didcot, UK) at -180 °C. Data processing was performed using Xia2 (3dRun). Statistics for the highest resolution shell are shown in parentheses.

5.5 The Sinefungin Binding Site of MmPRMT2

The cofactor mimic Sinefungin is a known pan-inhibitor of SAM-dependent methyltransferases that was isolated from the bacterium *Streptomyces griseolus* (Borchardt *et al.*, 1979). In Sinefungin, the SAM sulphur is replaced by a CH-group and an amine group is present in place of the SAM methyl-group (Figure 5-10). Sinefungin competes with SAM by binding to the cofactor binding pocket of the catalytic domain of PRMTs, which is well conserved (Zhang and Zheng, 2016).

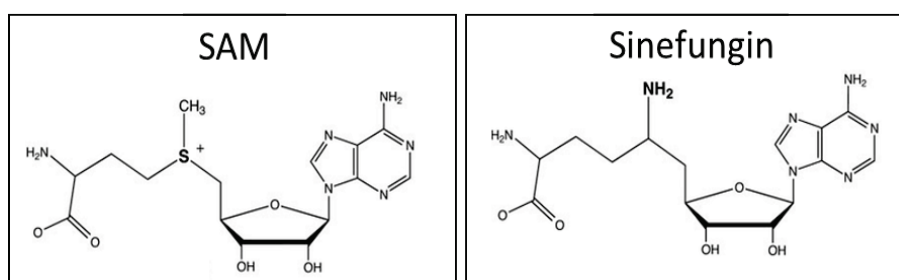


Figure 5-10 Chemical Structures of the PRMT cofactor SAM and the fungal-derived pan-PRMT Inhibitor Sinefungin.

Figure adapted from Zhang and Zheng (2015).

Figure 5-11 shows an overview of the SNF binding site of MmPRMT2. As already seen in other PRMT-SNF structures, the ligand binds in the cofactor binding pocket that is formed by three β sheets (β 1, 2 and 4) and the three α -helices (α -X-Z) that are located at the N-terminus. The ligand is buried in a deep pocket that is formed by the N-terminal α -helix (residues 107-130). All of the four invariant PRMT motifs are located in close proximity to the ligand. Protein-ligand interactions in the active site were analysed within 4 Å of SNF using CCP4mg and PoseView (Stierand and Rarey, 2010b) and are shown in Figure 5-12 A and B. The electron density for the ligand was well defined and SNF could be easily modelled into it. In previously obtained crystal structures of Type 1 PRMTs in complex with SNF, the binding mode of SNF to the cofactor pocket is very similar to SAH, and SNF forms interactions with active site residues of which most are highly conserved across all PRMT enzymes. This is also the case for PRMT2. Figure 5-12 A and B show the SNF binding site and the ligand interactions of two of the previously obtained crystal structures of zebrafish PRMT2 (PDB: 5G02) and human CARM1 (PDB: 5DXJ) for comparison. A glutamine residue Glu180 (CARM1: Glu215) which is located at the end of the β 2 strand forms hydrogen bonds with the hydrogen oxygens of the ribose moiety. In MmPRMT2 and CARM1, a serine

residue (Ser182 or Ser217, respectively) also interacts with one of the two oxygens. Another glutamine residue located in a loop after the β_3 sheet, Glu209 (CARM1: Glu244), forms a hydrogen bond with the amino group of the adenine.

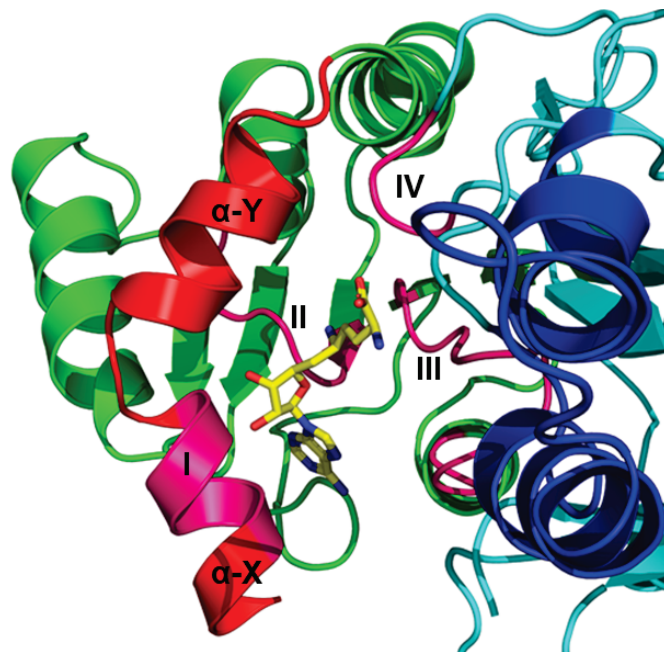


Figure 5-11 Overview of the SNF-binding Domain of MmPRMT2.

The SAM/SNF binding domain is shown in green and SNF as a stick model in yellow. The N-terminal α -helix consisting of the two element α -X and α -Y is shown in red. The β -barrel domain is coloured in cyan, whereas the dimerisation arm is shown in blue. The four PRMT motifs I, II, III and IV are highlighted in pink.

Additional interactions made by MmPRMT2 include Val208 that interacts with the nitrogen of the adenine moiety. The glutamine residue Glu223 (CARM1: Glu258) interacts in all structures with one of the nitrogen atoms of the Sinefungin. However, in MmPRMT2, the nitrogen atom has a different orientation to that seen in the other structures. It also interacts with Gly157 which is part of motif II (Asp155-Val156-Gly157-x-Gly159-x-Gly161).

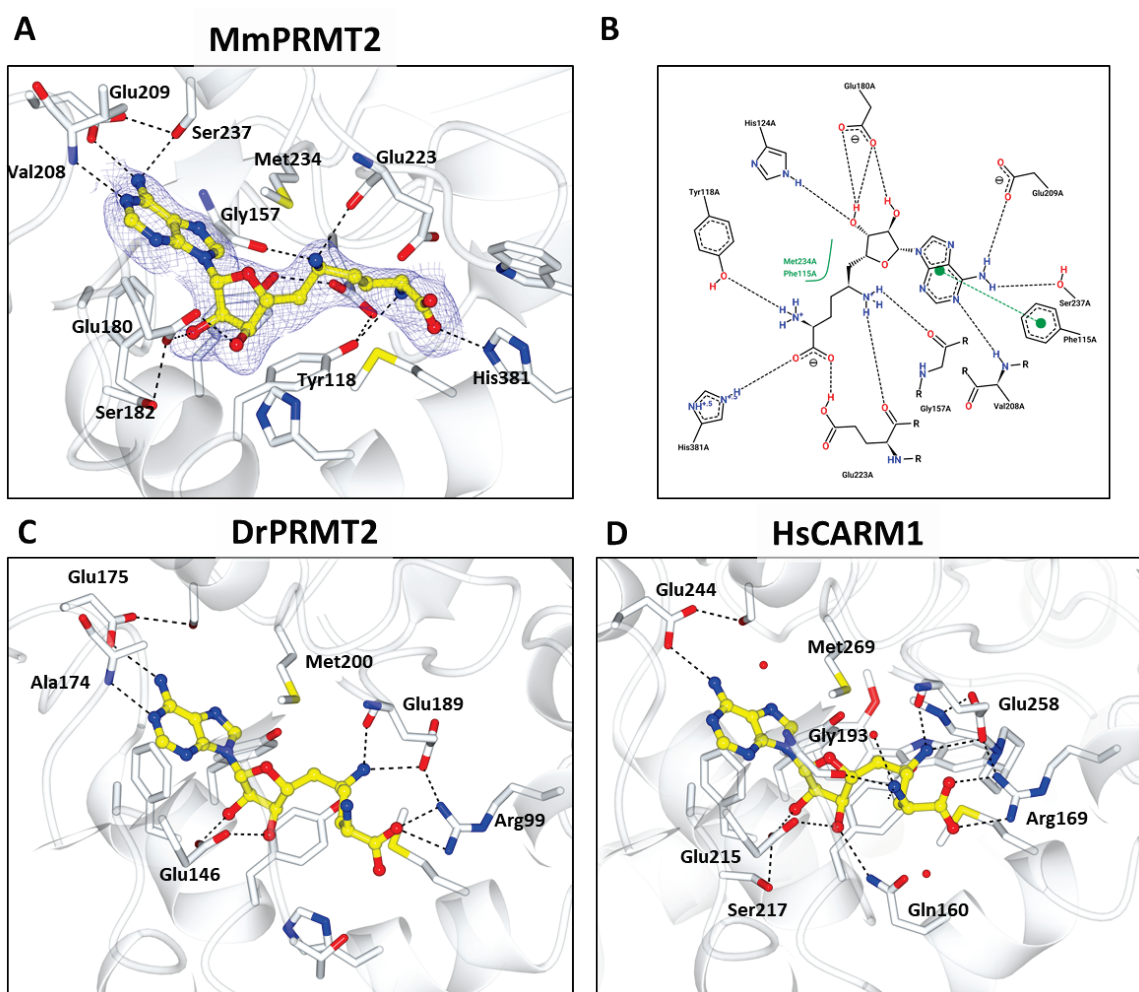


Figure 5-12 The Sinefungin Binding Pocket of Type 1 PRMTs.

A) Active site residues of MmPRMT2 within 4 Å of SNF and hydrogen binding interactions as calculated in CCP4mg (McNicholas et al., 2011). The electron density of SNF (yellow) is displayed as a blue mesh and contoured around the ligand at 1 σ . The final map was calculated using σ_A -weighted coefficients ($2mF_o - DF_c$) as calculated by the programme SigmaA (Read, 1986). **B)** Protein-ligand interaction diagram for PRMT2 derived from PoseView (Stierand and Rarey, 2010b). Hydrophobic interactions are represented as green smooth contour lines. Hydrogen bonds are shown as black dashed lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of 1.9 Å \pm 0.5 Å and the condition that the acceptor–hydrogen–donor angle must not fall < 120°. **C)-D)** Active site residues of DrPRMT2 (PDB: 5G02) and HsCARM1 in complex with SNF (PDB: 5DXJ). Dr= *Danio rerio/zebrafish*.

The structural comparison of our MmPRMT2-SNF structure and the previously published MmPRMT2-SAH structure (PDB: 5FUL) is shown in Figure 5-13. As for our MmPRMT2 structure, the N-terminal domain (AA 1-106) is missing in the electron density map, even if the full-length protein was crystallised, most likely due to high flexibility of the domain. The two structures align very well with an overall RMSD of 0.30/0.20 Å, calculated in Pymol (without/with outlier rejection). However, the different binding mode of SNF, which in other known PRMT structures occupies the SAH binding site, can

easily be seen. Moreover, two calcium ions are visible in the published MmPRMT2-SAH structure, one occupies the space where the guanidine group of the arginine substrate usually sits. The addition of calcium to the crystallisation drop was needed to obtain well diffracting crystals in the case of MmPRMT2-SAH (Cura et al., 2017). In our MmPRMT2-SNF structure no calcium ion was observed but the space is occupied by SNF.

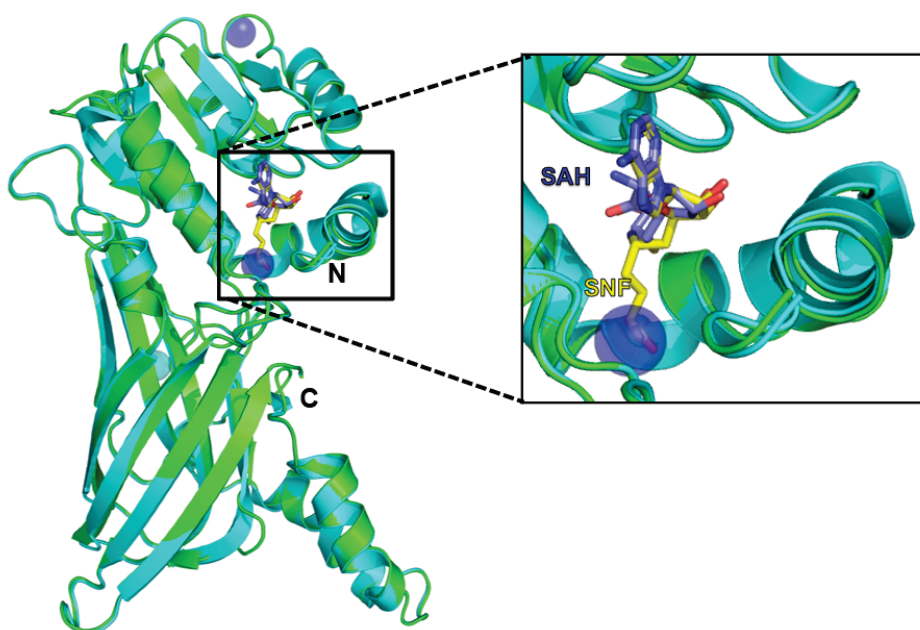


Figure 5-13 Structural Comparison of MmPRMT2 in Complex with SNF and SAH.

Structural alignment of MmPRMT2 (green cartoon) in complex with SNF (yellow sticks) and MmPRMT2 (cyan cartoon) and SAH (purple sticks) with a close-up of the SAH/SNF binding pocket. The structures align well but the two ligands adopt different binding positions. The two calcium ions in the MmPRMT2-SAH structure are shown as purple spheres. PDB code of MmPRMT2-SAH: 5FUL.

It is evident that the SNF molecule in the mouse PRMT2 structure determined in this study does not bind in the same position as SNF or SAH bound to zebrafish PRMT2, human CARM1 (PRMT4) or mouse PRMT2. Instead, the homocysteine moiety extends into the substrate arginine binding site. Usually, the guanidine moiety of the substrate arginine would be located in this position as shown in Figure 5-14.

In the other PRMT structures, the carboxylate atoms of the cysteine moiety interact with an arginine residue (CARM1: Arg169) located in the α -Z helix, that also interacts with one of the already mentioned glutamine residues (Glu258) of the double E-loop. In contrast, in MmPRMT2 the carboxylate group interacts with the histidine residue (His381) of the THW-loop. The other nitrogen atom of the adenine ring moiety

that interacts with another glycine residue of motif II in CARM1 (Gly193) interacts with a tyrosine in MmPRMT2 (Tyr118). This tyrosine is part of one of the four PRMT2 motifs called motif I (Tyr114-Phe115-x-x-Tyr118) which is located on the α -X helix.

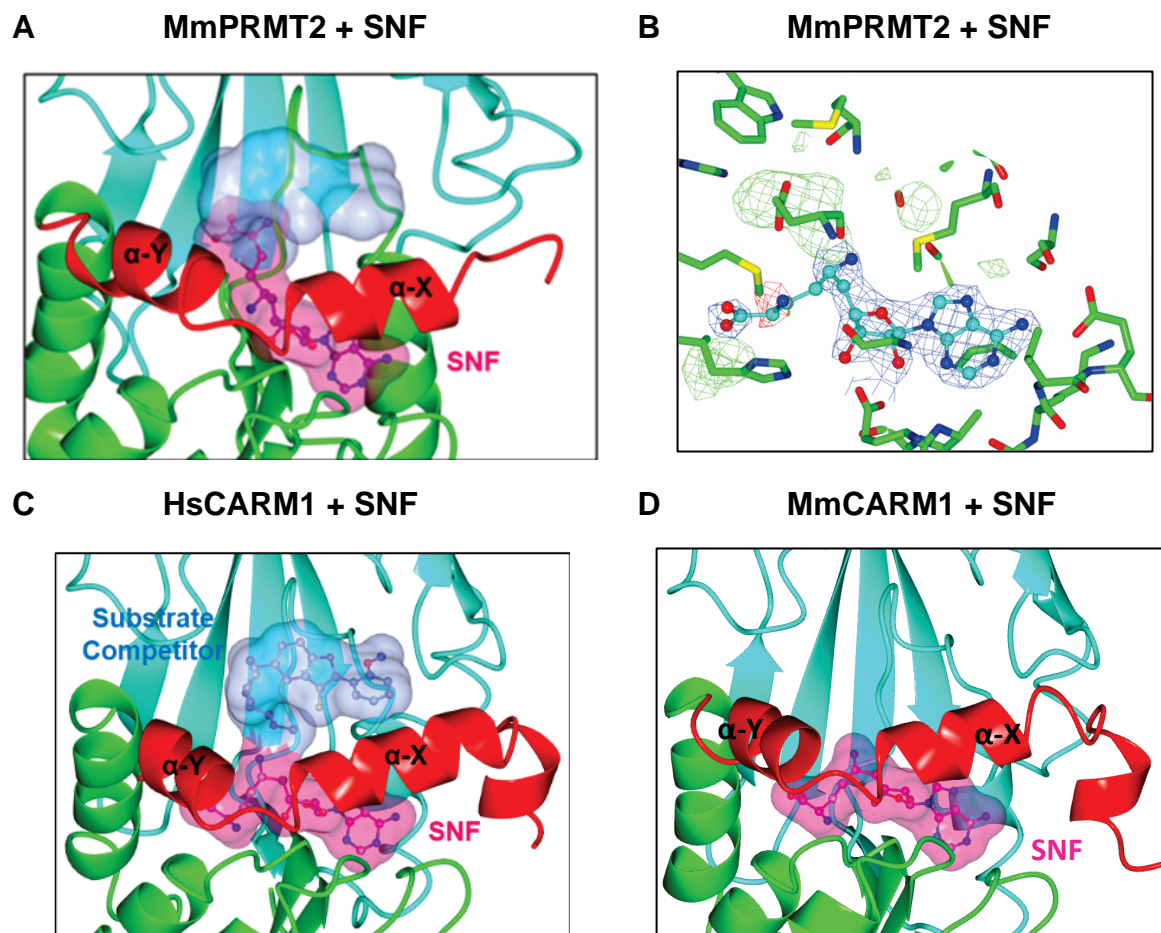


Figure 5-14 Cofactor and Substrate Binding Pockets of Different PRMTs.

A) In the case of MmPRMT2, SNF (pink sticks) reaches into the arginine substrate pocket visualised by a blue shell. **B)** SNF Binding Site in MmPRMT2 with SNF fitted in common PRMT orientation. Electron density maps are shown as chicken wire. The negative different density map is shown in red (-3σ), positive different density map in green ($+3\sigma$). The contour map of the ligand is shown in blue (1.5σ). **C)** HsCARM1 structure (PDB: 2Y1W) in complex with the substrate competitor CMPD-1 (blue sticks) and the cofactor competitor SNF (pink sticks) that occupies the cofactor binding pocket (pink shell). **D)** MmCARM1 structure (PDB: 5DXJ) co-crystallised with SNF that occupies the cofactor binding pocket.

Both tyrosine residues of the α -helix interact with the catalytic Glu223 residue of the double E-loop which is located in the substrate arginine pocket and facilitate the proper formation of the substrate binding pocket. Glu223 is essential for enzyme activity (Lee *et al.*, 2002). Similar binding has been observed for a number of nucleoside inhibitors of PRMT5 which interact with the Glu residues of the E-loop and reach into

the arginine substrate channel. As a result, some of the SAM-competitive PRMT5 inhibitors also compete with the substrate, as it is the case for JNJ64619178, a selective active PRMT5 inhibitor that is in clinic trials for different cancers (Wu *et al.*, 2018a, Lin and Luengo, 2019).

5.6 Comparison of the Mouse MmPRMT2 Apo and SNF-bound Structures

The apo MmPRMT2-10 structure was successfully crystallised in the space group C222₁ (86.6 % probability by Pointless) and was refined to a resolution of 2.2 Å with an R-factor of 19.5 % and an R-free value of 23.5 % (Table 5-3). A structural overview of one PRMT2 apo monomer is shown in Figure 5-15.

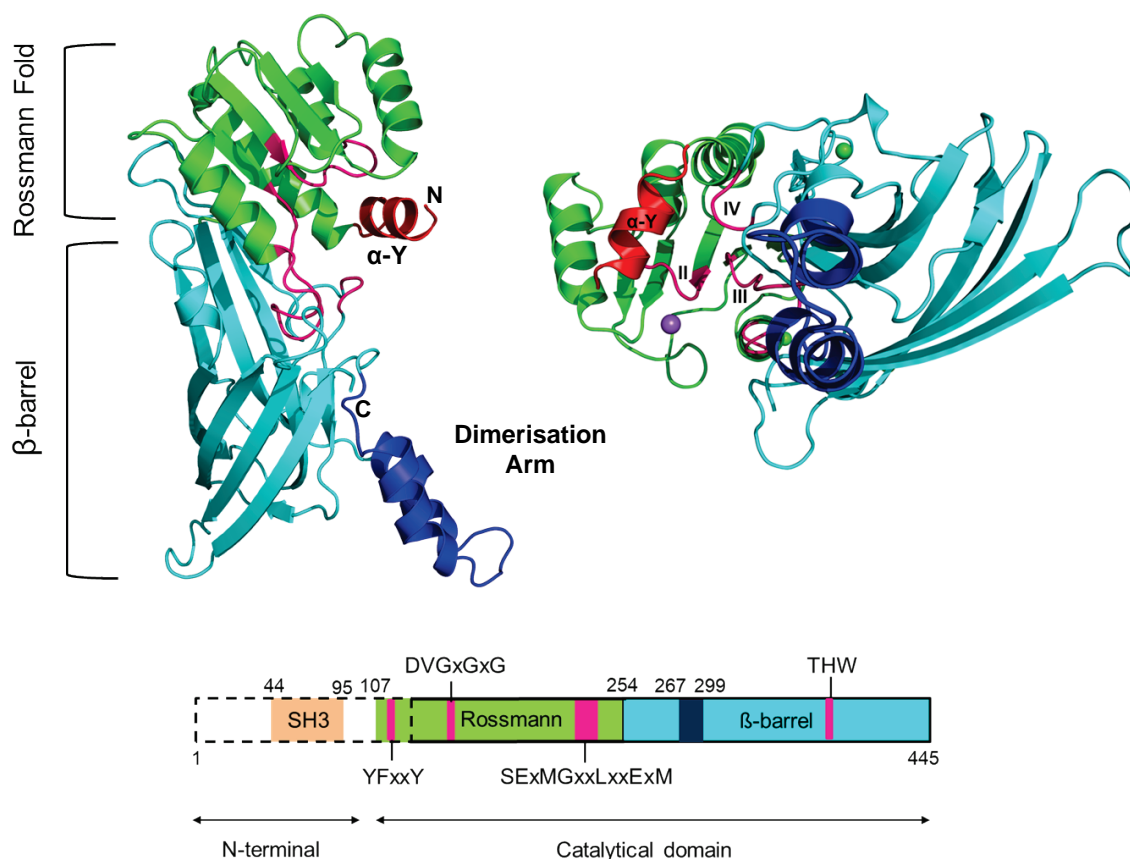


Figure 5-15 Structural Overview of Apo-MmPRMT2.

The PRMT2 monomer is shown as cartoon. The SAM binding domain that adapts a Rossmann fold is shown in green, its N-terminal α -helix in red. Only the α -Y element of the helix was visible in the electron density map. The β -barrel domain (cyan) is interrupted by the dimerisation arm (blue). The four invariant PRMT motifs I-IV are labelled. There is one Na⁺ ion (purple sphere) in the apo MmPRMT2 structure. A schematic illustration of the modular architecture of PRMT2 and the invariant motifs is shown at the bottom. Dashed lines indicate regions that were not visible in the electron density map.

Superposition of the apo and SNF-bound structures shows that the overall fold of both MmPRMT2 structures is very similar (Figure 5-16). Both structures align well with an overall RMSD value of 0.89/0.40 Å as calculated using Pymol (without/with outlier rejection).

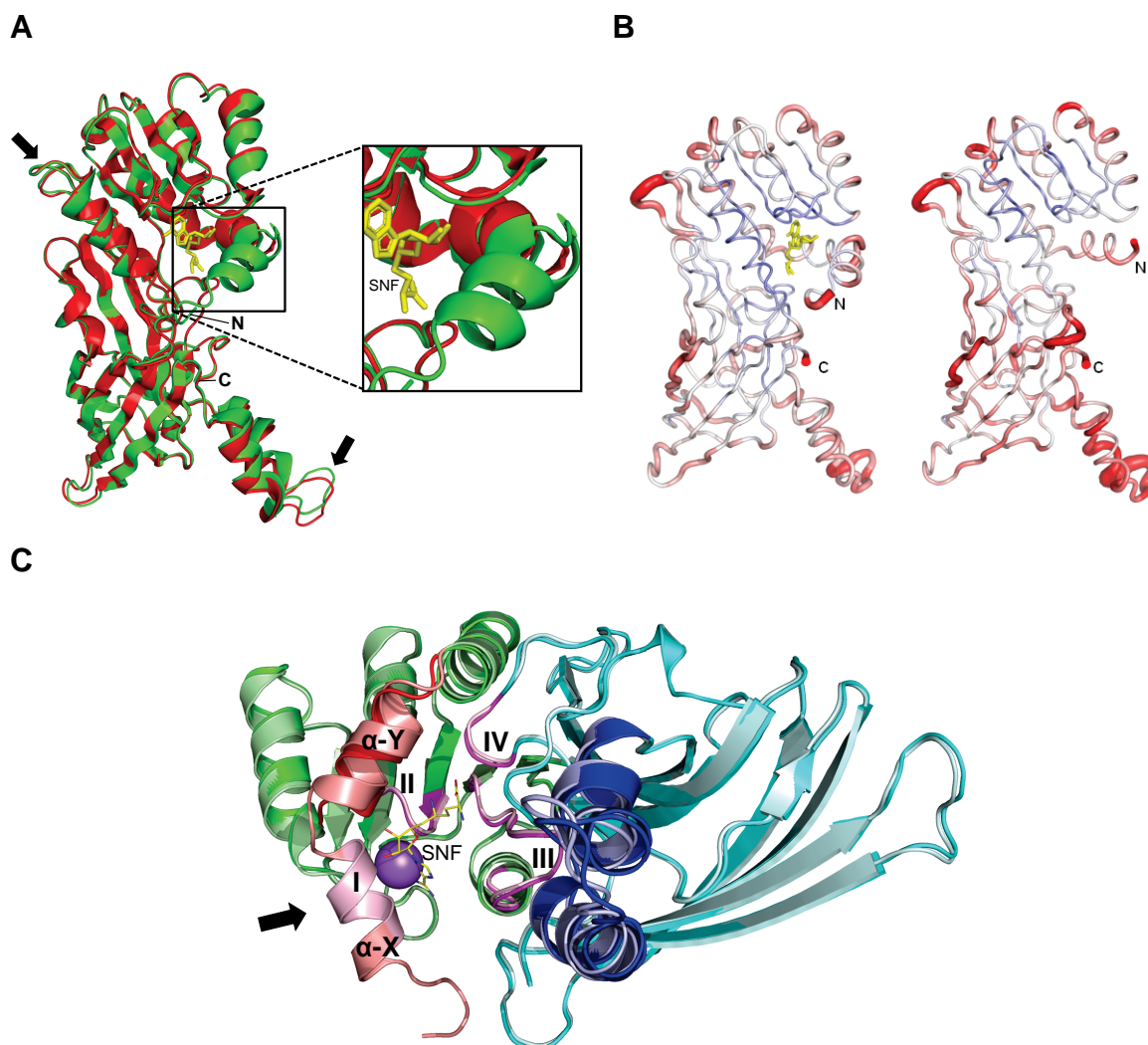


Figure 5-16 Structural Comparison of Apo-MmPRMT2 and MmPRMT2 in Complex with SNF.

A) Structural alignment of MmPRMT2 in complex with SNF (green) and apo PRMT2 (red) and a close-up of the α -helix. **B)** B-factor diagram of Apo-MmPRMT2 and MmPRMT2 in complex with SNF. The B-factor values are illustrated by colour; low (blue) to high (red). Both analysis were performed in Pymol (DeLano, 2002). **C)** Structural alignment of both MmPRMT2 structures. The Rossmann fold is shown in green with the n-terminal α -helix in red, the β -barrel domain is coloured in cyan and the inserted dimerisation arm in blue. The SNF-bound structure is shown in paler colours. The motifs I-IV that surround SNF (yellow) are highlighted in pink. There is one Na⁺ ion (purple sphere) in the apo MmPRMT2 structure.

However, in contrast to the superposition with the SAH bound MmPRMT2 structure (Figure 5-13), where no large structural differences could be seen, three structural changes were observed. Two loop regions, indicated by black arrows, one located in the dimerisation arm (Leu275- Lys289) between α -E and α -G, the other in the β -barrel domain (Phe359- Gln369) between two β -sheets, differ in their location. The RMSD value of the first loop is 0.91 Å, the second is 1.03 Å. B-factor analysis seen in Figure 5-16 B, showed that they are also the most flexible regions of the protein and show a higher degree of flexibility in the apo structure.

The biggest structural difference between the apo- and SNF-bound structures is at the N-terminal α -helix that consists of two elements α -X and α -Y. In the SNF bound PRMT2 structure, both α -helix elements are visible (Figure 5-9), whereas in the apo-structure the α -X element comprising residues 107-120 is missing in the electron density map. The apo-structure starts at Leu121 and ends at Trp445. Figure 5-17 shows the electron density maps of the last visible N-terminal residues of the MmPRMT2 structure in complex with SNF (Figure 5-17 B) and in its apo state (Figure 5-17 A).

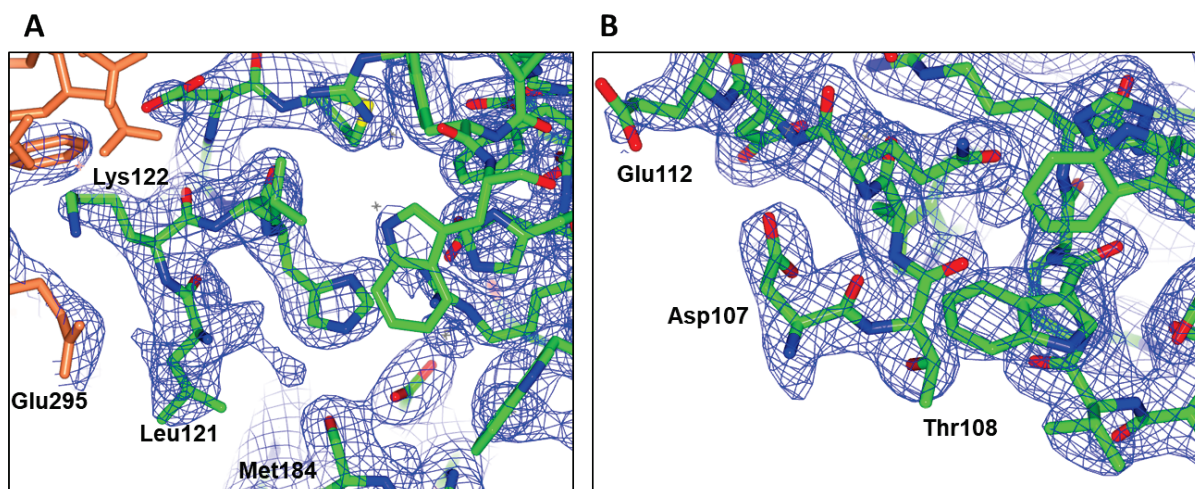


Figure 5-17 Observed Electron Density at the N-terminus of MmPRMT2.

The MmPRMT2 structures are shown as green ribbon, the $2mF_o-DF_c$ maps are shown as a blue mesh and contoured at 1σ . The MmPRMT2 apo structure (A) starts at Leu121, whereas the MmPRMT2 structure in complex with SNF (B) starts at residue Asp107.

This α -helix is a very important structural feature of Type 1 PRMTs, as it not only buries the cofactor upon binding but also participates in the binding and orientation

of the substrate arginine residue in the active site. And as previously mentioned contains motif IV, that interacts with Glu223 of the double E-loop, which is an essential residue for catalytical activity.

Table 5-3 Crystallographic Data and Refinement Statistic of Apo MmPRMT2-10.

Ligand	None
Wavelength (Å)	0.976
Data processing	
Resolution Range (Å)	58.03-2.20 (2.26-2.20)
Space Group	C222 ₁
Unit cell (Å, °)	67.13, 115.45, 131.0 90, 90, 90
Number of total reflections	193449 (16918)
Number of unique reflections	26269 (2251)
Multiplicity	7.4 (7.5)
Completeness (%)	100
I/σ (I)	14.4 (2.2)
Resolution	2.2
Rmeas	0.072 (0.873)
CC _{1/2}	0.996 (0.888)
Refinement	
R factor	0.195
R free	0.235
Number of non-hydrogen atoms:	
Protein	2598
Ligands	0
Water	95
Validation	
RMS Bonds (Å)	0.0145
RMS Angle (°)	1.89
Ramachandran favoured (%)	97.52
Ramachandran Outliers (%)	0
Average B factor (Å ²)	56.61
Rotamer Outliers (%)	3.08

X-ray diffraction data were collected from a single crystal on beamline I03 at the DIAMOND Light Source (Didcot, UK) at -180 °C. Data processing was performed using Xia2 (3dii). Statistics for the highest resolution shell are shown in parentheses.

As seen in Figure 5-18, in the apo-structure, the α -X element is not structured, the cofactor binding site is exposed and accessible for potential ligands including SNF and SAM. Upon SNF binding the α -X helix, shown in red, folds over the ligand, which is then deeply buried in the cofactor binding pocket.

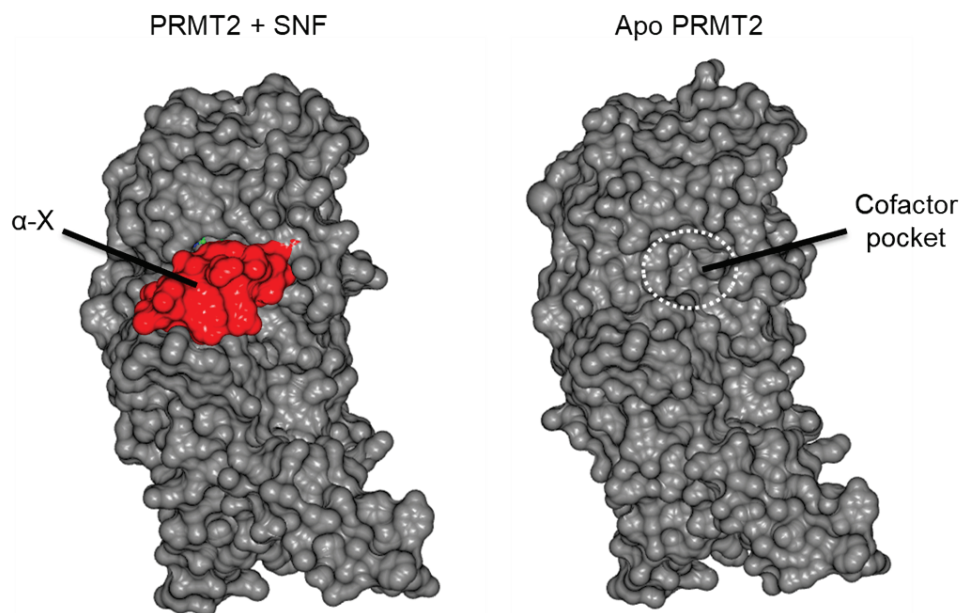


Figure 5-18 Surface Representation of the PRMT2 Structures.

The absence of the α -X helix in the apo-form has also been observed in other PRMT structures including the published apo-CARM1 structure (Boriack-Sjodin *et al.*, 2015). Figure 5-19 shows a comparison of the structural dynamics of the α -helix of the here solved PRMT2 structure in the SNF- and apo-bound state and the published SNF-CARM1 (PDB: 5DXJ) and apo-CARM1 structure (PDB: 3B3J). In the PRMT2 structure the position of the N-terminal α -helix in the SNF bound state (Figure 5-19 C) is very similar to the one observed in CARM1, in which the helix folds over the ligand (Figure 5-19 D). In the PRMT2 apo structure the α -X element of the α -helix is not visible and most likely disordered. The cofactor binding pocket is accessible for SAM. Upon binding structural changes are induced and the α -X element of the α -helix (red) folds over the cofactor, which is buried deep inside the generated pocket.

However, this is not the case for the CARM1 structures. The location of the α -Y element of the helix differs in apo CARM1 (PDB code: 3B3J) and in the apo PRMT2

structure (compare Figure 5-20 A and B). Moreover, in the apo-state of PRMT2 the α -X element is missing which indicates a high degree of flexibility. In the inactive apo-form of CARM1 only the α -Y element of the helix is present, the α -X element is partially transformed into a β -sheet (Figure 5-19 A and B).

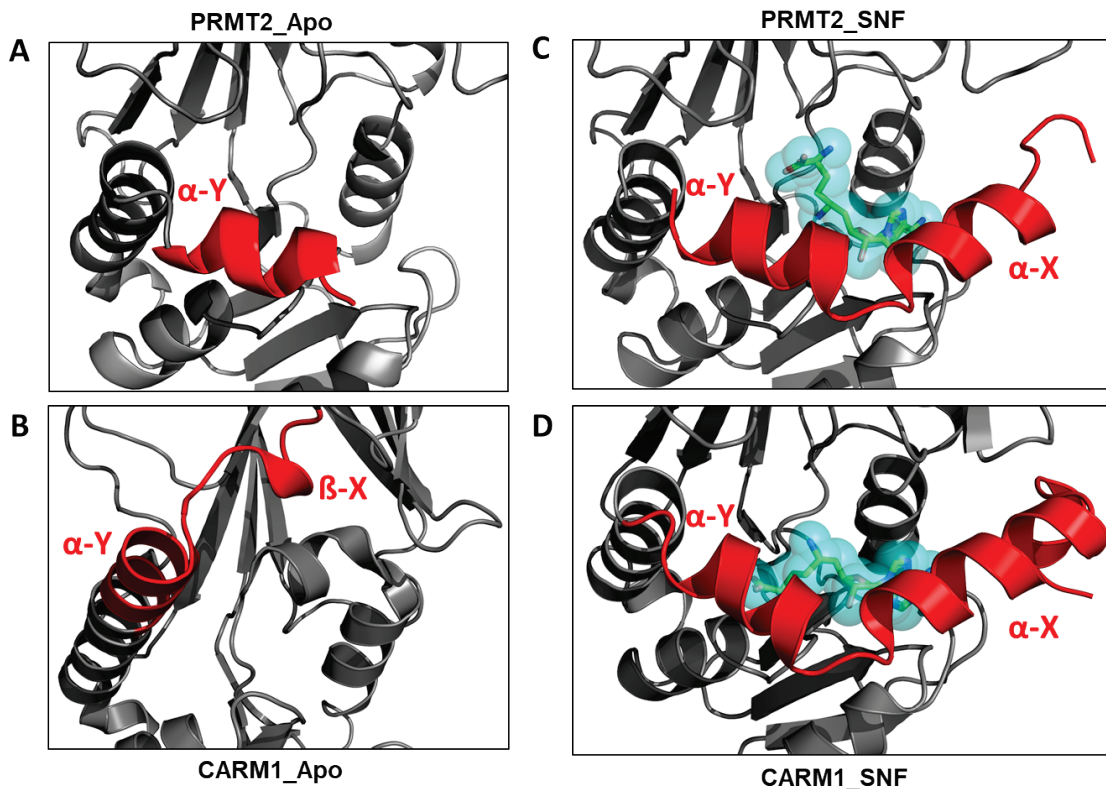


Figure 5-19 Structural Dynamics of the α -helix in PRMTs.

The dynamic α -helix that forms the cofactor binding pocket, consist of two units, α -Y and α -X, which are shown in red in (C) and (D). In the PRMT2 apo structure its α -X segment is not visible most likely due to its flexibility. However, the α -Y is present (A). In contrast, in case of CARM1 the cofactor binding pocket is not present in its apo state structure (PDB code: 3B3J) (B), whereas both α -helical segments are visible in the Sinefungin bound state of PRMT2 (C) and CARM1 and (D; PDB: 5DXJ).

A structural mechanism induced by cofactor binding has been suggested for CARM1 in which this element assembles upon cofactor binding into an α -helix. At the same time, α -Y moves by 180° so that the cofactor gets buried in a deep pocket (Schapira and de Freitas, 2014).

In the case of CARM1 the structural rearrangement of the α -helix would result in a repositioning of the N-terminal domain of the protein, which is a PH domain in CARM1 (Schapira and de Freitas, 2014). In contrast, in the PRMT2 apo-state the α -Y element of the N-terminal helix stays in the same location as in the active cofactor or

SNF-bound state (Figure 5-19 A). This observation would suggest that the regulatory mechanism proposed for CARM1 and other PRMTs in which SAM binding and thus methyltransferase activity correlates with repositioning of the N-terminal domain and its binding partners into proximity of the catalytic core, would not be conserved in PRMT2. Figure 5-20 shows the structural dynamics of the α -helix of both proteins and in the case of CARM1, the reorientation of its N-terminal domain that is caused by structural changes induced upon cofactor binding.

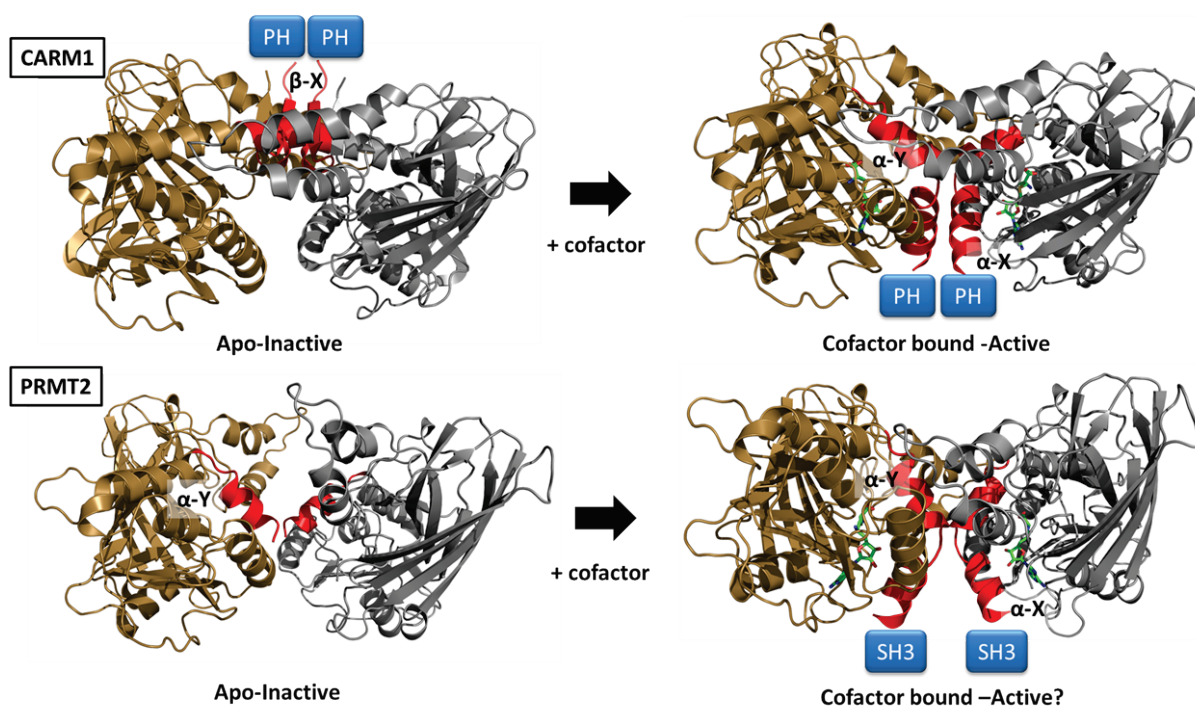


Figure 5-20 Structural Dynamics of the α -helix of CARM1 and PRMT2.

The α -helix of CARM1 is very dynamic and undergoes drastic conformational changes upon cofactor binding. The structural changes from the apo-CARM1 state (PDB: 3B3J) and the cofactor bound/active state, repositions the N-terminal PH domain of the protein. In contrast, no structural changes of the α -Y element of the helix were observed for PRMT2 upon cofactor binding.

In both mouse PRMT2 structures the N-terminal SH3 domain is missing in the electron density map. The Matthew coefficient and solvent content have been calculated for both structures using MATTPROB (Adams-Cioaba and Min, 2009). The solvent content for the full-length mouse PRMT2 (1 monomer per asymmetric unit) was 50.81 % ($V_m = 2.5 \times \text{\AA}^3 \text{ Da}^{-1}$) for the apo structure, and 49.87 % ($V_m = 2.45 \times \text{\AA}^3 \text{ Da}^{-1}$) for the SNF bound structure. Thus, it should not be problematic to fit the N-terminal domain into the unit cell. Its absence is most likely due to the high flexibility of the domain and not degradation and has been reported in previous structures (Cura *et al.*, 2017).

5.7 Summary

The aim was to solve the crystal structure of PRMT2 and to establish a stable crystal system, which could be used to characterise the protein and to study its substrate recognition behaviour and could also be used in the future to identify and optimise new substrates and interaction partners.

Full-length mouse PRMT2 was successfully crystallised in complex with the known pan-inhibitor Sinefungin. Surprisingly, the binding conformation differed from the position usually observed in other PRMT family members. The SAM-mimic reaches into the arginine binding channel, which has been previously observed for different SAM-competitive PRMT5 inhibitors (Lin and Luengo, 2019). Moreover, the first apo-structure of PRMT2 was solved. It was suggested that cofactor binding is needed for formation of the substrate binding pocket. However, this seems not to be the case for PRMT2. Additionally, the orientation of the α -helix, which is located N-terminal to the Rossmann fold and buries the cofactor SAM, differs in the PRMT2 apo structure from other published PRMT apo structures (Cura *et al.*, 2017).

The N-terminal SH3 domain was not visible in the electron density of both solved structures, this was also observed in other known PRMT2 structures, suggesting that it is disordered. However, to eliminate the possibility that the missing parts are due to proteolysis, the crystals would have to be dissolved and analysed by SDS-PAGE to estimate their molecular size. Unfortunately, no more MmPRMT2-10 crystals were available to perform this experiment.

In the future, it would be interesting to try co-crystallisation experiments with known SH3 domain binding partners. Structural data of the SH3 domain with ligands could help to identify potential interactions and might be also used as a start point for drug discovery with the aim to disrupt PRMT2 SH3 domain mediated protein-protein interactions that are involved in different diseases.

Unfortunately, it was not possible to further optimise PRMT2 crystallisation or establish a stable crystal system for screening due to the limited amount of PRMT2 protein.

Chapter 6 Development of Small-Molecule Inhibitors of CARM1

The Coactivator Associated Arginine Methyltransferase 1 (CARM1), also known as PRMT4, was the first identified PRMT that regulates gene transcription. It was identified in 1999 in a two-hybrid screen as a binding partner of the p160 steroid receptor co-activator, glucocorticoid receptor-interacting protein-1 (GRIP1) (Chen *et al.*, 1999), and was later shown to activate transcription together with the histone acetyltransferase p300 (Lee *et al.*, 2002). CARM1 can also influence gene transcription by adding histone methylation marks or acting as a co-activator for several other transcription factors and co-regulators, including c-Fos (Fauquier *et al.*, 2008), NF- κ B (Miao *et al.*, 2006), oestrogen receptor and E2F1 (El Messaoudi *et al.*, 2006), β -catenin (Koh *et al.*, 2002), and p53 (An *et al.*, 2004), reviewed by Bedford *et al.* (2009). CARM1 can also methylate the C-terminal domain of RNA polymerase II (Sims *et al.*, 2011). As discussed in Section 1.9, methyl-marks are recognised by Tudor-domain-containing proteins. As an example, the effector molecule TDR3 binds to the asymmetrically di-methylated Arg17 residue of histone H3 and Arg3 of histone H4 - marks produced by CARM1 (Yang *et al.*, 2010).

Beside its role in epigenetic regulation, CARM1 also regulates splicing events by methylation of different splicing factors (Cheng *et al.*, 2007). It can also methylate and influence the activity of other non-histone proteins such as the chromatin-remodelling factor BAG155 and the nuclear receptor coactivator 3 (NCOA3) (Feng *et al.*, 2006). Its importance in epigenetic and gene regulation is reflected by the fact that CARM1-knockout mice and enzyme inactive CARM1-knockin mice die after birth (Yadav *et al.*, 2003, Kim *et al.*, 2010a).

In human cells, two isoforms exist, the full-length CARM1 gene (CARM1FL) which has 16 exons and CARM1 Δ 15, in which exon 15 is excluded through alternative splicing (Wang *et al.*, 2013a). Both isoforms, show methyltransferase activity, but in the case of CARM1 Δ 15, the CARM1 auto-methylation site is missing. CARM1FL is mainly expressed in brain, heart, skeletal muscle, and testis but CARM1 Δ 15 is the main isoform in all other tissues (Wang *et al.*, 2013a). The same study also identified CARM1 Δ 15 as the main isoform in breast cancer tissue. However, it was previously shown that CARM1 Δ 15 has impaired co-activator activity for ER α transcription regulation for which auto-methylation is necessary (Kuhn *et al.*, 2011).

CARM1 is similar to other PRMTs in that they share a highly conserved catalytic core domain which includes the SAM binding and MTase domains. However, compared to other Type 1 PRMTs, CARM1 sequence conservation is lower: CARM1 shares only 34 % sequence identity whereas other members typically have almost 50 % (Schluckebier *et al.*, 1995a). Different crystal structures of the CARM1 catalytic domain have been solved to collectively show that it adopts the typical Type 1 PRMT fold (Section 1.4) but also has some unique structural features (Tewary *et al.*, 2019, Boriack-Sjodin *et al.*, 2015).

The overall dimeric mouse CARM1 structure in complex with SAH (PDB: 5IH3) superposed with rat PRMT1 is shown in Figure 6-1.

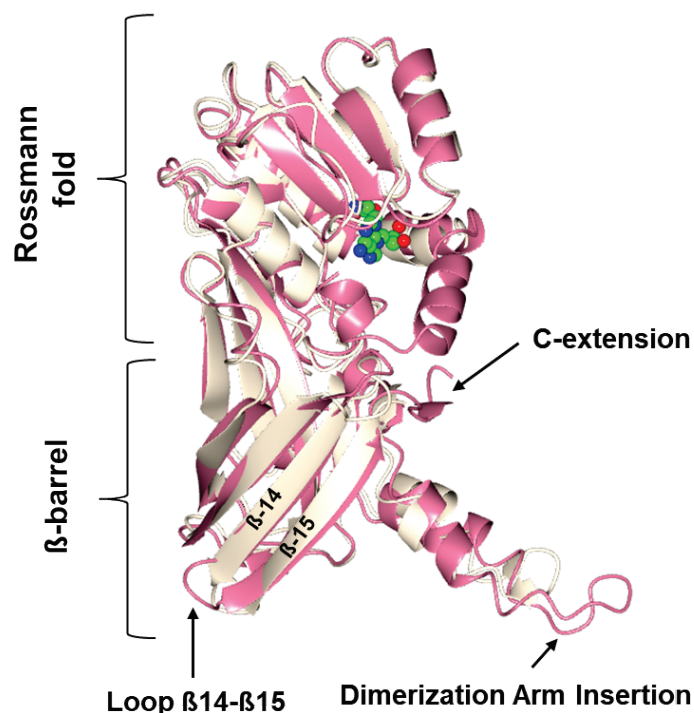


Figure 6-1 Unique Features of the Catalytic Domain of CARM1.

Superposition of rat PRMT1 (cream ribbon) and mouse CARM1 (pink ribbon) in complex with SAH (green ball and sticks). In CARM1 nine additional amino acids are inserted in the dimerisation arm and it has an additional extension at the C-terminus of the catalytic domain. There are also differences in the loop between β 14 and β 15 (MmPRMT2: 5IH3, RnPRMT1: O1R8).

The catalytic CARM1 domain has an additional C-terminal extension which is located after the last β -strand of the β -barrel domain and nine additional residues are inserted in the dimerisation arm. Moreover, differences in the loop between β -14 and β -15 can be seen. The C-extension changes the size of the central dimer cavity, which

is larger in CARM1 and the additional residues in the dimerisation arm increase the overall dimer size compared to other Type 1 PRMTs (Yue *et al.*, 2007). The C-extension incorporates a unique aromatic motif (Pro472-Phe473-Phe474-Arg475-Tyr476, human CARM1 numbering) which interacts with residues of $\beta 7$ and is conserved across all CARM1 proteins (Yue *et al.*, 2007). Deletion of this motif abolishes CARM1 methyltransferase activity on histone H3 Arg17 (Yue *et al.*, 2007). Compared to PRMT1 and PRMT3 that have a negative surface-charge due to the presence of 20 aspartate and glutamate residues on the protein surface, CARM1 only has 5 of the acidic patch residues and as a result, is less negatively charged (Zhang and Cheng, 2003, Weiss *et al.*, 2000).

The catalytic domain is sufficient for protein dimerisation (via the dimerisation arm) and for GRIP1 binding. Moreover, methyltransferase activity and the N- and C-terminal domains were shown to be necessary for CARM1's co-activator function (Teyssier *et al.*, 2002). The N- and C-terminal domains of CARM1 differ from other PRMT family members (Figure 1-7) and, to date, there are no crystal structures for the full-length protein as the domains are missing in the electron density maps and are most likely disordered (Troffer-Charlier *et al.*, 2007, Yue *et al.*, 2007, Boriack-Sjodin *et al.*, 2015) The C-terminal domain of CARM1, also known as activation domain (AD), has its own autonomous activation activity that is needed for CARM1's coactivator function but not for methyltransferase activity or specificity (Teyssier *et al.*, 2006a). The AD in the C-terminus of CARM1 was shown to interact with the transcriptional intermediary factor 1 α (TIF1 α) and, together with GRIP1, form a stable ternary complex. It was speculated that the AD domain contributes to the co-activator assembly complex by the formation of additional protein-protein interactions but more research is needed to elucidate its role (Teyssier *et al.*, 2002, Teyssier *et al.*, 2006b).

The isolated N-terminal domain of mouse CARM1 has been crystallised (Figure 6-2 (Troffer-Charlier *et al.*, 2007)). It crystallised as a dimer and is also a dimer in solution (Figure 6-2, Troffer-Charlier *et al.*, 2007). The N-terminal domain adopts a pleckstrin homology domain (PH)-like fold. PH domains are approximately 100-120 amino acids in size and are common in proteins involved in signal transduction pathways. The common fold consists of 7 β -strands folded into two almost perpendicular β -sheets which form a β -barrel like structure, followed by a C-terminal amphipathic α -helix (Lemmon *et al.*, 1996, Lemmon, 2007). The three loops located between β -strands $\beta 1$ and 2 (VL1), $\beta 3$ and $\beta 4$ (VL2), and $\beta 6$ and $\beta 7$ (VL3) vary greatly in length. The dimer

interface is formed by β 5-7 of both monomers, these strands are known ligand binding sites in all PH domains (Troffer-Charlier *et al.*, 2007). Within the superfamily of PH domain-containing proteins, CARM1 has very high similarity to the family of *Drosophila*-enabled/vasodilator-stimulated phosphoprotein homology 1 (Evh1) domains which are found in many multi-domain signalling proteins (Ball *et al.*, 2002). The EHV1 domains are always N-terminal and bind proline-rich motifs with weak affinity but high specificity via conserved surface exposed aromatic residues. These typical aromatic clusters are missing in the CARM1 N-terminal domain (Ball *et al.*, 2002, Troffer-Charlier *et al.*, 2007).

The CARM1 PH domain is required for the recognition and methylation of most CARM1 substrates and can directly interact with substrates that contain proline-rich motifs, for example BAG155, PABP1 and NCOA3 (Shishkova *et al.*, 2017). The authors of the study also speculated that the disordered PH domain, found in full-length CARM1 crystal structures, might assist in CARM1 substrate capturing and fold into the PH domain fold after substrate binding. Thus, it would be useful to try to co-crystallise full-length CARM1 with substrates that bind the PH domain (Shishkova *et al.*, 2017).

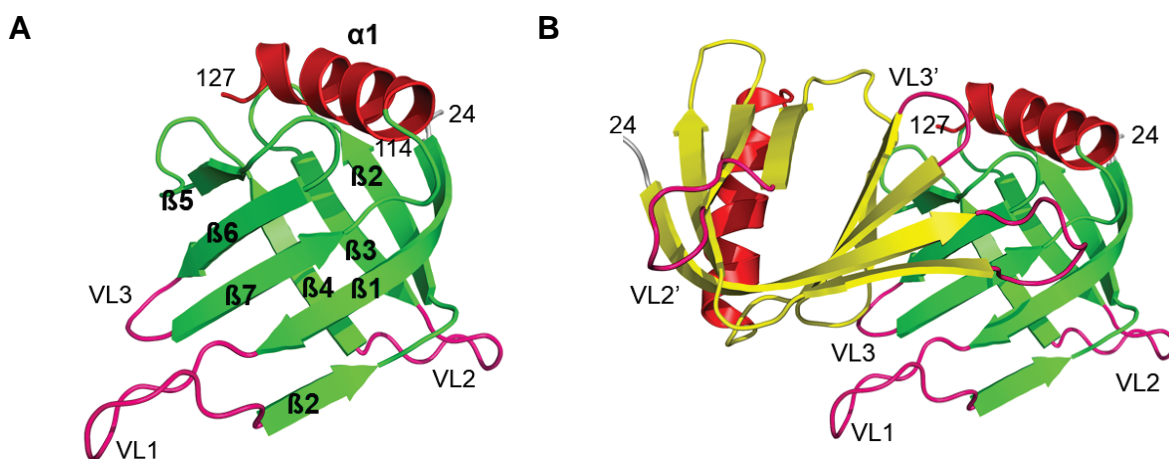


Figure 6-2 N-terminal PH-Domain of CARM1.

A) CARM1 PH domain monomer. Seven anti-parallel β -strands 1-7 (green) form a pseudo β -barrel fold followed by a C-terminal amphipathic α -helix (red). The loops between the β -strands highlighted in pink vary in size and conformation in known PH domains. **B)** Non-crystallographic dimer: During purification the PH domain of CARM1 behaves as a dimer. The dimerisation interface is formed by β 5 - β 7 of each monomer, hiding the ligand-binding site of other PH domains (PDB: 2OQB).

CARM1 overexpression has been reported in a variety of different cancer types, including prostate and colorectal cancer (Kim *et al.*, 2010b), lung cancer (Elakoum *et al.*, 2014), and breast cancer (Morettin *et al.*, 2015). Overexpression of the protein leads to the activation of a variety of oncogenic pathways including E2F1, WNT- β -catenin and AIB1 (Yang and Bedford, 2013). For this reason, a variety of different CARM1 inhibitors are currently being developed (Section 1.12.2). However, it needs to be assessed whether direct PRMT inhibition is desired or if *in vivo* toxicity will be observed considering the multiple roles of PRMTs in the cell (Yadav *et al.*, 2003, Kim *et al.*, 2010a).

6.1 PRMT Inhibitor Design

A variety of PRMT inhibitors are currently being developed, the advances over the last year have already been summarised in Section 1.12. The different mechanisms of action of PRMT inhibitors are illustrated in Figure 6-3 (Luo, 2015). Most of them target the enzyme substrate binding site and compete directly with substrate. SAM/SAH-uncompetitive (also known as SAM-dependent) and non-competitive inhibitors can be distinguished. An example of a SAM-uncompetitive PRMT inhibitor is the PRMT5 inhibitor GSK3368715 (Fedoriw *et al.*, 2019). In contrast, the CARM1 inhibitor MS023 has a SAM/SAH-non-competitive mechanism of action (Eram *et al.*, 2016).

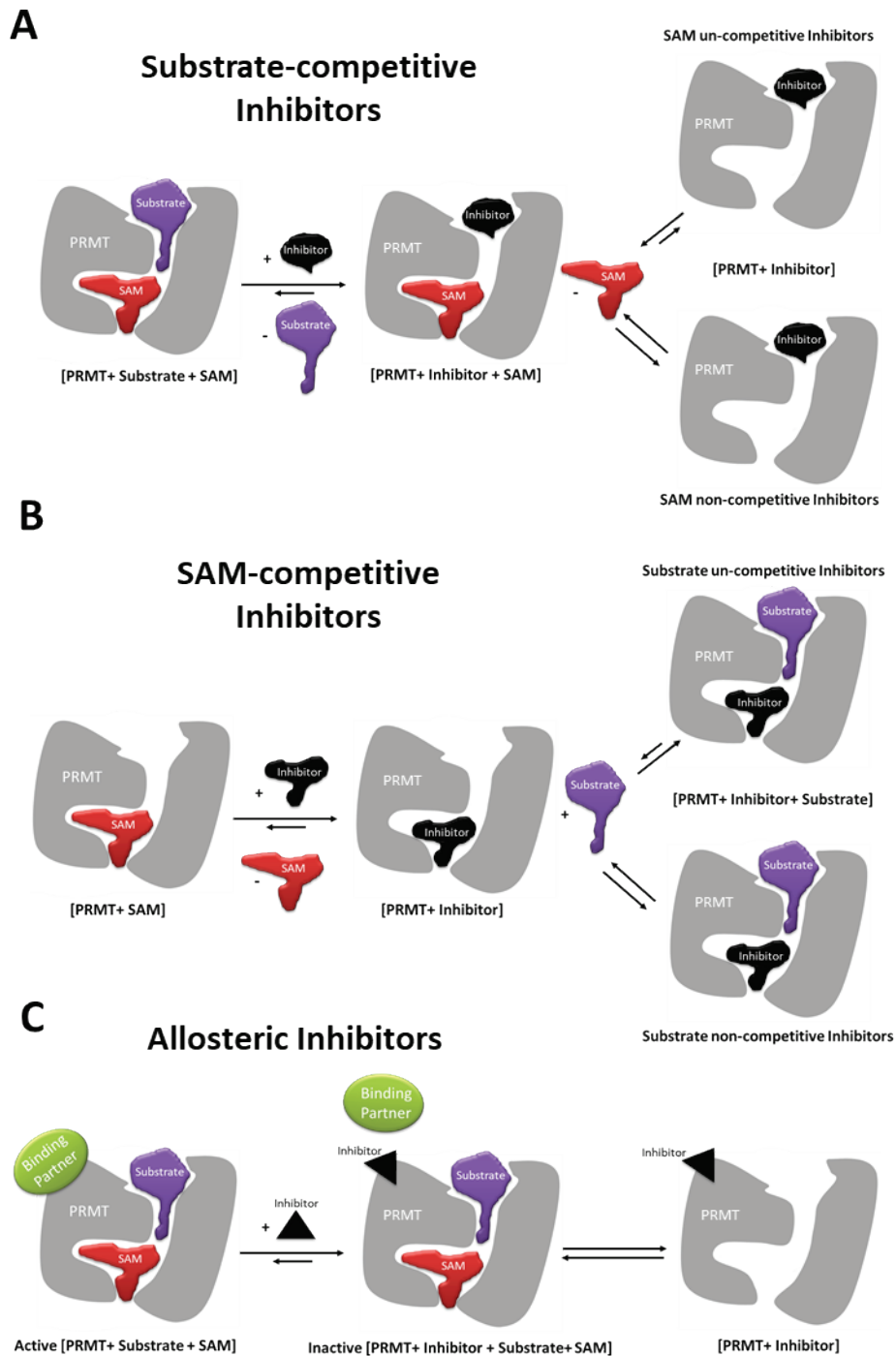


Figure 6-3 Mechanism of Action of PRMT Inhibitors.

Three different classes of PRMT inhibitors have been characterised: **A** Substrate-competitive, **B**) SAM-competitive and **C**) Allosteric inhibitors. Substrate-un-competitive and SAM-un-competitive mechanisms have also been observed for PRMT inhibitors.

Fragments have also been developed that directly or indirectly target the SAM binding site or occupy allosteric pockets. These compounds include a PRMT3 and PRMT6 inhibitor (Section 1.12.4). The locations of the distinct two binding sites are illustrated using the structure of human CARM1 where the substrate binding pocket and cofactor binding site are occupied by a histone mimic and by Sinefungin respectively (PDB: 2Y1W, Figure 6-4). As already described in detail in Section 1.5, the α -helix, located at the N-terminus of the Rossmann fold, folds over the cofactor (mimic) and participates in the formation of the substrate binding site.

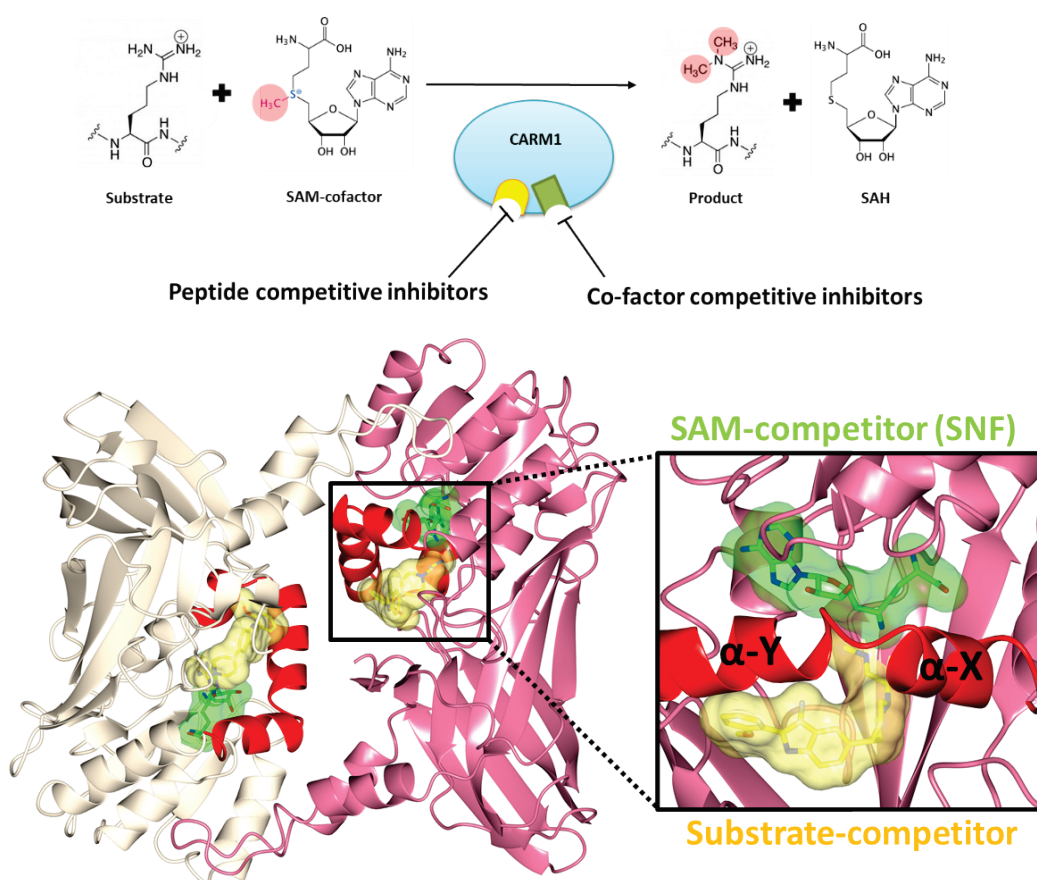


Figure 6-4 Targeting CARM1 Binding Pockets.

CARM1 catalyses the methyl-transfer from the cofactor SAM to the guanidino group of the substrate arginine residue. Different CARM1 inhibitors target the peptide and SAM-binding pocket. The dimeric crystal structure of the catalytic domain of CARM1 (PDB: 2Y1W) in complex with the SAM-competitive inhibitor Sinefungin (green) and a substrate-competitive inhibitor (yellow), and their binding pockets, is shown.

In the following sections the structural diversity of the two targeted CARM1 binding pockets will be discussed.

6.2 Targeting the CARM1- Cofactor Binding Site

An overview of the cofactor binding site of human CARM1 (PDB: 2Y1W) in complex with the SAM-mimic Sinefungin and analysis of its sequence conservation is shown in Figure 6-5. The binding pocket of the very polar cofactor SAM is mostly hydrophilic, but small lipophilic patches are present. The high hydrophilicity of the pocket will make the design of cell-permeable inhibitors challenging, because they need to bind the SAM pocket with high affinity but also need some lipophilicity to cross the cell membrane. The cofactor SAM has a calculated partition coefficient ($\log P$) value of -5.8 and a polar surface area (PSA) value of 182.6 Å² (Ferreira de Freitas *et al.*, 2019). The $\log P$ value is defined as the ratio between the concentration ratio of the ligand in octan-1-ol: aqueous phase. Thus, positive $\log P$ values mean that the ligand is more hydrophobic, negative values more hydrophilic (Bhal, 2007). Traditionally, small molecule inhibitor development aims for a $\log P$ value < 5, in agreement with “Lipinski’s Rule of 5” (Lipinski *et al.*, 1997). However, fragments are most often assessed against the Astex “Rule of Three” which proposes that fragments should have a mass ≤300 Da, ≤3 hydrogen bond acceptors and donors, ≤ 3 rotatable bonds, a $\log P$ ≤ 3 and a PSA < 60 Å² (Congreve *et al.*, 2008, Congreve *et al.*, 2003). SAM and the pan-inhibitor Sinefungin have similar $\log P$ and PSA values (-5.1, 209 Å², calculated with ChemAxon), explaining the low cell-permeability of both. Thus, it is important when designing new SAM-competitive inhibitors to target spaces which show higher $\log P$ and lower PSA values.

Figure 6-5 A shows the side-chains located within 5 Å of the SAM-competitive inhibitor Sinefungin in human CARM1 (PDB: 2Y1W). Figure 6-5 B shows the key residues that are involved in SAM binding (CARM1 numbering). A sequence alignment of residues in the SAM binding pocket of different Type 1 PRMTs is shown in Figure 6-5 C (Schapira and de Freitas, 2014). Sequence conservation in the pocket is generally high, but a few positions including Phe137 (Phe138 in the CARM1 crystal structure), Met163, and Ala215 are diverse and ligand interaction with the residues might drive selectivity. The CARM1 crystal structure shows that Phe137 is located on the α -X element of the α -helix that is located at the N-terminus of the Rossmann fold, and Met163 is on the α -Y element. In the position equivalent to human CARM1 Phe137, human and mouse PRMT2 have a glutamic acid residue (Glu90/Glu102 respectively), which is an arginine residue (Arg35) in human PRMT6, and a lysine (Lys41) in human PRMT1. Analysis of the pocket geometry and hydrophobicity previously performed by Campagna-Slater *et al.*, indicated that the cofactor binding site is druggable but in the

case of CARM1 less than 5 % of the SAM surface is accessible to the solvent (Campagna-Slater *et al.*, 2011).

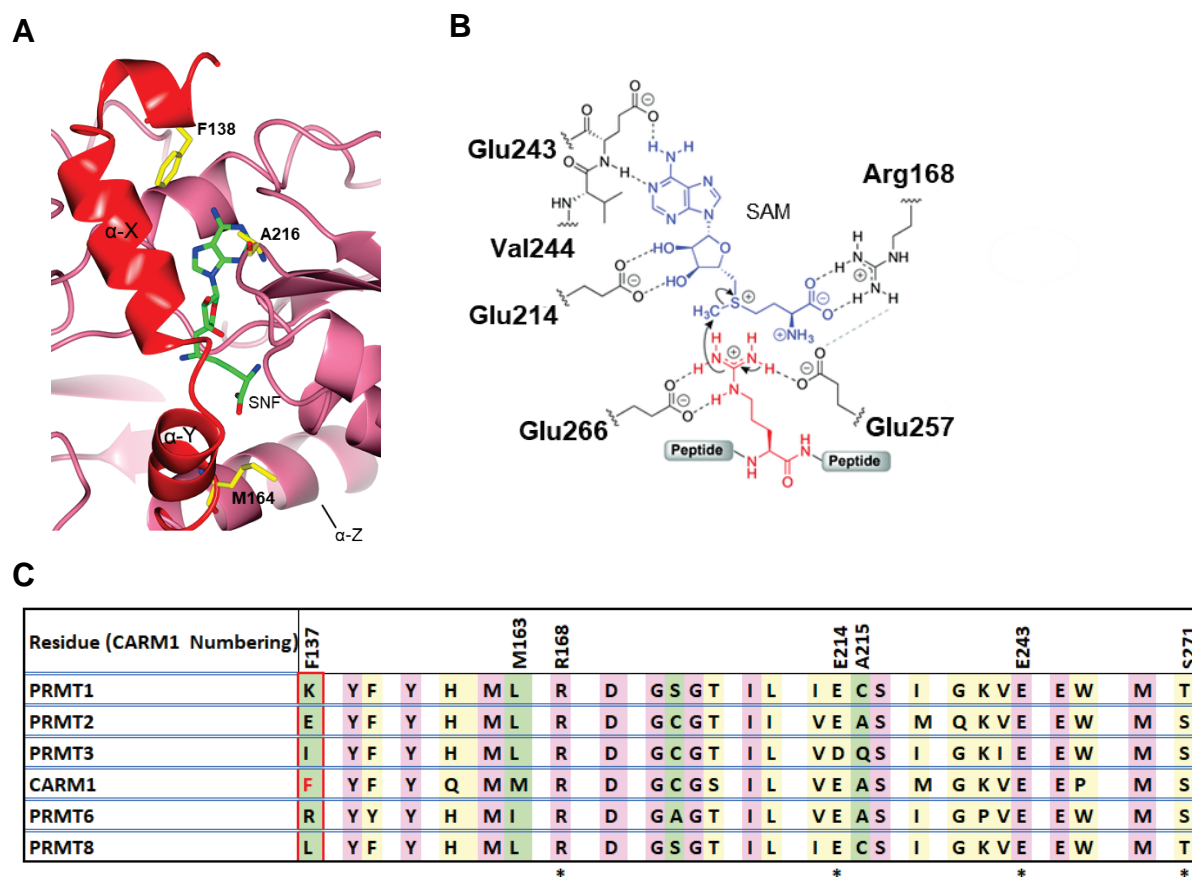


Figure 6-5 Structural Diversity of the Cofactor Binding Pocket.

A) Human CARM1 structure (purple ribbon) in complex with Sinefungin (carbon atoms in green). Residues with low sequence conservation are shown in yellow. The α -X helix is shown as red ribbon. **B)** Position of the SAM Pharmacophore in the Active Site. Key residues are highlighted (HsCARM1 numbering). The cofactor SAM is coloured in blue, the arginine residue of the substrate peptide red. Figure adapted, with permission, from van Haren *et al.* (2017). **C)** Sequence diversity of the SAM-binding pocket of the PRMT family. Via structure-based multiple sequence alignment the side-chains in 5 Å proximity of the cofactor SAM were analysed (Schapira and de Freitas, 2014). Selected CARM1 residues are highlighted. The asterisk indicates residues that form direct hydrogen bonds with SAM (pink= high sequence similarity; green = low; yellow = medium).

The SAM pocket shows little sequence diversity across the wider SET domain-containing PMTs (Section 1.2). When PMTs were clustered based on the conformational diversity of SAM/SAH as observed in structures deposited in the PDB, the conformation of the bound cofactor was highly similar to that seen in other DNA and RNA methyltransferases including the histone methyltransferase DOT1L (Campagna-Slater

et al., 2011). As a consequence, obtaining selective inhibitors for the PRMT SAM binding site could be difficult and the risk of off-target effects is high, similar to the kinase ATP binding sites for which selective inhibitors exist (Rees, 2016).

Developing ligands that target the SAM binding site is also challenging because the intra-cellular concentration of the cofactor is very high (approx. 20–40 μM) (Copeland, 2018). The high SAM concentration needs to be taken into account when designing cell-free ligand screening assays, otherwise apparent binding affinities could be drastically reduced when testing the ligands in a cell-based assay where the SAM concentrations are higher than under *in vitro* assay conditions (Copeland, 2018). However, potential initial fragment hits that bind with μM affinity to the binding site might not be detected if the concentration of the authentic cofactor is too high (Rees, 2016). Consequently, PRMTs with low binding affinity for the cofactor might be better targeted with peptide-site directed or bisubstrate competitive inhibitors.

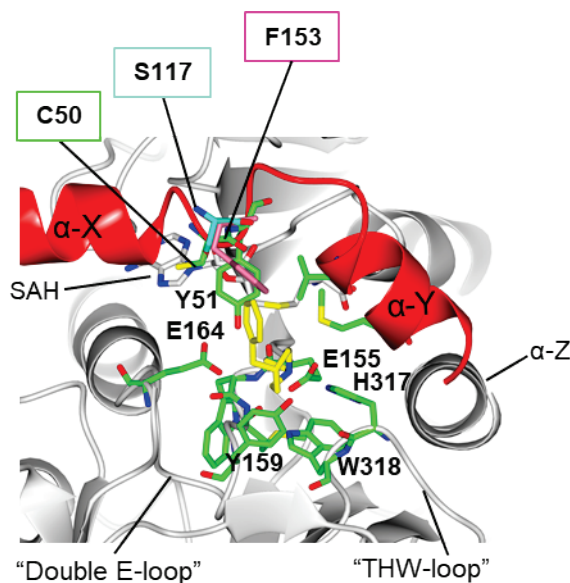
However, selective, cell-permeable SAM-competitive inhibitors have been successfully developed for a few PMTs including Tazemetostat (TAZVERIK, Epizyme, Inc.) a first-in-class inhibitor of the PKMT Enhancer of Zeste Homolog (EZH) 2 which was recently approved by the FDA for advanced epithelioid sarcoma treatment (Knutson *et al.*, 2014, Italiano *et al.*, 2018). Other SAM-competitive inhibitors that have entered clinical trials target EZH1/2 (Toshiki Watanabe, 2016), and DOT1L (Daigle *et al.*, 2013, Stein *et al.*, 2018). Three potent and selective PRMT5 inhibitors that share a deazapurine ring and show cell-activity have already been mentioned (LLY-283, JNJ-64619178, and PF-06939999). JNJ-64619178 has very high potency (IC_{50} *circa* 0.13 nM; EC_{50} *circa* 0.25 nM) and in the absence of a structure a recent modelling study proposed that the inhibitor not only occupies the SAM pocket but also extends into the arginine side channel in a pseudo-irreversible binding mode (Lin and Luengo, 2019). If this is the case, it is likely that it is also substrate-competitive. In contrast, LLY-283 does not reach into the substrate pocket (Lin and Luengo, 2019).

Virtual screening has been successfully used to identify SAM-competitive ligands which are non-nucleosides (Mao *et al.*, 2017). These compounds were less potent but they show higher $\log P$ and lower PSA values, which is very promising (Ferreira de Freitas *et al.*, 2019). In summary, development of selective SAM competitive inhibitors, even if challenging, is achievable.

6.3 Targeting the CARM1-Substrate Binding Site

A multiple sequence alignment of residues within 5 Å of a potent human PRMT6 inhibitor that occupies the substrate-binding pocket (PDB: 5EGS) showed high sequence conservation within the substrate arginine channel, but significant diversity at the distal region of the pocket (Figure 6-6 B).

A



B

Residue (CARM1 Numbering)	F152	E257	E266	H414
PRMT1	Y F D S Y A H F G	I S E W M G Y C L F Y Q S M L N T V L		H W K Q T V F Y
PRMT2	Y F G S Y G T L K	V S E W M G T C L L F E F M I E S I L		H W K Q T L F M
PRMT2	Y F D S Y G T L K	V S E W M G T C L L F E F M I E S I L		H W K Q T L F M
CARM1	Y F Q F Y G Y L S	I S E P M G Y M L F N E R M L E S Y L		H W Y Q V R C L
PRMT6	Y Y E C Y S D Y S	V S E W M G Y G L L H E S M L S S V L		H W K Q A L L Y

Figure 6-6 Structural Diversity of the Substrate-Binding Pocket.

A) Crystal structure of human PRMT6 (PDB: 5EGS, grey ribbon) in complex with a substrate competitive inhibitor (carbon atoms in yellow). High sequence diversity at position 50 (cysteine in PRMT6), should drive selectivity: serine residue in PRMT2 (PDB: 5FUL), phenylalanine in CARM1 (PDB: 3B3F). Important binding site residues are shown with carbon atoms coloured green. The α -helix is shown as a red ribbon. **B)** Sequence diversity of the substrate-binding pocket of selected PRMT family members. Via structure-based multiple sequence alignment the side-chains in 5 Å proximity of the substrate inhibitor were analysed. Residues that make direct hydrogen bonds with the inhibitor are marked with an asterisk, they include the two glutamic acid residues of the “Double-E-loop” (Glu257 and Glu266, CARM1 numbering) and His414 of the THW-loop.

The position equivalent to Cys50 in HsPRMT6 is occupied by Ser117 in MmPRMT2 and Phe152 (Phe153 in the crystal structure) in HsCARM1. The substrate

binding sites of PRMTs are not only much more diverse than the SAM binding pocket they are also more hydrophobic. Thus, the development of less polar compounds with higher target-specificity is more likely.

In the next section, the production of the CARM1-SAH crystals that were used for ligand soaking experiments is described, followed by a short overview of the SPR binding assay that was used to rank the designed fragments. Thereafter, the results of the characterisation of the fragments that were designed to target either the SAM or substrate binding sites, or both sites simultaneously will be presented.

6.4 Crystallisation Trials of the Catalytic Domain of HsCARM1

Heterologous expression of human CARM1 in insect cells yielded sufficient pure protein to carry out crystallisation trials (Figure 6-7, Section 3.5.3). For initial crystallisation trials, the purified protein was used at 2.4 and 7 mg/mL and incubated +/- 0.2 mM SAH for 2 h at 4 °C. Hampton Index crystal screens were performed at 20 °C and 4 °C. All samples were centrifuged at 10 000 g for 10 min at 4 °C to remove insoluble particles before setting up the trays.

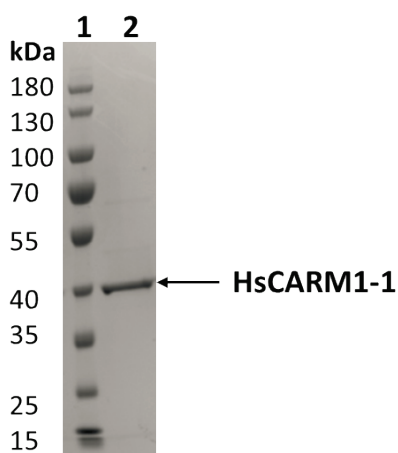


Figure 6-7 HsCARM1-1 Crystallisation Trials. Protein used for Crystallisation Experiments.

12 % SDS-PAGE gel stained with InstantBlue dye. Lane 1, PageRuler protein ladder; Lane 2, HsCARM1-1 (MW= 39.78 kDa).

The HsCARM1-1 construct comprises only the catalytic domain (amino acids 135-482) and thus will be named HsCARM1^{CAT}, it has been successfully crystallised previously (Sack *et al.*, 2011). The plates were analysed for crystal growth every few

days and the first crystals appeared after two days. In total, eight crystals were harvested and tested at the Diamond Light Source, UK. All CARM1^{CAT} crystals incubated with SAH diffracted, the best to 1.96 Å. In contrast, crystals with no SAH bound showed only weak diffraction to a resolution of 2.7-3.4.

Table 6-1 Growth and Harvest Conditions for HsCARM1^{CAT} Crystals used for Structure Determination

	Protein Complex	Ligand	Growth Conditions	Harvest Conditions
A	HsCARM1 ^{CAT} (2.4 mg/mL)	SAH	0.2 M Ammonium sulfate, 0.1 M HEPES pH 7.5, 25 % (w/v) PEG 3,350	15 % (v/v) Glycerol
B	HsCARM1 ^{CAT} (2.4 mg/mL)	SAH	0.2 M Lithium sulfate monohydrate, 0.1 M HEPES pH 7.5, 25 % (w/v) PEG 3,350	20 % (v/v) Glycerol
C	HsCARM1 ^{CAT} (7 mg/mL)	None	0.1 M BIS-TRIS pH 6.5, 28 % (w/v) mPEG 2,000	-
D	HsCARM1 ^{CAT} (7 mg/mL)	SAH	0.2 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 25 % (w/v) PEG 3,350	15 % (v/v) Glycerol
E	HsCARM1 ^{CAT} (7 mg/mL)	SAH	0.1 M BIS-TRIS pH 5.5, 25 % (w/v) PEG 3,350	20 % (v/v) Glycerol
F	HsCARM1 ^{CAT} (7 mg/mL)	SAH	0.1 M HEPES pH 7.5, 25 % (w/v) PEG 3,350	20 % (v/v) Glycerol

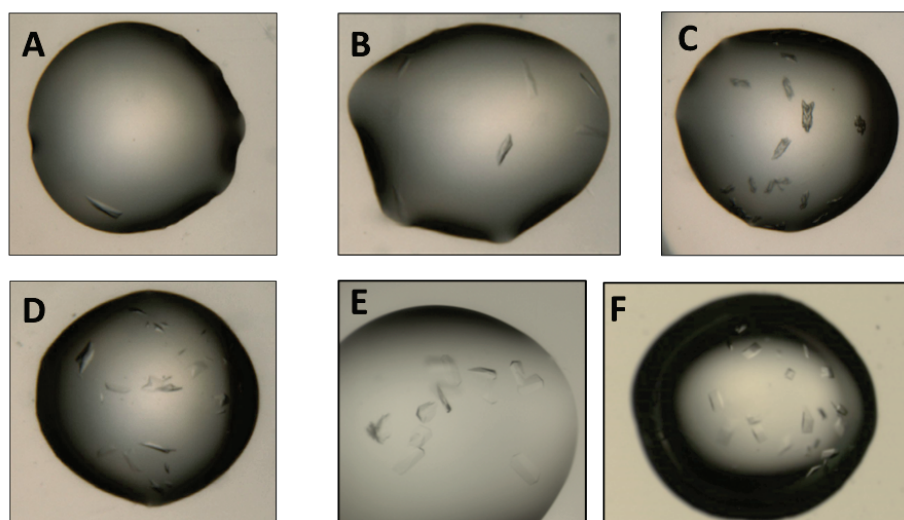


Figure 6-8 HsCARM1^{CAT} Crystallisation Trials with and without SAH.

CARM1 crystals were observed after a few days in different buffer conditions (Table 6.1 A-F).

Structures were determined for CARM1 in the presence and absence of added SAH. The data was collected at the Diamond Light Source (Didcot, UK) on I03 beam-line by Dr Arnaud Basle (Newcastle University, UK). However, despite adding no SAH

to the protein, all structures showed density for SAH in the cofactor binding pocket, which is most likely a result of intracellular SAH uptake during heterologous expression. It is often reported that substrate-competitive inhibitors are only able to bind and inhibit the PRMT if SAH or SAM is bound in the cofactor pocket. One reason is that especially in the case of CARM1, SAM or SAH binding is needed for structural rearrangements and the formation/stabilisation of the substrate-binding pocket (Section 1.4). Additionally, some inhibitors were shown to interact directly/indirectly with the cofactor or with the reaction by-product SAH (Ferreira de Freitas *et al.*, 2019). One example is the PRMT5 inhibitor EPZ015666 (Chan-Penebre *et al.*, 2015): cation- π -interactions are formed between its phenyl ring and the sulfonium methyl-group of the cofactor. This observation highlights that crystal studies which are commonly performed in the presence of SAH should in some cases also be done using SAM (Ferreira de Freitas *et al.*, 2019).

In the initial CARM1 crystal screen, crystals grown in either 0.1 M BIS-TRIS pH 5.5, 25 % (w/v) PEG 3350 or 0.1 M HEPES pH 7.5, 25 % (w/v) PEG 3350 using the higher CARM1 concentration, diffracted the best. In the next step, optimisation screens were performed. One screen varied PEG 3350 (21- 31 %) and AmSO₄ (0-0.3 M) concentrations in 0.1 M HEPES buffer pH 7.5, another screened pH (from 7.5-5.5 with 0.1 M HEPES or Bis-Tris buffer), against PEG 3350 (21- 31%). However, crystal diffraction did not improve, so the original conditions were used to grow CARM1-SAH crystals for ligand soaks (Section 2.22.4).

Prior to performing crystal-ligand soaks, the DMSO tolerance of the CARM1-SAH crystals was determined, as described in Section 2.22.3. Diffraction quality was not affected by a DMSO concentration up to 15 % and it was decided to use ligand solutions with 10 % DMSO for the soaking experiments. Beside structural studies, the fragments were screened via SPR, the principle will be discussed in the next section.

6.5 SPR Analysis

Surface Plasmon Resonance analyses binding between two non labelled molecules in real-time. One of the binding partners is immobilised on a sensor chip surface, while the other is free in solution and is flown over the chip surface. The technology is based upon a quantum physical phenomenon that occurs on the sensor chip and which is visualised by an optical detector (Figure 6-9). The sensor chip consists of a glass

surface which is covered with a thin gold layer. The CM5 chip has a carboxymethylated dextran surface which forms a hydrophilic matrix that can covalently bind biomolecules. The optical detection unit consists of a electroluminescent diode, prism and a photodiode array. An integrated flow channel that has direct contact with the chip surface is used to transport the injected sample to and from the chip surface. The monochromatic and polarised light is positioned is such way, that it passes through the glass prisms (medium with higher refractive index) into the gold foil covered flow cell (medium with lower refractive index), and a total internal reflection occurs. Under total internal reflection conditions, so called evanescent waves are produced in the lower refractive index medium. At a specific wavelength and incident angle of the light beam, the waves excite a surface plasmon at the surface of the metal film named surface plasmon resonance (SPR). This leads to a decrease in the intensity of the light which is reflected at a specific angle from the sensor surface. Binding of molecules at the chip surface, causes a change of the refractive index and causes an alteration of the angle of minimum reflected intensity (resonance angle).

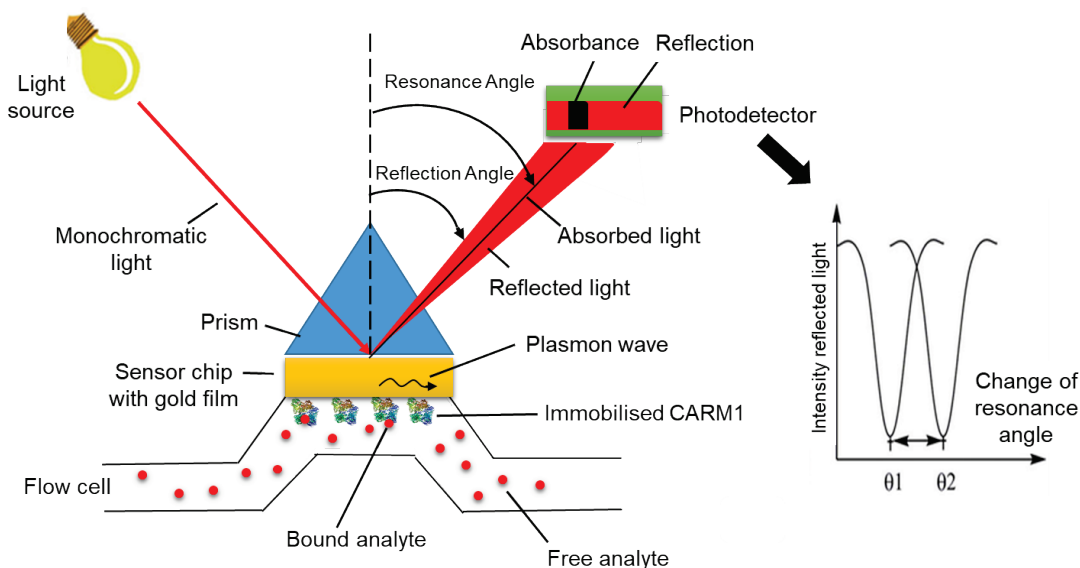


Figure 6-9 Principle of Surface Plasmon Resonance Analysis.

The optical unit consists of a light source that produces a monochromatic light beam that passes through a glass prism and is totally reflected at the sensor surface. At a specific wavelength and incident angle, the evanescent waves, excite the valence electrons of the sensor chip surface and a surface plasmon is produced (Surface Plasmon Resonance). The SPR causes a decrease in reflected light beam intensity. The detector records the intensity of reflected light and calculates the resonance angle. Binding of the analyte to the immobilised ligand causes an increase in refractive index and thus a change in resonance angle. The change of the resonance angle is detected by the SPR instrument in real-time.

The sensor detects the change in refractive index that occurs during the interaction reaction of the immobilised ligand and the mobile binding partner. The change in SPR angle is proportional to the mass of material bound to the chip surface and is expressed as resonance units (RU). One RU unit is defined as the mass of 1 pg/mm^2 and corresponds to a change of resonance angle of $\sim 0.00001^\circ$. The chip consists of four cells, of which one is used as a reference cell. The reference signal which shows the unspecific binding to the chip surface is then subtracted from the measured binding signal (Myszka, 1997, Jason-Moller *et al.*, 2006, Tang *et al.*, 2010). Measuring the change in the SPR signal over time and plotting the RU against time, produces a graph, a so called sensogram (Figure 6-10).

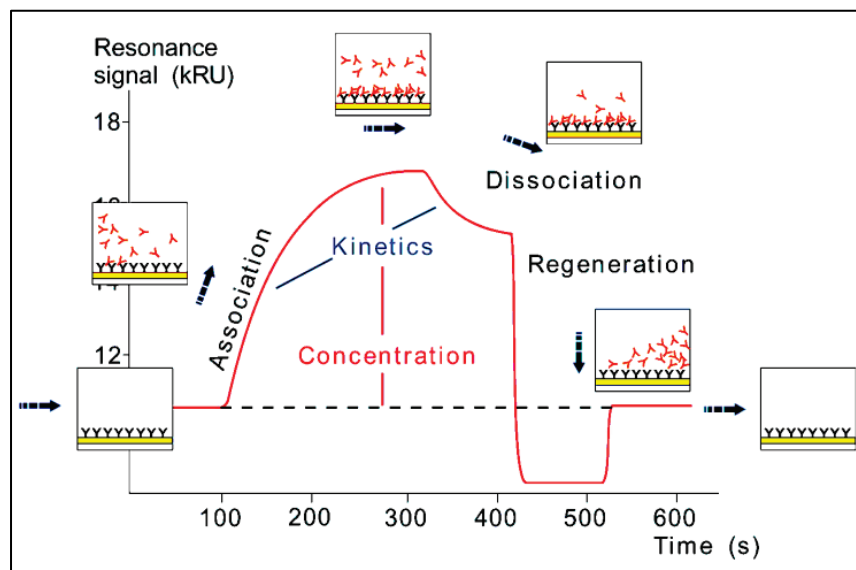
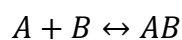


Figure 6-10 SPR Sensorgram.

The response units are measured over time and a plot is generated (sensorgram). The sensorgram can be divided into different phases. During the association phase, the analyte that is flowed over the chip surface binds to the immobilised ligand until a binding equilibrium has been reached (steady-state phase). Next, buffer is flowed over the surface which causes the dissociation of the binding complex and a decrease of RU signal.

A sensogram consists of three different phases: association, steady-state and dissociation. After analyte injection the RU signal increases due to complex formation by the binding partners.

The complex formation between two binding partners A and B can be described with the following equation:



Equation 6.1

The curve flattens when the steady-state phase is reached, where the number of molecules that form a binding complex equals the number of molecules that dissociate. During the dissociation phase, the complexes dissociate. Analysis of the different phases gives information about the association rate (k_a), dissociation rate (k_d) of the molecules and their binding affinity (K_d). SPR assays are often used for initial fragment screening but also for hit validation because they produce accurate and fast measurements of binding affinities and kinetics and no labelling is required.

For SPR analysis of fragment binding to CARM1, the catalytic domain of CARM1 (CARM1-1), purified as described in Section 3.5.3, was immobilised on a SPR chip by amine coupling. The method protocol can be found in Section 2.20. An exemplary CARM1 immobilisation sensogram can be seen in Figure 6-11.

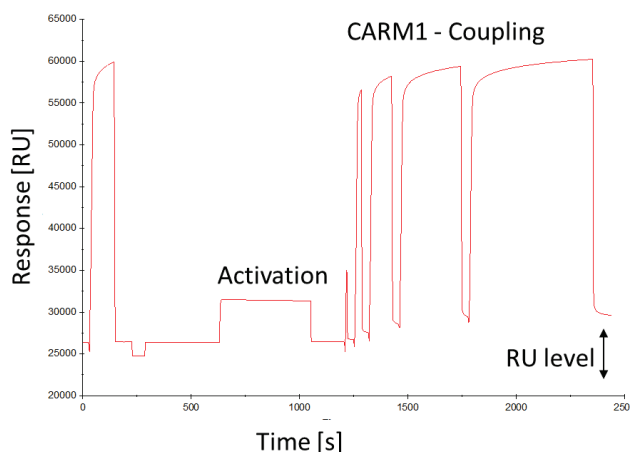


Figure 6-11 HsCARM1^{CAT} Amine-Coupling to CM5 Sensor Chip.

After activation of the chip surface with EDC/NHS, the ligand HsCARM1^{CAT} was immobilised via amine coupling. This step was followed by deactivation of the free esters with ethanolamine (not shown). Immobilisation levels achieved for CARM1 were typically 3500-6000 RU.

One drawback of direct coupling via amine groups is that the ligand is immobilised in a random position, which can result in some binding sites being inaccessible for analyte binding. Moreover, a low pH (pH 5.0-5.8) is needed during the coupling process which could lead to the inactivation or denaturation of some of the ligand. Ligand screening experiments were performed as described in Section 2.20.3.

6.6 Small-Molecules Targeting the Cofactor Binding Pocket

Different fragments that target only the cofactor binding pocket of CARM1 were designed, and binding to CARM1 was analysed by SPR and by co-crystallisation and ligand soaking into existing CARM1-SAH crystals. The chemical structures of SAM and the fragment designed by Dr Edwige Picazo are shown in Figure 6-12. The aim was to maintain the hydrogen bonding motifs of the aminopyrimidine of the adenine portion of the SAM molecule but to modify or even delete the imidazole ring to be able to introduce new substituents that can interact with the surrounding binding pocket residues. The ring oxygen of the ribose moiety did not appear to make productive contacts; thus, it was replaceable with a carbocyclic ring, increasing the stability and reducing the polarity of the analogues.

The aim was to increase the potency of the fragments, once potent and selective warheads have been identified, by growing the molecule towards the methionine binding region of the SAM site. Unfortunately, none of the initial fragments yielded a CARM1-fragment co-crystal structure either by attempting soaking or co-crystallisation experiments. Crystals appeared after a few days but SAH was always bound to the cofactor pocket.

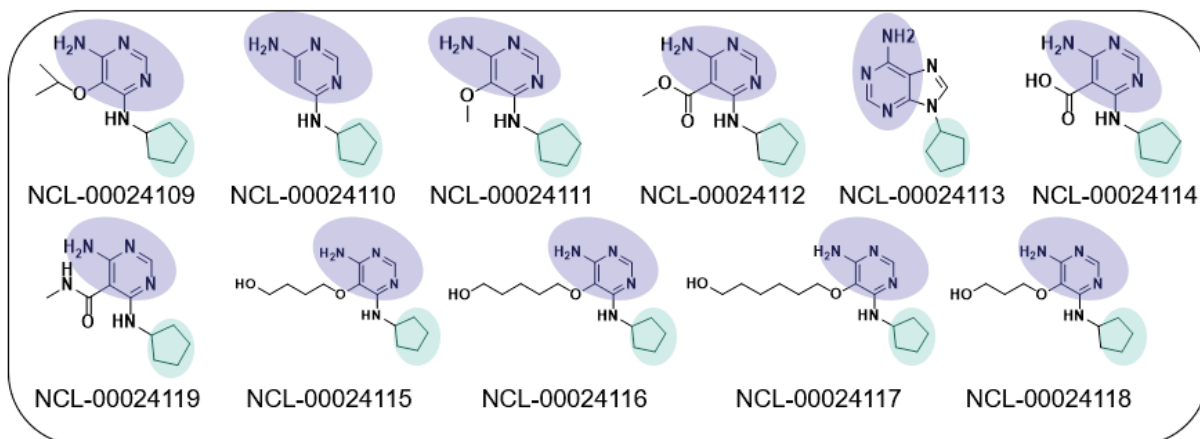
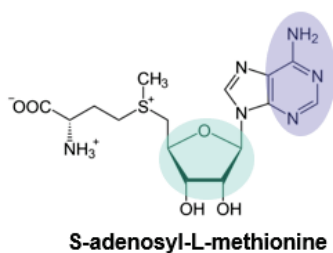


Figure 6-12 Fragments Designed to Target the Cofactor Binding Pocket.

SPR experiments were performed with and without SAH in the buffer, but fragment binding could not be detected indicating that they are either not binding or are not able to compete and replace the bound cofactor. The exception was NCL-00024109 (Figure 6-13). This fragment has an estimated K_d value of 72 μM measured in the absence of SAH in the SPR buffer (Figure 6-13 A). One explanation could be that the SAM binding pocket is not accessible during the SPR experiment due to the immobilisation of CARM1 to the SPR chip. However, the SAM-competitive inhibitor Sinefungin was used as a positive control and gave interpretable results with high response units (Figure 6-13 B). The calculated K_d values for Sinefungin were 5.4 μM (no SAH) and 2.0 μM (plus SAH) in the SPR buffer.

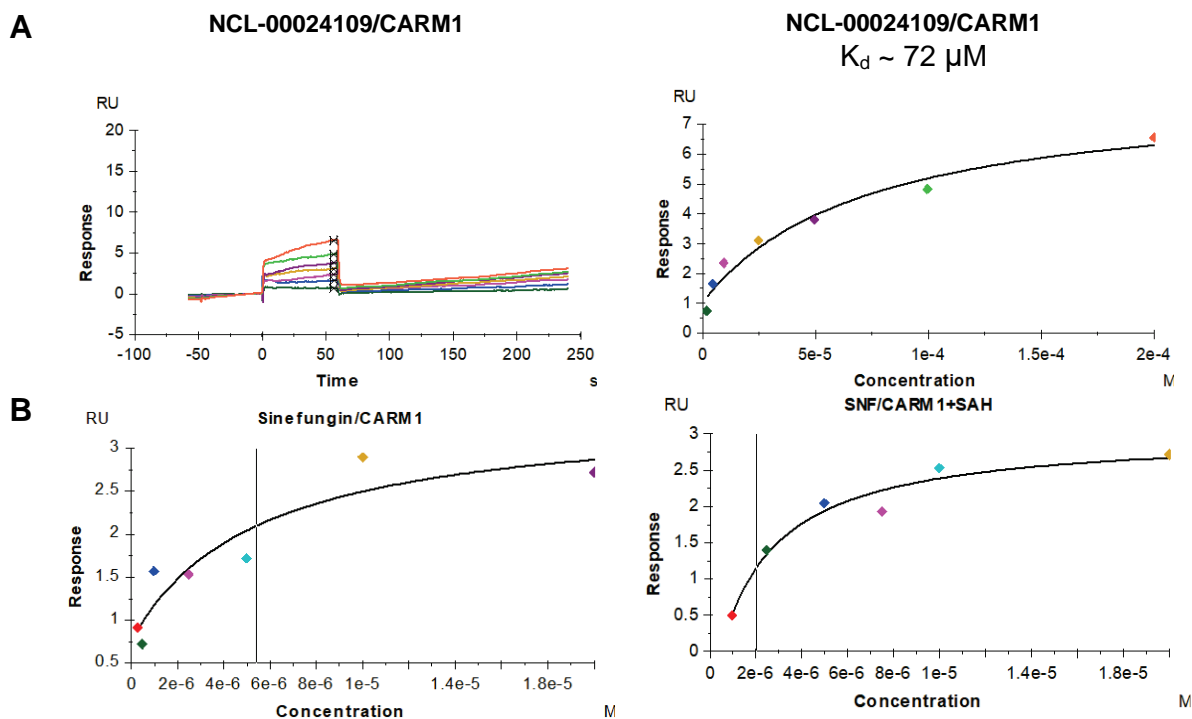


Figure 6-13 SPR Analysis of NCL-00024109 and SNF.

A) SPR analysis of NCL-00024109 binding to CARM1 in the presence of SAM. **B)** SPR analysis of Sinefungin binding to CARM1 without or with SAH in the SPR buffer.

To ensure that the co-crystallisation protocol was working, Sinefungin was co-crystallised with CARM1 using both a literature protocol (Section 2.22.5) and the protocol used to co-crystallise CARM1 with SAM competitive ligands (Section 2.22.6). In both cases, a co-crystal structure was obtained. The CARM1 catalytic domain in complex with SNF crystallised in space group $P2_12_12$ (94.2 % probability by Pointless) and was refined to a resolution of 1.96 Å with an R-factor of 20.2 % and an R-free value of 24.1 %. The crystallographic data and refinement statistics are presented in Table 6-2. Sinefungin is bound to the SAM-binding pocket in the position usually occupied by SAM/SAH (Sack *et al.*, 2011, Boriack-Sjodin *et al.*, 2015) as shown in Figure 6-14. Sinefungin binding to CARM1 has already been discussed in detail in Sections 5.4 and 6.2, and thus will not be discussed here. The CARM1-SNF structure superposed with the published structure very well (RMSD of 0.24 Å over 342 Cα atoms). No conformational differences in the binding pocket could be observed and analysis of the sidechains within 5 Å of the inhibitor showed that they are all in the same orientation.

ITC was also used to confirm the binding of Sinefungin and SAH to CARM1 (described in Section 2.22, Figure 6-14 C and D). Unexpectedly, the number of binding sites was not close to 1 as would be expected for a 1:1 binding interaction. In the case

of Sinefungin titration the number of binding sites was 0.46 and for SAH 0.33. The reason for this is most likely that SAH is already bound to the CARM1 cofactor pocket. This result is in agreement with ITC titrations of SAH and SNF vs CARM1 reported in the literature (Schapira and de Freitas, 2014).

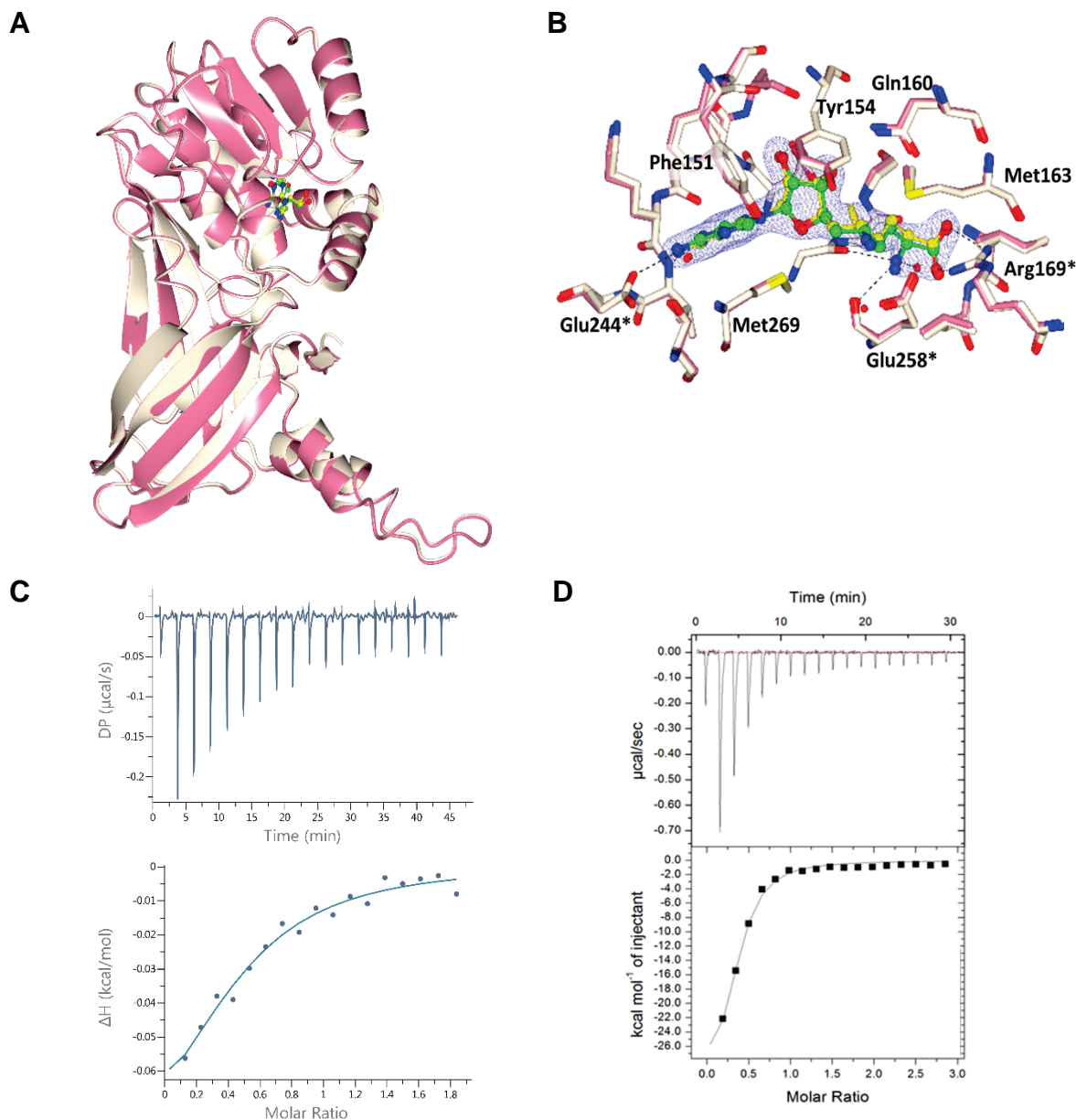


Figure 6-14 Characterisation of the CARM1-Sinefungin Complex.

A, B) CARM1-Sinefungin crystal structure. **(A)** Superposition of the solved CARM1-SNF structure (pink ribbon) with the published structure (cream ribbon, PDB: 2Y1W). The inhibitor is shown with carbon atoms coloured green and yellow, respectively. **(B)** Close up of the residues involved in ligand binding, key residues are highlighted with an asterisk. The $2mF_o-DFc$ map for the solved SNF structure is contoured around the ligand at 1σ . **(C, D)** ITC experiments were performed at 20°C with $20 \mu\text{M}$ CARM1 in the cell and $200 \mu\text{M}$ Sinefungin in the syringe **(C)**, or $15 \mu\text{M}$ CARM1 in the cell and $230 \mu\text{M}$ SAH in the syringe **(D)**. Experiments were carried out once.

Table 6-2 Crystallographic Data and Refinement Statistics of Human CARM1 in Complex with Sinefungin.

X-ray diffraction data were collected from a single crystal on beamline I03 at the DIAMOND Light Source (Didcot, UK) at -180 °C. Data processing was performed using Xia2 (3dii). Statistics for the highest resolution shell are shown in parentheses.

Ligand	Sinefungin
Wavelength (Å)	0.980
Data processing	
Resolution Range (Å)	70.23- 1.96 (2.01-1.96)
Space Group	P2 ₁ 2 ₁ 2
Unit cell (Å, °)	74.72, 98.81, 205.68 90 90 90
Number of total reflections	774860 (38734)
Number of unique reflections	109868 (5417)
Multiplicity	7.05 (7.15)
Completeness (%)	100 (100)
I/σ (I)	10.31 (1.5)
Resolution limit	1.96
R _{meas}	0.121 (1.366)
CC _{1/2}	0.859 (0.633)
Refinement	
R factor	0.202
R free	0.241
Number of non-hydrogen atoms:	
Protein	10987
Ligands	108
Water	240
Validation	
RMS Bonds (Å)	0.00089
RMS Angle (°)	1.63
Ramachandran favoured (%)	96.18
Ramachandran Outliers (%)	0
Average B factor (Å ²)	36.75
Rotamer Outliers (%)	4.36

6.7 Fragments Targeting the Substrate Binding Pocket

The first set of fragments was based on three potent PRMT fragment inhibitors that were discovered after deconstruction of a set of previously developed PRMT6 and CARM1 inhibitors. These inhibitors are CMPD-1 and CMPD-2 (CARM1, IC_{50} circa 0.027 μ M) (Sack *et al.*, 2011), and the PRMT6-selective inhibitor EPZ020411 (Mitchell *et al.*, 2015). Their chemical structures are shown in Figure 6-15.

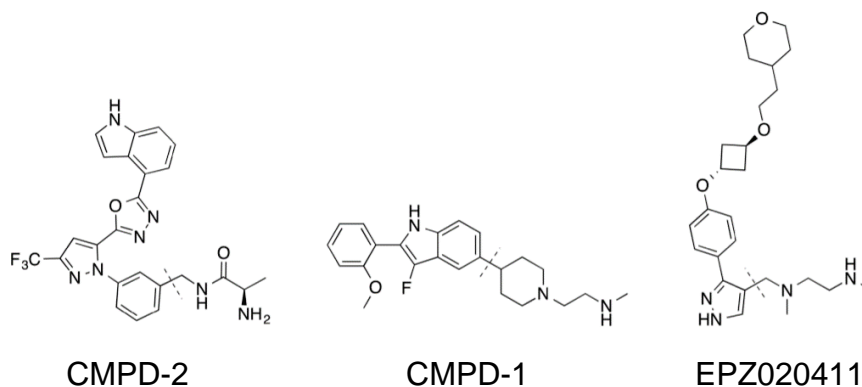


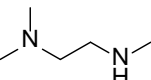
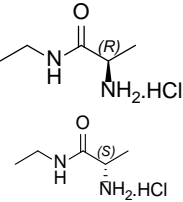
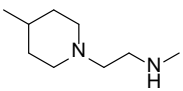
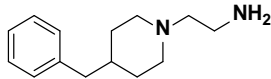
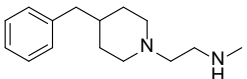
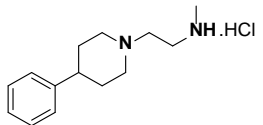
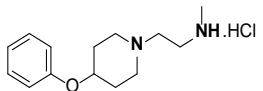
Figure 6-15 PRMT6 and CARM1 Inhibitors Used for Original Fragmentation Study.

Two potent CARM1 inhibitors CMPD-2 and CMPD-1 and one PRMT6 inhibitor EPZ020411 were used for the original fragmentation study conducted by Ferreira *et al.* (2016). The dashed lines indicate the fragmentation points.

The fragments (Table 6-3, compounds 1-4) were shown to have inhibitory activity against full-length CARM1 and PRMT6 and showed excellent ligand efficiency values (Ferreira de Freitas *et al.*, 2016). In fragment-based drug discovery, identified molecules are usually small, which brings many advantages including that the possibilities for fragment growing during optimisation are more diverse by which the chemical space around the fragment can be more efficiently used in contrast to the much larger lead compounds identified in HTS screening (Bembenek *et al.*, 2009). HTS hit compounds often do not have optimal physicochemical characteristic, for example show a poor solubility and instability, which is why the use of fragment-based drug discovery, often combined with structural characterisation of the target-fragment complex, is increasing in recent years (Lipinski, 2004). However, due to the small size and lower number of interactions of the identified fragments, binding affinities are low, but they can still be efficient binders. The ligand efficiency (LE) can be used to rank those fragments according to their potency. LE is defined as the binding free energy for a ligand

divided by its molecular size (Hopkins *et al.*, 2004). The number of non-hydrogen atoms is used to define the fragment molecular size (Kuntz *et al.*, 1999). LE is very useful for fragment optimisation as it allows calculation of the impact on activity of the addition of more molecular bulk (Reynolds *et al.*, 2007). Further SAR studies conducted by Shen *et al.* (2016) on the fragments identified fragment 5 and 6.

Table 6-3 Literature Fragments Targeting PRMT6 and CARM1.

No	Ligand Structure	CARM1		PRMT6		NCL ID
		IC ₅₀ [μM]	LE	IC ₅₀ [μM]	LE	
1		3 ± 0.2	1.08	5 ± 0.6	1.04	
2		105 ± 7	0.68	>200	<0.63	NCL-00024058
3		0.2 ± 0.2	0.83	1 ± 0.3	0.71	
4		1 ± 0.04	-	0.3 ± 0.04	0.56	NCL-00025069
5		0.89 ± 0.14	-	0.17 ± 0.04	-	NCL-00024056
6		6.3 ± 0.28	-	9.34 ± 1.82	-	NCL-00024057
7		0.15 ± 0.02	-	0.09 ± 0.02	-	

Fragments which appeared to give the best inhibitory concentration (IC₅₀) values were ethylenediamine-based compounds. A first set of ligands was produced by Jade Illingworth, including the S-stereoisomer of the literature fragment 2 (NCL-00024058), 4 (NCL-00025069), 5 (NCL-00024056), and 6 (NCL-00024057). In the case of fragment 2, both stereoisomers had been reported in the literature, and it was

not clear which fragment was used for the SAR studies, thus the S-stereoisomer was chosen.

Additionally, compound NCL-00025070, which was based on fragment 3, but had the methyl-group removed was designed by Dr Edwige Picazo. An overview of the structures of the alanine-based (NCL-00024052, NCL-00024053, NCL-00024058, NCL-00024854, NCL-00024058, NCL-00025070) and ethylenediamine-based fragments (NCL-00024055, NCL-00024057, NCL-00024860, NCL-00025069, NCL-00024056) can be seen in the upper part of Figure 6-16.

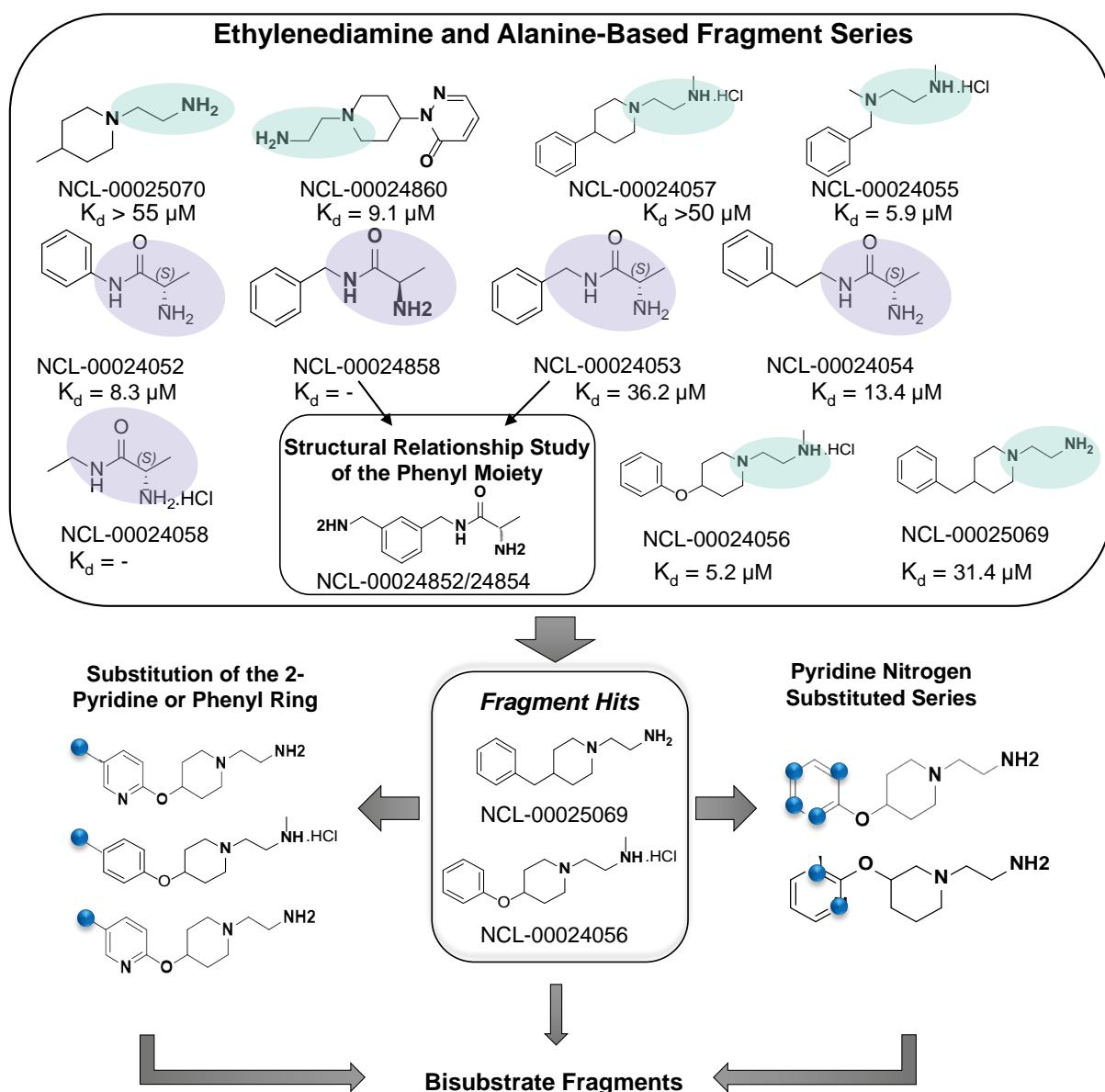


Figure 6-16 Overview of Fragment Design.

In the first step, determination of the crystal structures of these fragments bound to the human CARM1 catalytic domain was attempted via ligand soaking into existing CARM1-SAH crystals (Section 2.22.4). In addition, they were analysed by SPR and in some cases ITC. X-ray crystallography data collection and refinement statistics for all the CARM1-ligand structures can be found in Appendix O.

In the case of NCL-00024058, (Fragment 2, Table 6-3), binding to the CARM1 catalytic domain was not detectable by SPR (Figure 6-17 A), but a complex crystal structure was determined, as shown in Figure 6-17 B. The previously reported half maximal IC₅₀ value for full-length CARM1 was $\geq 100 \mu\text{M}$, which is still surprising for such a small fragment (Ferreira de Freitas *et al.*, 2016). The glutamic acid residue of the E-loop, Glu257 and Met259 form hydrogen bonds with the free amine of the fragment. The carbonyl oxygen of Glu266 forms a hydrogen bond with the secondary amine group. Moreover, the side chain of histidine His414 of the THW-loop participates in hydrogen bonding with the ligand. Figure 6-17 C shows a 2D PoseView Diagram of the ligand interactions (Stierand and Rarey, 2010a). Hydrogen bonds are shown as dashed lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$.

Ligand NCL-00025070 was based on the literature fragment 3, which had shown very good ligand efficiencies for CARM1 and PRMT6 and had an IC₅₀ values of $0.2 \mu\text{M}$ for CARM1 (Ferreira de Freitas *et al.*, 2016). No crystal structure of fragment 3 had been solved, but it was postulated that the high binding affinity and ligand efficiency resulted from buried electrostatic interactions between one glutamic acid residue of the E-loop, Glu257 in CARM1, and the secondary amine group of fragment 3 (Ferreira de Freitas *et al.*, 2016). This interaction had already been observed in the CARM1-CMPD-1 complex structure (PDB: 2Y1W).

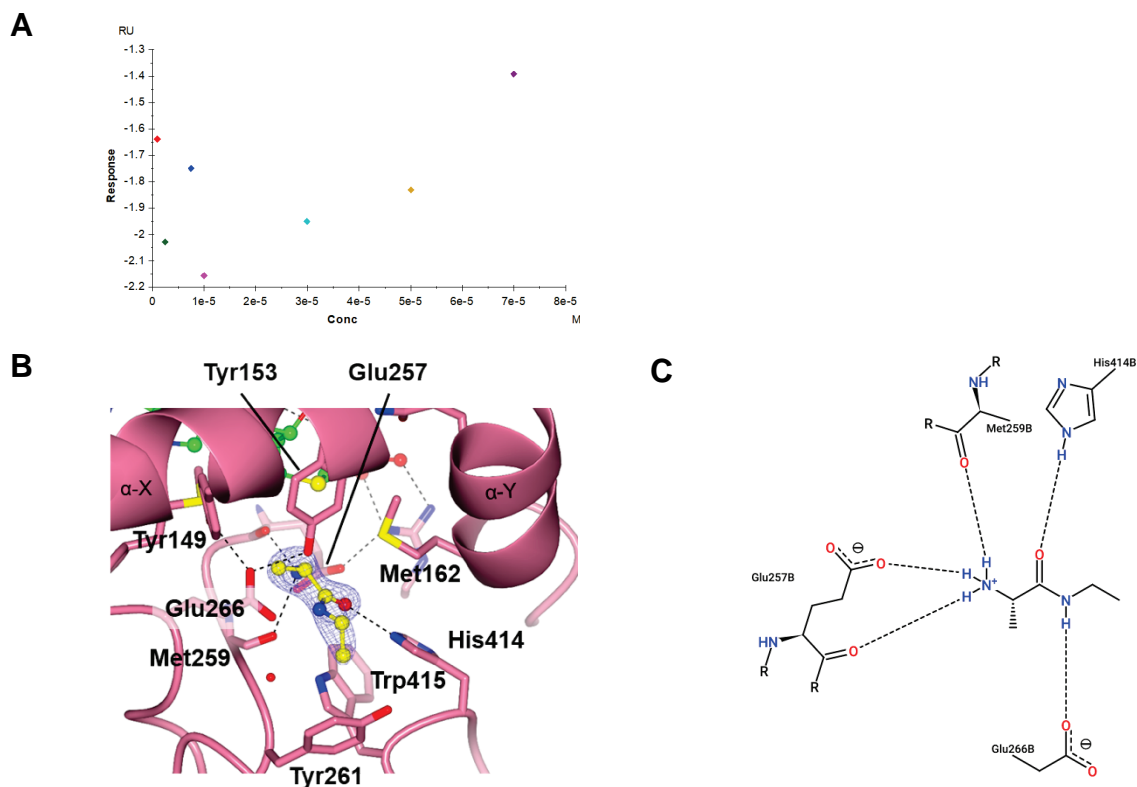


Figure 6-17 Characterisation of NCL-00024058 Binding to CARM1.

A) SPR-Analysis of NCL-00024058. **B)** Crystal Structure of CARM1-NCL-00024058. The $2mF_o-DF_c$ map is shown as a blue chicken wire and contoured around the ligand at 1σ . **C)** 2D PoseView Diagram. Hydrogen bonds are shown as dashed black lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$.

The crystal structure of CARM1-NCL-00025070 was solved (Figure 6-18 B), and fragment binding was also confirmed by SPR, from which a K_d value of $54.8\text{ }\mu\text{M}$ was determined (Figure 6-18 A). The structures of CARM1-NCL-00025070 and CARM1-CMPD-1 (PDB: 2Y1W) are superposed in Figure 6-18 C and show that the fragments bind in an almost identical position. The amine groups form direct hydrogen bonds with His414 and Glu257. The 4-methylpiperidine ring forms hydrophobic interactions with the three amino acid residues His414, Met162 and Trp415.

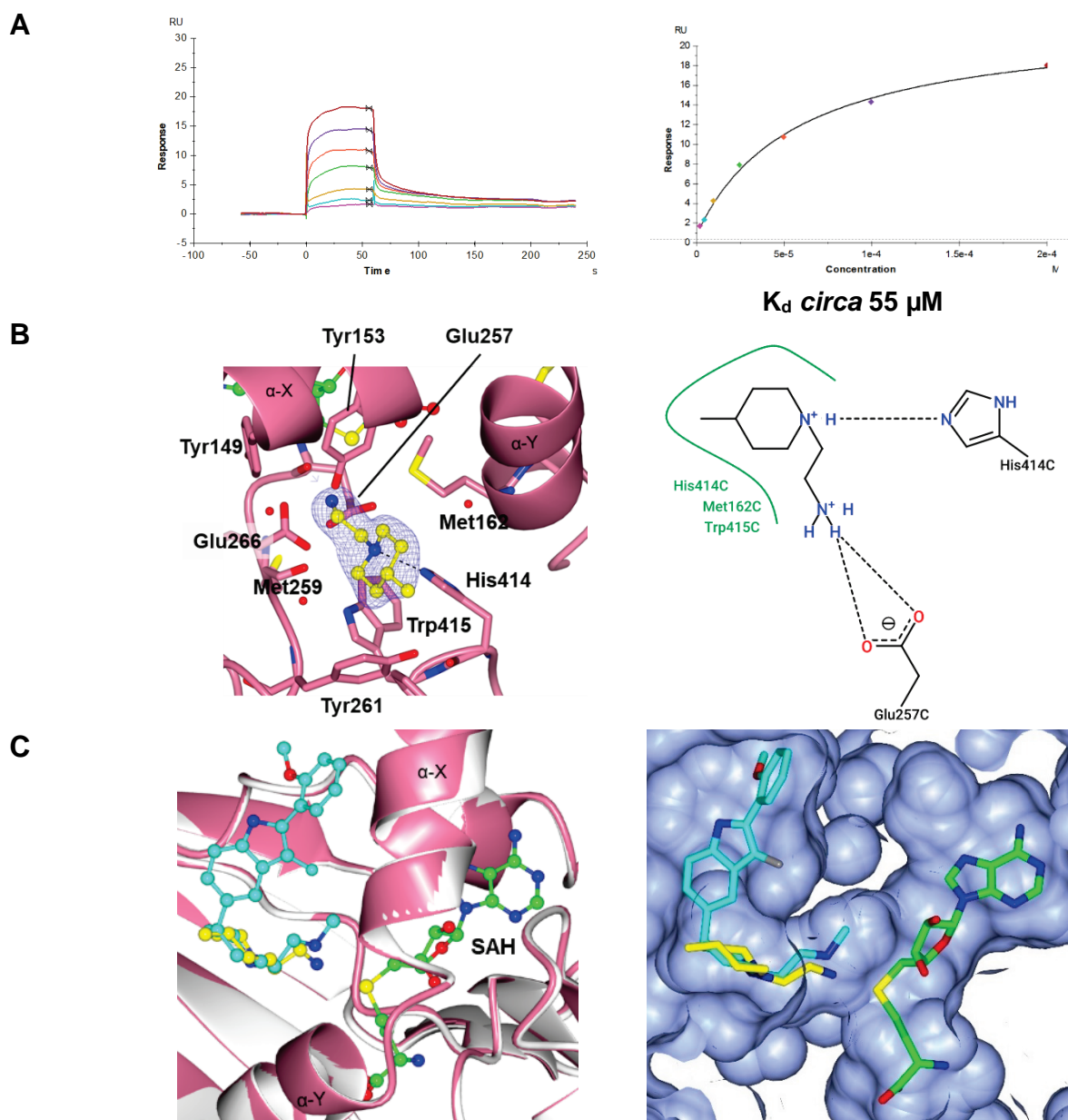


Figure 6-18 Characterisation of the Binding of NCL-00025070 to CARM1.

A) SPR analysis of NCL-00025070 binding. **B)** Crystal Structure of CARM1-NCL-00025070. 2D PoseView Diagram. Green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$. The $2mF_o - DF_c$ map is shown as a blue chicken wire and contoured around the ligands at 1σ . **C)** Superposition of CARM1-NCL-00025070 (purple ribbon, carbon atoms in yellow) with CARM1-CMPD-1 (PDB: 2Y1W, white ribbon and carbon atoms coloured cyan). For clarity, only the SAH molecule of the CARM1-NCL-00025070 is shown.

NCL-00025069 (LE= 0.56) is a fragment that was also identified in the screen against PRMT6 (Ferreira de Freitas *et al.*, 2016). It inhibits CARM1 with an IC_{50} of $1 \mu\text{M}$ and PRMT6 with IC_{50} of $0.3 \mu\text{M}$ and is also active against PRMT8 (IC_{50} circa $2.1 \mu\text{M}$).

Using SPR, a K_d of 31.5 μM was measured (Figure 6-19 A), and the crystal structure bound to CARM1 was determined (Figure 6-19 B).

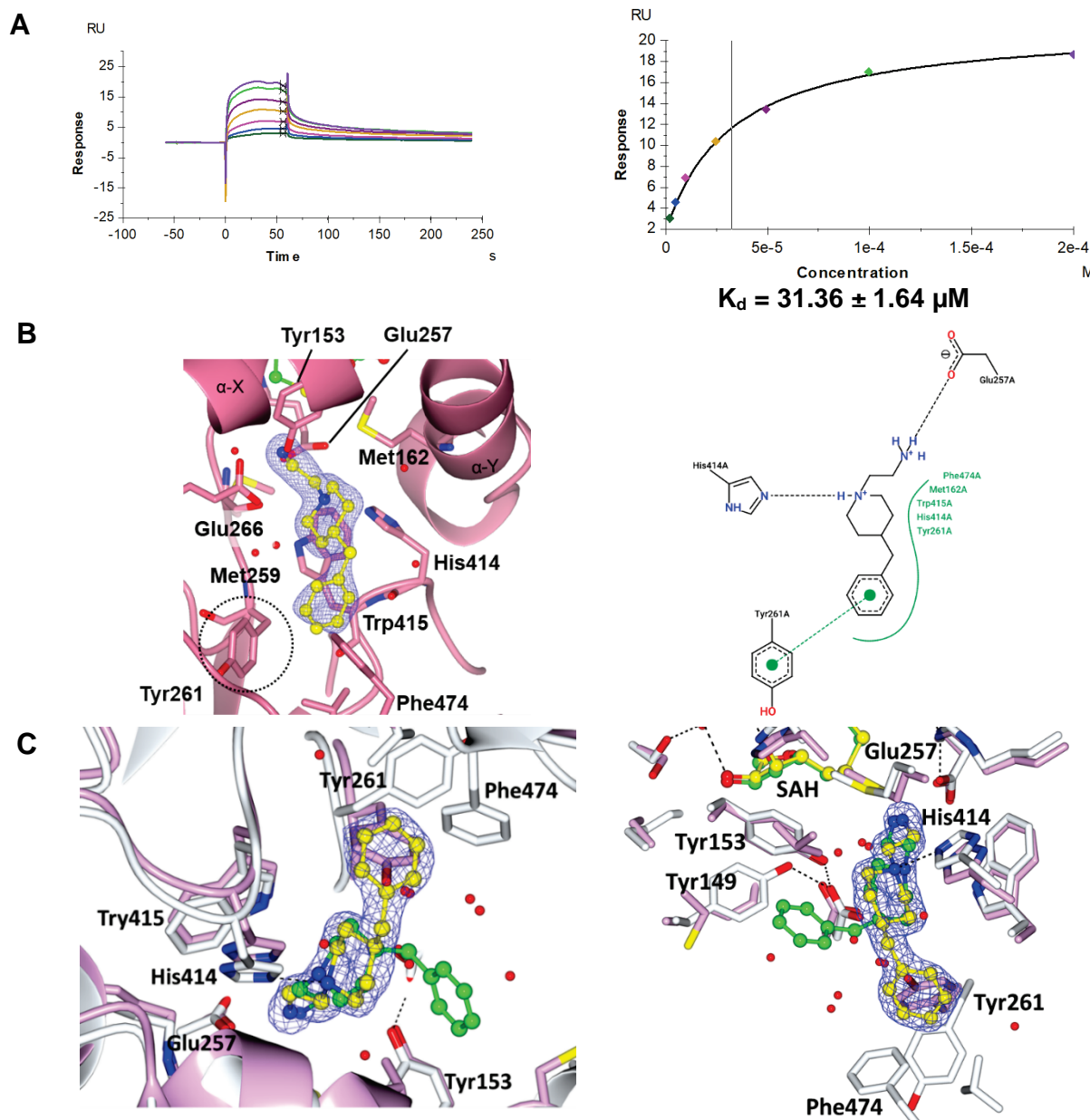


Figure 6-19 Comparison of the NCL-00025069 Binding Site of CARM1 and PRMT6.

A) SPR Analysis of NCL-00025069 The black vertical line in the right-handed plot indicates the value of the calculated K_d . **B)** CARM1-NCL-00025069 structure and 2D PoseView Diagram. Green dotted line: π -stacking; green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor-hydrogen-donor angle must not fall $< 120^\circ$. The $2mF_o - DF_c$ map is shown as a blue chicken wire and contoured around the ligands at 1σ . **C)** Superposition of the crystal structures of CARM1 (white) and PRMT6 (purple, PDB: 5EGS). The aromatic ring of the ligand adopts a different position in CARM1 (yellow) when compared to PRMT6 (green). The $2mF_o - DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured at 1σ .

In this structure, shown superposed with the published PRMT6-ligand structure (PDB: 5EGS) (Ferreira de Freitas *et al.*, 2016), the fragment binds to the arginine binding channel and the alkylamino side chain extends deep into the arginine binding pocket. The amine group forms a hydrogen bond with a glutamate residue (CARM1: Glu257) and the secondary amine forms a hydrogen bond with a histidine residue (CARM1: His414). However, the phenyl ring adopts a different conformation when bound to CARM1 compared to PRMT6. In the CARM1 structure, a tyrosine ring (CARM1: Tyr261) flips out of the pocket to accommodate the ligand, which allows π -stacking with the aromatic ring of the ligand, as visible in Figure 6-19 C.

In the PRMT6-bound structure, a second lysine rich binding pocket next to the arginine binding site was partially occupied by the ligand (Ferreira de Freitas *et al.*, 2016). This site is not observed in the CARM1-complex structure. The differences in the ligand binding mode may explain why the fragment is a less potent inhibitor of CARM1 compared to PRMT6. Figure 6-20 shows an overlay of the PRMT6 and CARM1 structures bound to NCL-00025069, and the CARM1 structure bound to CMPD-1 (PDB: 2Y1W). The alkylamino sidechain aligns well in all structures but the aromatic rings are in different positions. Moreover, the positions of the NCL-00025069 phenyl ring bound to CARM1 (yellow) and to PRMT6 (green) differ and reach into opposite located hydrophobic pockets.

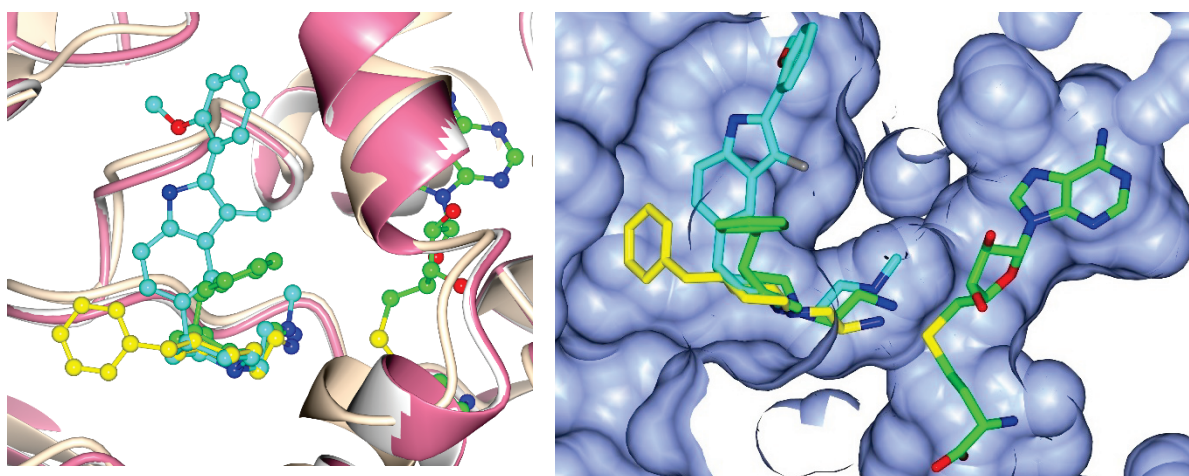


Figure 6-20 PRMT6 and CARM1 Superposition with NCL-00025069 and CMPD-1.

CARM1 in complex with NCL-00025069 (pink ribbon, carbon atoms in yellow) was superposed with the published PRMT6- NCL-00025069 structure (cream ribbon, carbon atoms in green), and CARM1 in complex with CMPD-1 (white ribbon, carbon atoms in cyan). For clarity, only the SAH molecule of the CARM1-NCL-00025069 structure is shown (CARM1 PDB code: 2Y1W, PRMT6: 5EGS).

Next, the literature ligand NCL-00024056 was tested, in which the terminal amino group of NCL-00025069 was substituted with a methyl-group. SPR analysis of the fragment (Figure 6-21 A), showed that it binds to CARM1 with 10× increase in affinity compared to NCL-00025069 ($K_d = 54.8 \mu\text{M}$), with a K_d value of $5.18 \mu\text{M}$. The reported literature IC_{50} for CARM1 is $0.8 \pm 0.14 \mu\text{M}$ (Ferreira de Freitas et al., 2016).

To analyse the binding mode, which might explain the large affinity differences between the two fragments, the crystal structure in complex with CARM1 was solved (Figure 6-21 B). The aromatic ring and the secondary amine form the same hydrogen bond and π -stacking interactions seen for NCL-00025069. Superposition of the CARM1 crystal structure with NCL-00025069 and NCL-00024056 (Figure 6-21 C) confirmed that the aromatic ring and the piperidine ring align almost perfectly. However, the alkylamino side chain adopts a different conformation, the overall RMSD value of the ligands is 1.04 \AA . However, the ligands form the same interactions, in both cases the amine group of the alkylamino side chain forms a hydrogen bond with Glu257.

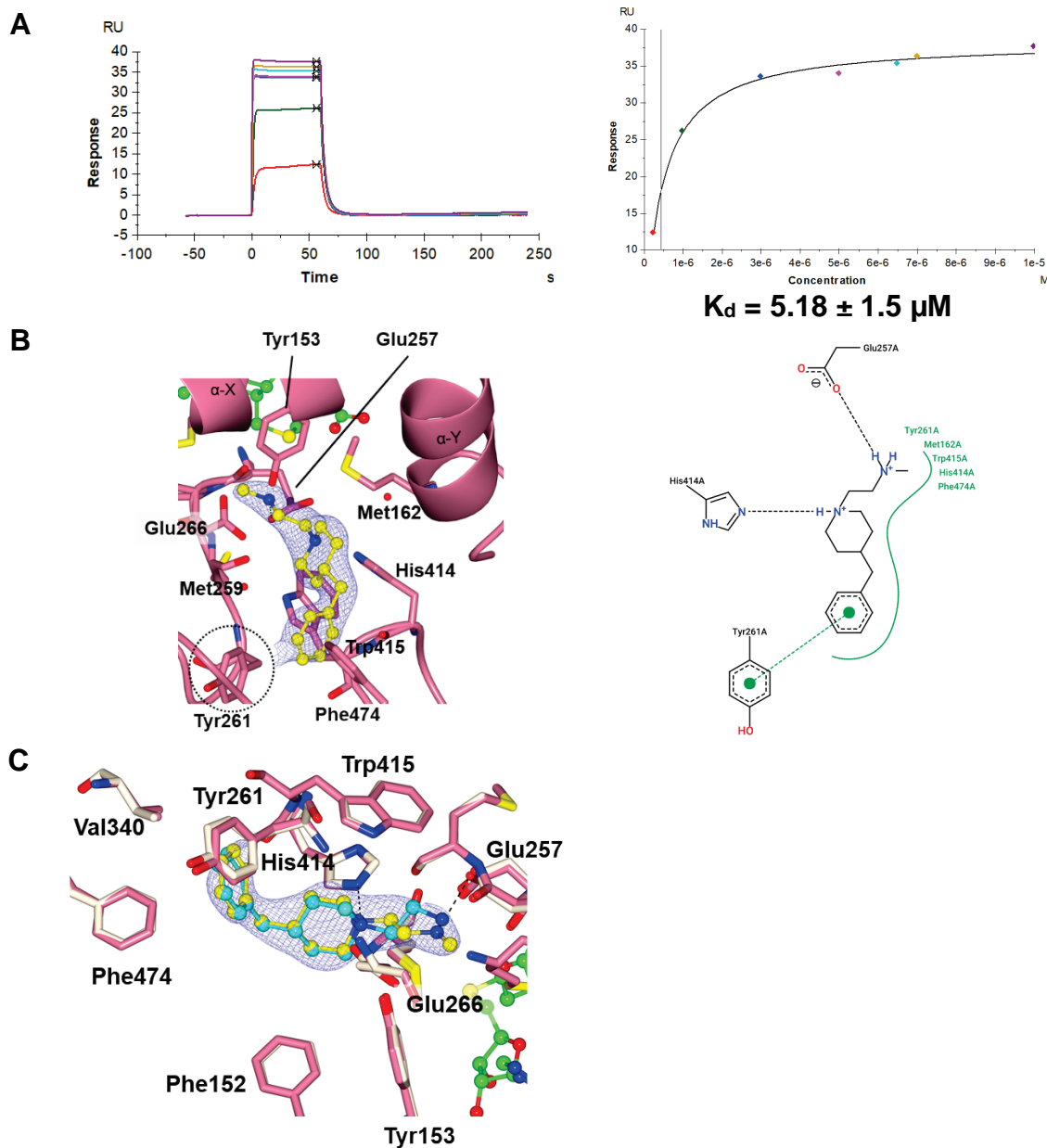


Figure 6-21 Binding Analysis of NCL-00024056 to CARM1.

A) SPR Analysis of NCL-00024056 binding to CARM1. The black vertical line in the right-handed plot indicates the value of the calculated K_d . **B)** Crystal structure of a CARM1- NCL-00024056 complex. The $2mF_o-DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured to 1σ . 2D PoseView Diagram. Green dotted line: π -stacking; green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor-hydrogen-donor angle must not fall $< 120^\circ$. **C)** Superposition of CARM1- NCL-00024056 (yellow) and CARM1- NCL-00025069 (cyan). The $2mF_o-DF_c$ map is shown for NCL-00025069 and contoured around the ligand at 1σ .

Binding was also analysed by ITC (Figure 6-22). A K_d of $2.59 \pm 8.47 \mu\text{M}$ was calculated, which is in good agreement with the SPR results.

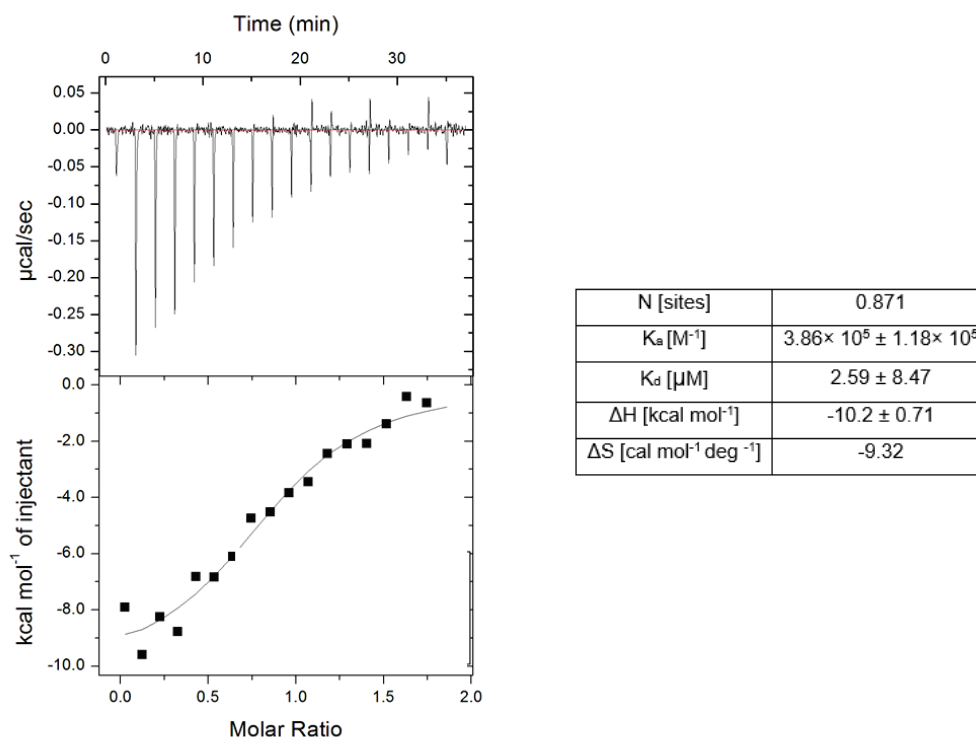


Figure 6-22 ITC Analysis of NCL-00024056 Binding to HsCARM1^{CAT}.

ITC experiments were performed 20 °C with 20 μM CARM1 (pre-incubated with 200 μM SAH) in the cell and 200 μM ligand in the syringe. Experiments were carried out once and analysed using a one set of sites model. A K_d of $2.59 \pm 8.47 \mu\text{M}$ was determined.

A crystal structure of NCL-00024057 bound to CARM1 could not be determined, which is surprising because the IC_{50} value of the fragment reported in the literature is $6.3 \pm 0.28 \mu\text{M}$. SPR analysis was performed and showed a low binding affinity to CARM1 (Figure 6-23). The ligand bound to CARM1 with an estimated K_d of *circa* 50 μM and did not reach saturation. One reason for the low binding could be the fact that all SPR and crystal experiments were performed with the catalytic domain of CARM1 and not the full-length protein, which was used in the study conducted by Shen *et al.* (2016). However, the SPR response is odd as it is already near the maximal theoretical RU of 45. To reconfirm ligand binding, ITC analysis was performed (Figure 6-23 B). However, binding could not be detected.

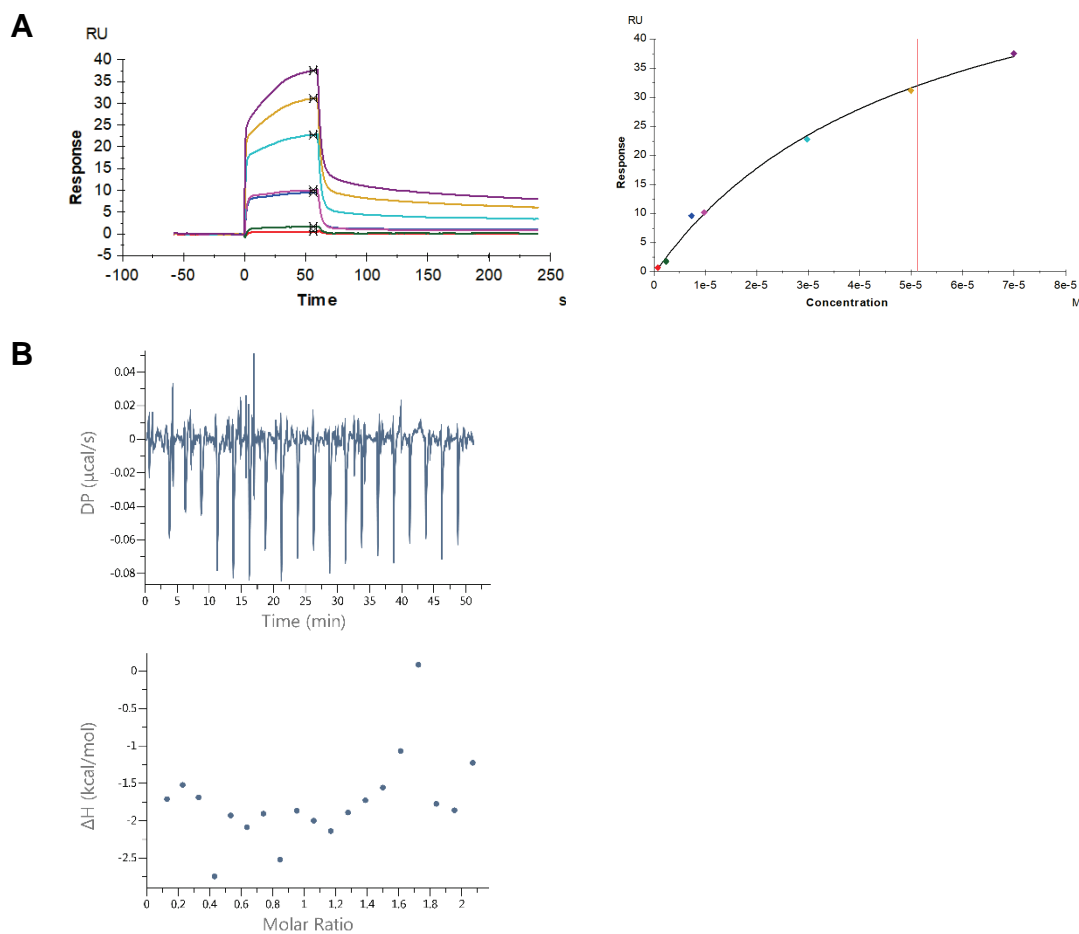


Figure 6-23 SPR and ITC Analysis of NCL-00024057 Binding to CARM1.

A) SPR analysis of NCL-00024057 binding to CARM1. Experiments were performed at 20 °C using 0-100 μM of ligand with SAH present in the SPR buffer. The red vertical line indicates that the reported K_d should not be trusted as it is higher than half the highest analyte concentration. **B)** ITC titration experiment of NCL-00024057 against CARM1. ITC experiments were performed at 20 °C with 20 μM CARM1 (pre-incubated with 200 μM SAH) in the cell and 200 μM ligand in the syringe.

Moreover, fragment NCL-00024860 was designed. SPR analysis showed that it binds to CARM1 with a K_d value of $9.11 \pm 4.76 \mu\text{M}$ (Figure 6-24). Unfortunately, a CARM1-ligand structure could not be determined.

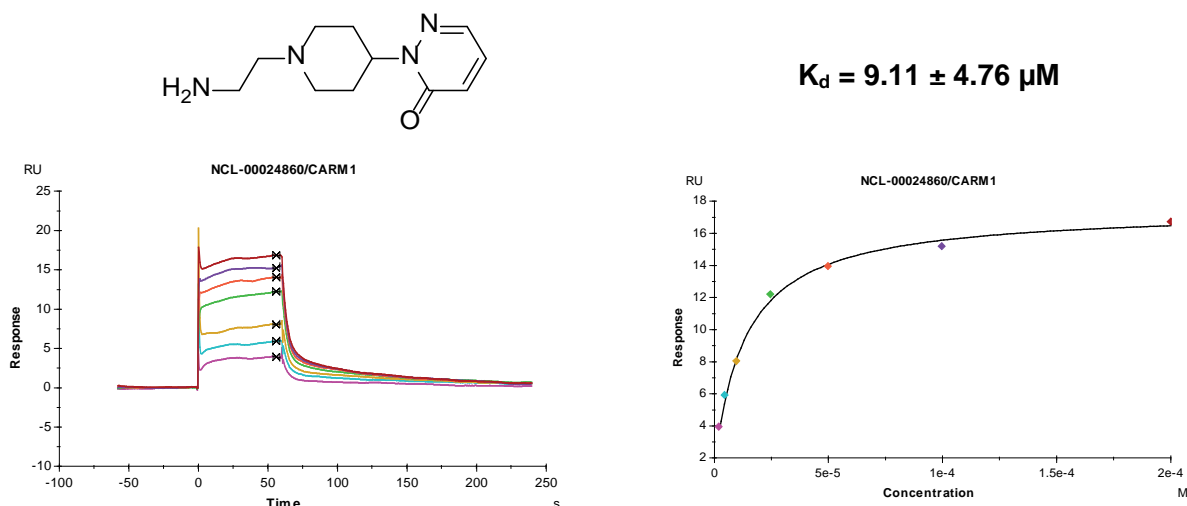


Figure 6-24 SPR Analysis of NCL-00024860 Binding to CARM1.

In the next step, the other newly designed alanine- or ethylenediamine-based fragments were screened by SPR. The ligand structures and SPR results can be seen in Table 6-4 and Figure 6-25.

Table 6-4 Newly Designed Alanine- and Ethylenediamine-based Fragments.

NCL ID	NCL-00024052	NCL-00024053	NCL-00024054	NCL-00024055
Structure				
K_d [μM]	8.26 ± 0.92	36.21 ± 1.89	13.37 ± 0.76	5.91 ± 1.57

Different lengths for the middle linker between the phenyl ring and the secondary amine group were tested (NCL-00024052/53/54). NCL-00024052 and NCL-00024054 had similar binding affinities, 8.26 and 13.37 μM . NCL-00024053 showed the weakest binding with a K_d of 36.21 μM . In addition, the alanine-based ligand NCL-00024055 was analysed. NCL-00024052 and NCL-00024055 showed very similar binding affinities to CARM1 with measured K_d s of 8.26 and 5.91 μM , respectively.

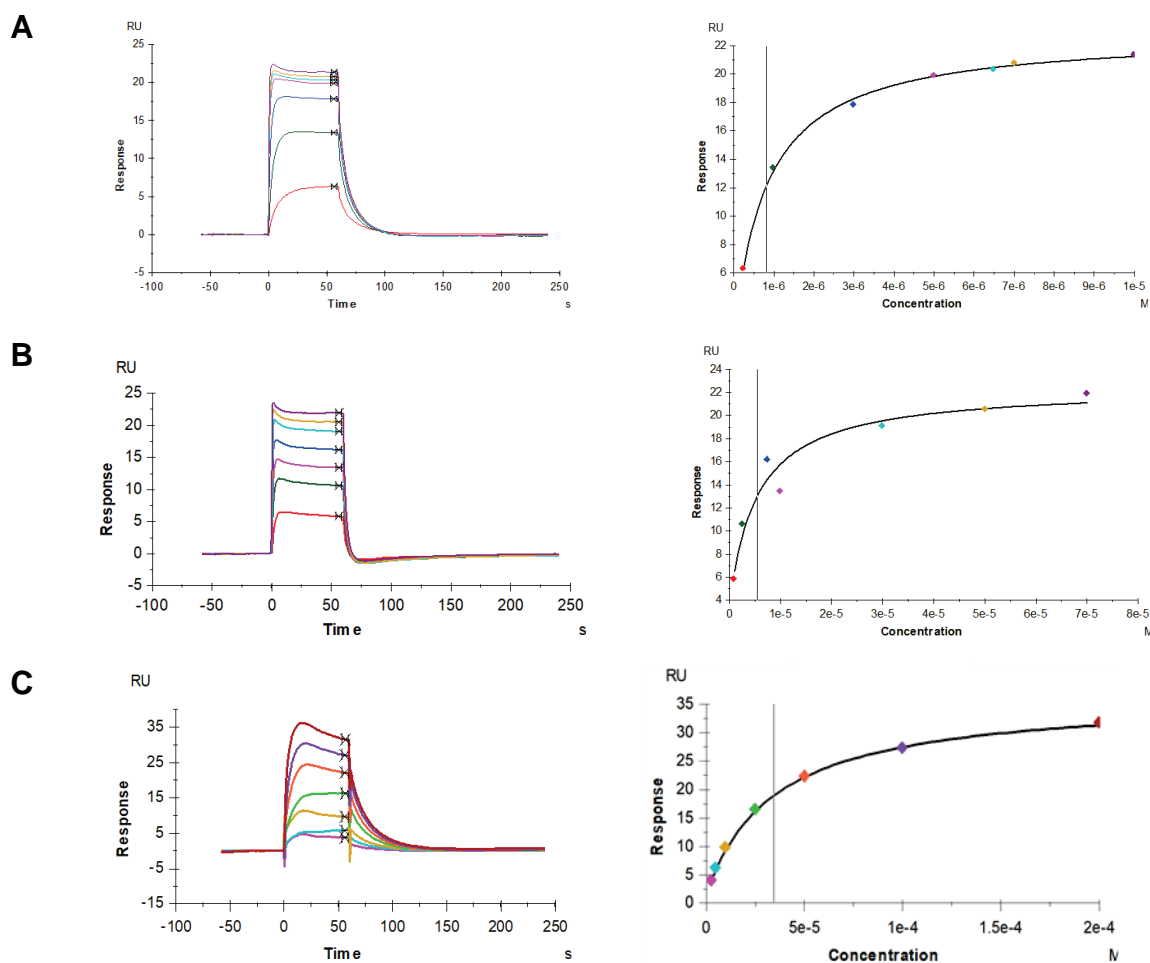


Figure 6-25 Analysis of Fragment Binding to CARM1 by SPR.

SPR analysis to measure the binding of **A)** NCL-00024052; **B)** NCI-00024053 and **C)** NCL-00024054 to CARM1. The black vertical line in the right-handed plot indicates the value of the calculated K_d .

Crystal structures were determined for all three ligands bound to CARM1 (Figure 6-26). As previously observed, the alkylamino chain forms hydrogen bonds with Glu257 and His414 and all three ligands. However, in the case of NCL-00024052 an additional hydrogen bond is formed between the free amine group and a methionine residue (Met259), which could explain the higher binding affinity compared to NCL-00024053. The phenyl ring sits deep in a hydrophobic pocket. In NCL-00024052 and NCL-00024053 the secondary amine can form a hydrogen bond with a tyrosine residue of the structural motif IV (Tyr153). In contrast, in NCL-00024054 a hydrogen bond is formed with Glu266 and its phenyl ring sits deep in a hydrophobic pocket.

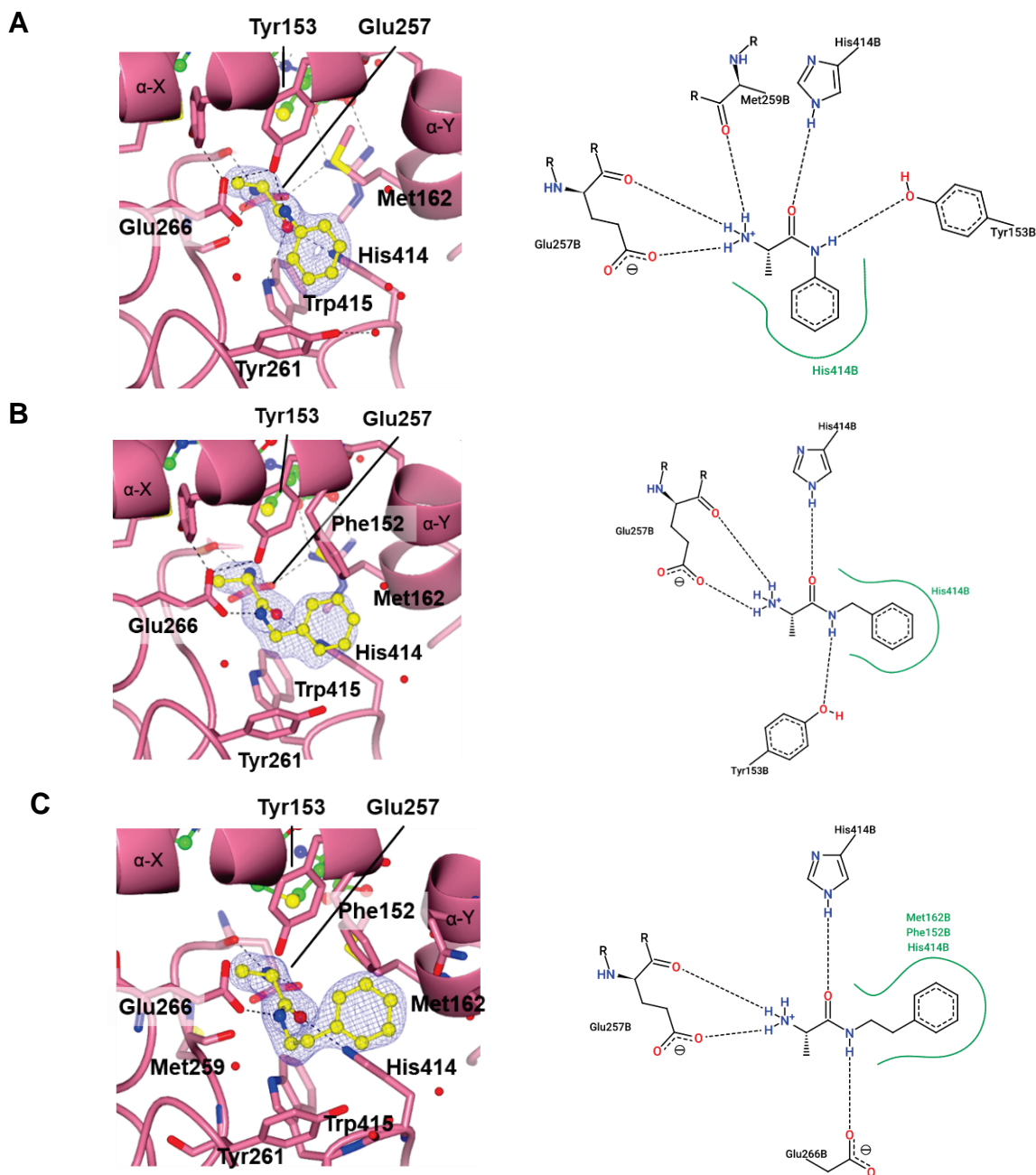


Figure 6-26 Characterisation of the Binding of Ethylenediamine-Based Fragments to CARM1 by X-Ray Crystallography.

Crystal structures of CARM1 in complex with NCL-00024052 (A), NCL-00024053 (B), and NCL-00024054 (C). 2D PoseView Diagram. Green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$.

SPR analysis of NCL-00024055 gave a K_d value of $5.91 \pm 1.57 \text{ \mu M}$, very similar to NCL-00024052. The SPR analysis and crystal structure can be seen in Figure 6-27. The secondary amine forms a hydrogen bond with Met259 and not with Glu257 as seen in the other ethylenediamine-based ligand structures, the tertiary amine with

His414. The aromatic ring is located in a hydrophobic pocket, formed by His414, Phe152 and Met162.

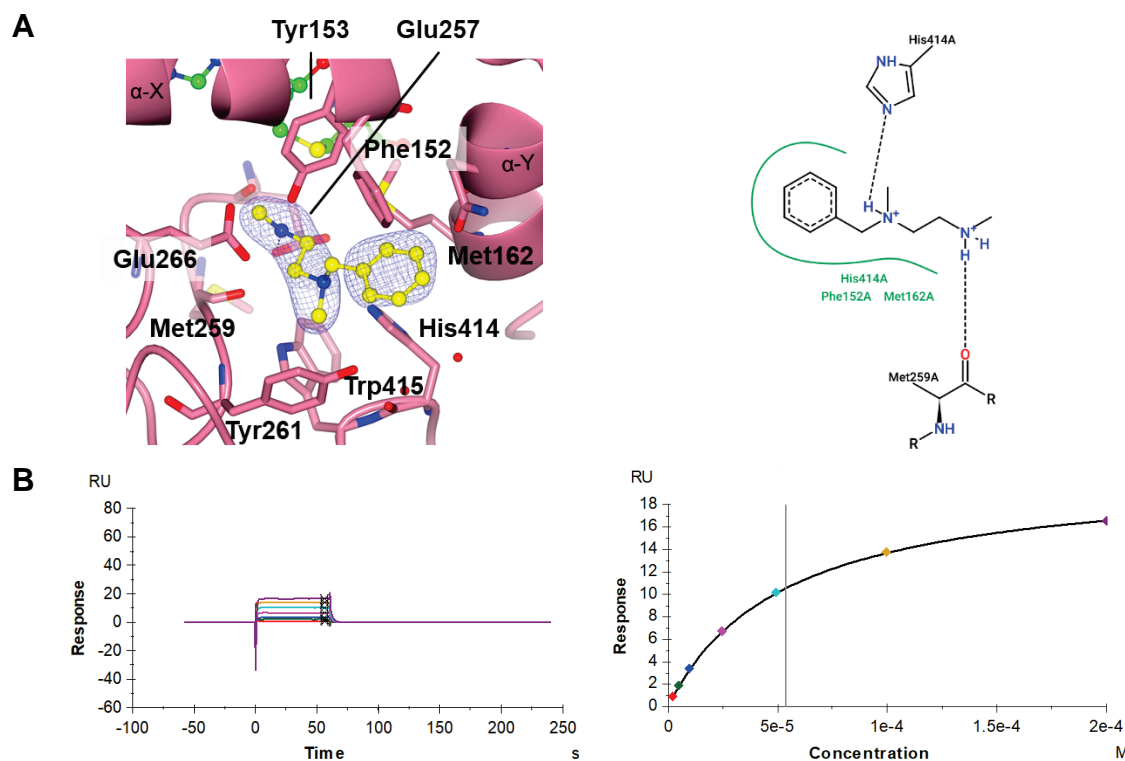


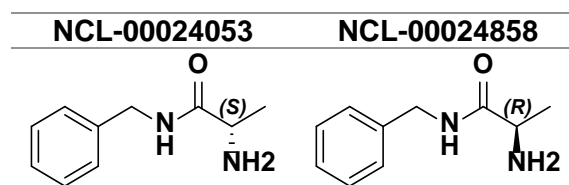
Figure 6-27 Characterisation of the Alanine-Based Ligand NCL-00024055.

A) CARM1-NCL-00024055 crystal structure. The $2mF_o-DF_c$ map around the CARM1 ligands is shown as blue chicken wire and contoured at 1σ . 2D PoseView Diagram. Green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$. **B)** SPR-analysis of NCL-00024055 binding to CARM1. A K_d value of $5.91 \pm 1.57 \mu\text{M}$ was determined. The black vertical line in the right-handed plot indicates the value of the calculated K_d .

6.7.1 Stereochemical Study of the Secondary Amine of the Ethylenediamine

NCL-00024858 was designed to test whether the stereochemistry of the secondary amine group of ligand NCL-00024053, which has an S configuration that has already been described in the last section, influences binding affinity.

Table 6-5 Chemical Structure of the Stereoisomers.



Binding of the R-stereoisomer to CARM1 was observed by SPR. However, it did not reach saturation and thus no K_d value could be calculated (Figure 6-28 A). The crystal structure of CARM1 bound to NCL-00024858 was determined (Figure 6-28 B), the overlay of the S- and R-stereoisomer is shown in Figure 6-28 C. Due to the different stereochemistry they adopt a slightly different orientation. However, the space can accommodate both stereoisomers. The free amine group and the oxygen form the same hydrogen bond interactions. However, the secondary amine group in NCL-00024858 forms hydrogen bond with Glu266, in the case of NCL-00024053 it is formed with Try153 (Figure 6-28 B).

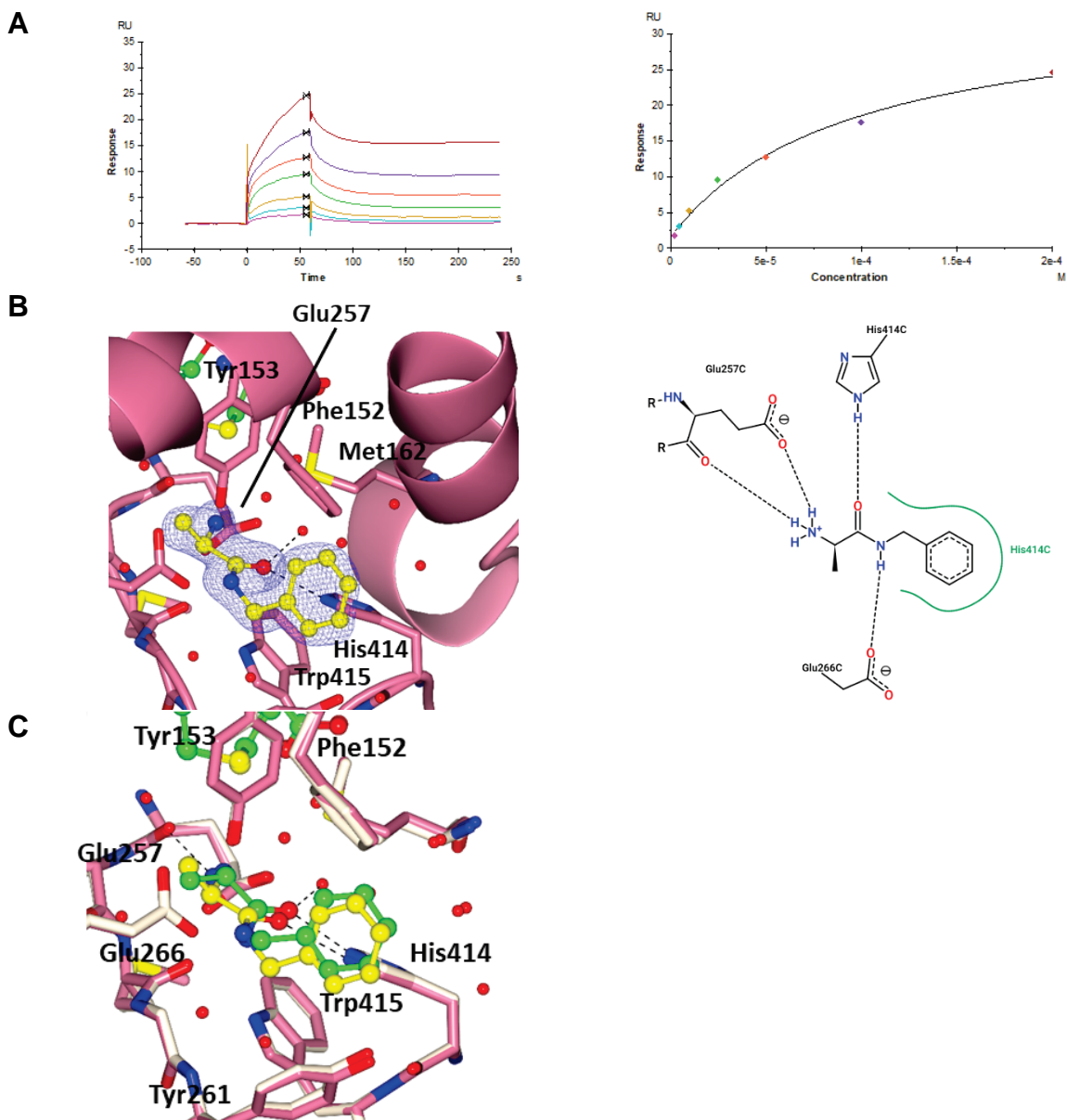


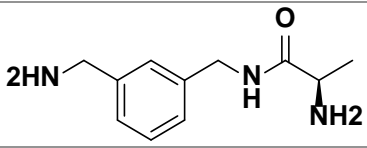
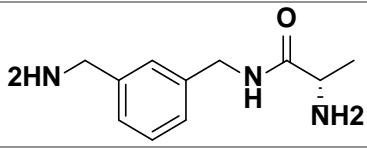
Figure 6-28 Comparison of the Binding of NCL-00024053 and NCL-00024858 to CARM1.

A) SPR analysis of NCL-00024053 binding to CARM1. **B)** CARM1 crystal structure in complex with NCL-00024858. The $2mF_o-DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured at 1σ . 2D PoseView Diagram. Green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$. **C)** Superposition of CARM1-NCL-00024053 (cream ribbon, carbon atoms in green) and CARM1-NCL-00024858 (pink ribbon, carbon atoms in yellow).

6.8 Structural Relationship Study of the Phenyl Moiety

Fragment NCL-00024852 and NCL-00024854 were designed based on NCL-00024858 and NCL-00024053 with the aim to test whether binding affinities can be increased by substitution of the Phenyl moiety with an amine group. The chemical structures are shown in Table 6-6. SPR analysis showed that the R-stereoisomer binds to the catalytic domain of CARM1 with a K_d of $3.21 \pm 1.43 \mu\text{M}$, and the S-stereoisomer with a K_d of $24 \pm 2.83 \mu\text{M}$ (Figure 6-29 A).

Table 6-6 Chemical Structure and Affinity Data of the Stereoisomers.

	NCL-00024852 (R)	NCL-00024854 (S)
		
K_d [μM]	3.21 ± 1.43	24 ± 2.83

The crystal structure of the R-stereoisomer could not be determined, but the crystal structure of CARM1 in complex with NCL-00024854 (S) was solved (Figure 6-29 B). The oxygen atom and the free amine group form the same hydrogen bonds observed in the CARM1-NCL-00024052 structure, and the phenyl ring is located in a similar position and forms hydrophobic interactions with His414. However, the secondary amine forms a hydrogen bond with Glu260, in contrast to NCL-00024052, where a hydrogen bond to Tyr153 was observed. The added free amine group forms hydrogen bond interactions with Tyr261. The additional observed interactions do not agree with the observed 10x loss in binding affinity compared to NCL-00024052.

From the first set of ethylenediamine-based fragments NCL-00024055, NCL-00024056 and NCL-00024060 showed the best binding affinities. Unfortunately, no crystal structure of NCL-00024060 was obtained, which would be needed to analyse the binding mode. Additionally, the alanine-based fragment NCL-00024052 showed good binding affinities.

The most promising ethylenediamine-based fragment that was successfully crystallised in complex with CARM1 was NCL-00024056 ($K_d = 5.2 \mu\text{M}$) values. In the next step two different fragment series were designed with the aim to increase binding

affinities of these identified fragments. The results will be discussed in the following section.

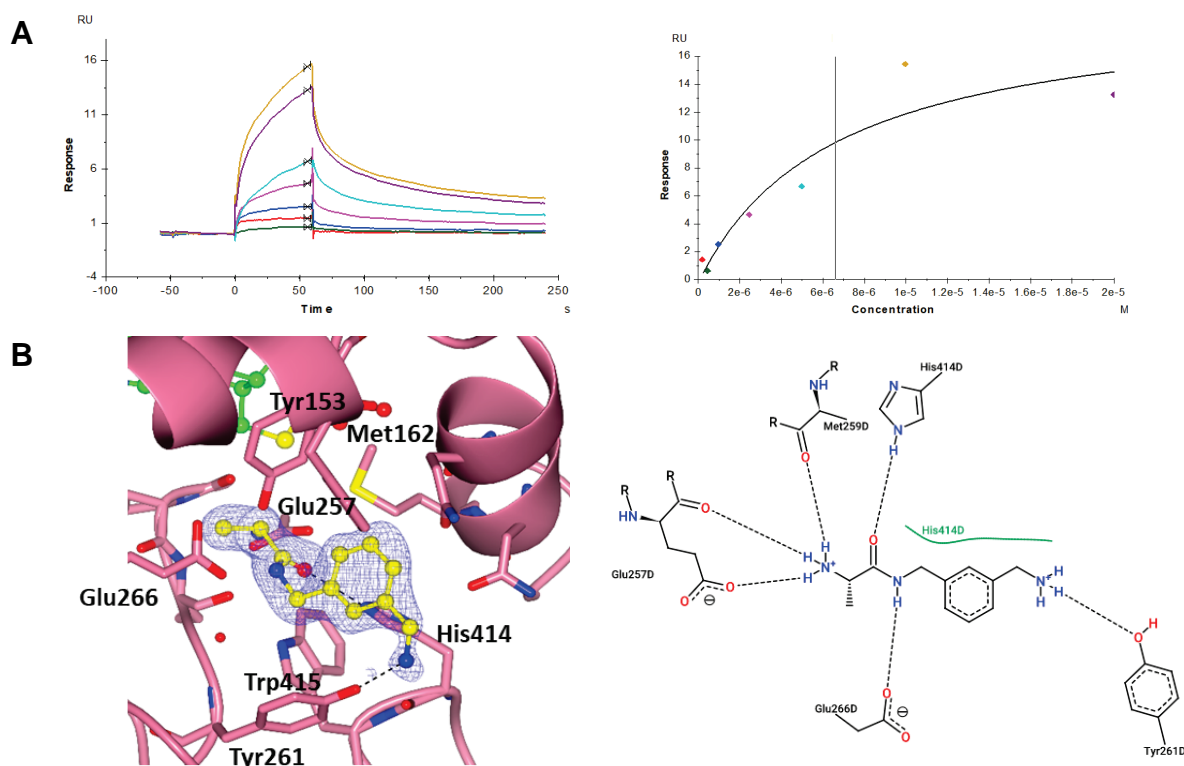


Figure 6-29 Characterisation of the Binding NCL-00024852 and NCL-00024854 to CARM1.

A) SPR analysis of NCL-00024852 binding to CARM1. The black vertical line in the right-handed plot indicates the value of the calculated K_d . **B)** CARM1 crystal structure in complex with NCL-00024854. The $2mF_o-DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured at 1σ . 2D PoseView Diagram. Green dotted line: π -stacking; green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$.

6.8.1 Characterisation of the Pyridine Nitrogen Substituted Series

A pyridine nitrogen substituted series and a substitution series of the 2-pyridine and phenyl ring were prepared, as shown in Figure 6-30.

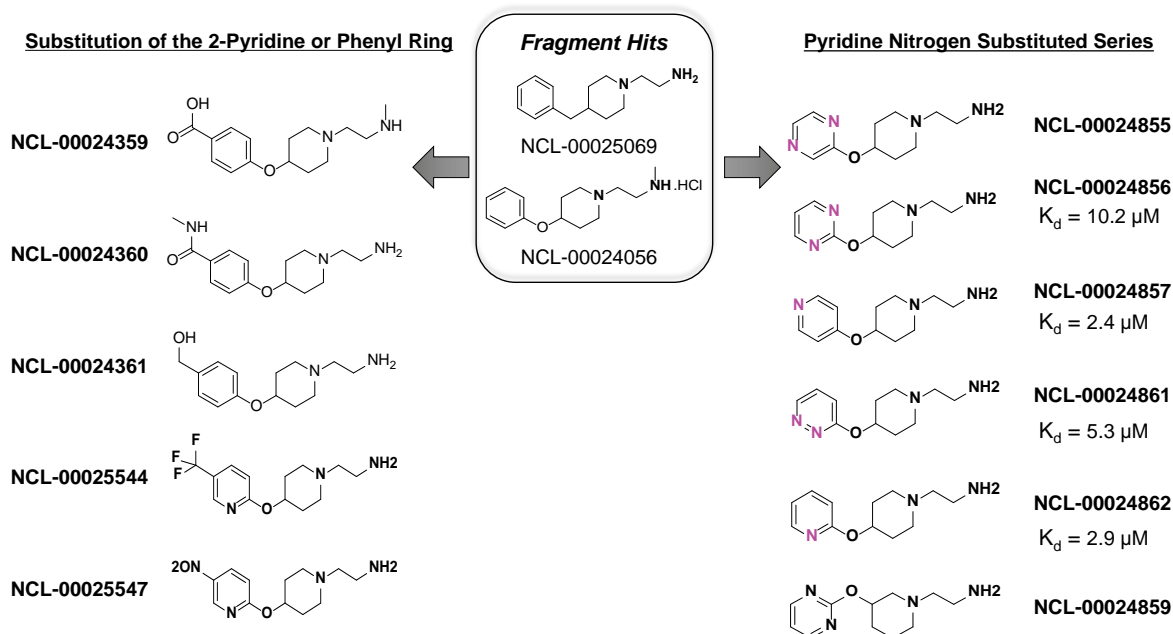


Figure 6-30 Structures of the Designed Fragments.

A SAR study of the middle linker of fragment NCL-00024056, performed by Shen *et al.* (2016), identified fragment 7 (Table 6-3) as the compound in the set with highest affinity for CARM1. This fragment has an O linker, which increased the IC_{50} for CARM1 slightly from 0.89 to 0.15 μM . Based on these results, a series of fragments substituted with a pyridine nitrogen was designed based on ligand NCL-000250069 and NCL-00024056, including a middle O-linker, that only differ in the position and number of nitrogen atoms substituted into the pyridine ring. Their binding to CARM1 was characterised by SPR (Figure 6-31). NCL-00024857 and NCL-00024862 showed the highest binding affinity with K_d values of 2.4 μM and 2.9 μM , respectively. Both are pyridines with the ring nitrogen at position 2 or 4, respectively. Changing the ring nitrogen position did not affect affinity. A second nitrogen atom was added to the ring, resulting in a 3-pyridazinyl-containing compound (NCL-00024861). A K_d value of 5.3 μM was measured by SPR. Thus, addition of the second nitrogen at the 3-position caused a slight increase in affinity compared to NCL-00024862. A 2-pyrimidinyl-group was not well tolerated (NCL-00024856) and resulted in a decreased binding affinity (K_d circa

10.2 μM). NCL-00024855, which has a 2-pyrazinyl group showed drastically reduced binding to CARM1, such that a K_d value could not be determined.

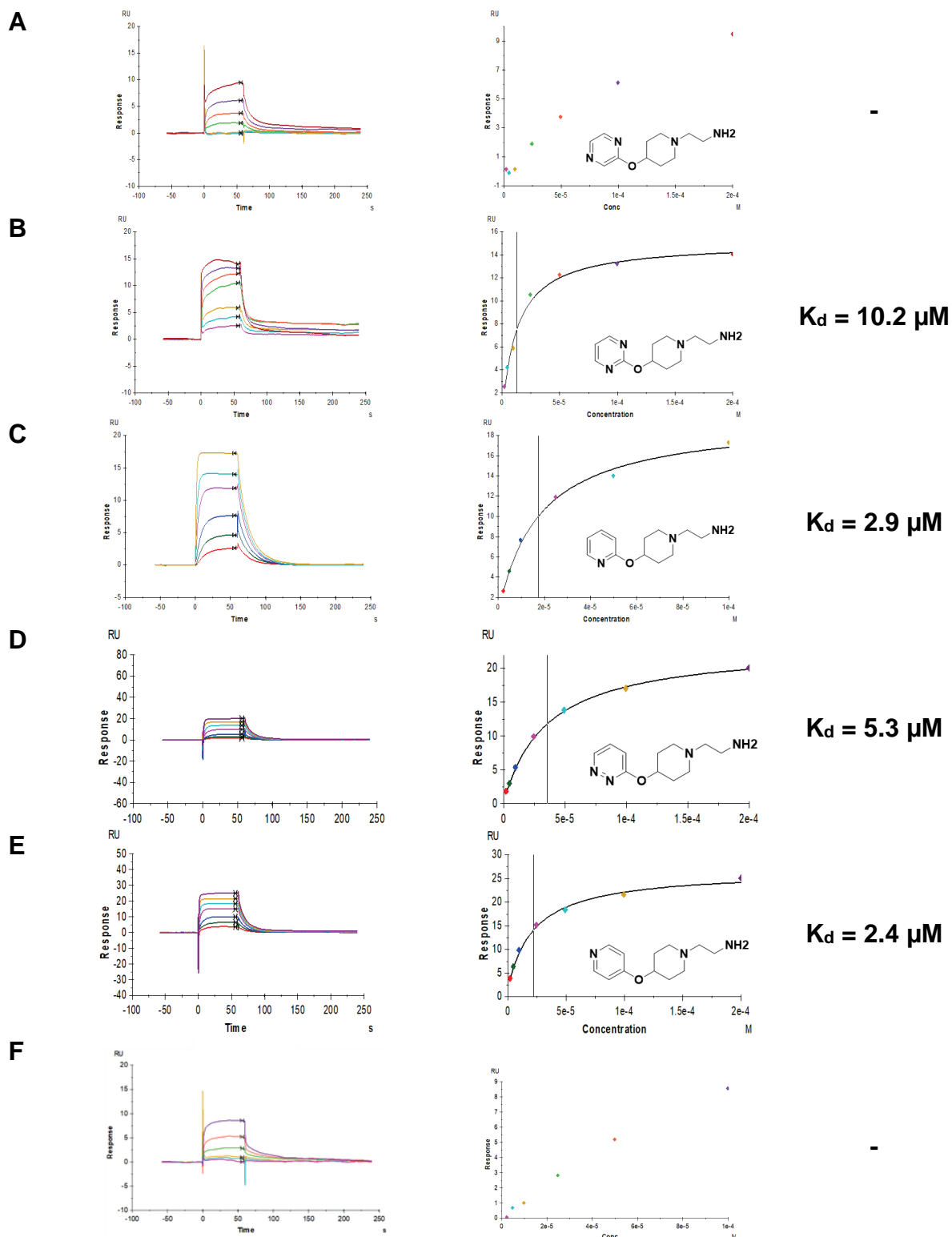


Figure 6-31 SPR Binding Analysis of Pyridine Nitrogen Substituted Series.

A) NCL-00024855 B) NCL-00024856 C) NCL-00024862 D) NCL-00024861 E) NCL-00024857 F) NCL-00024859. The black vertical line in the right-handed plot indicates the value of the calculated K_d .

The crystal structures of several CARM1-ligand complexes were solved. Figure 6-32 shows the crystal structures of NCL-00024855 and NCL-00024856.

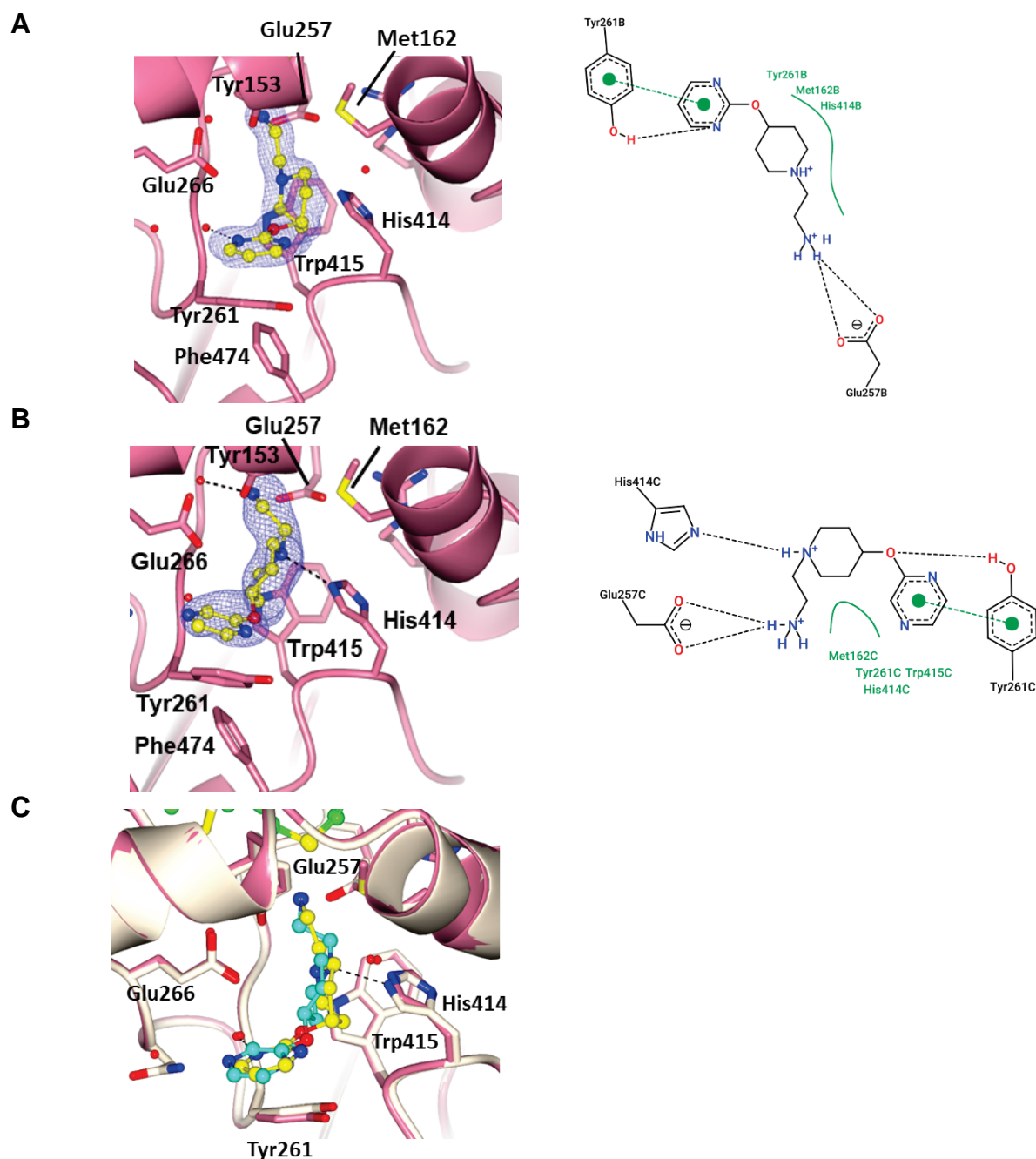


Figure 6-32 Characterisation of the Binding of NCL-00024856 and NCL-00024855 to CARM1.

A) Binding of NCL-00024856 to CARM1. **B)** Binding of NCL-00024855 to CARM1. **C)** Superposition of CARM1-NCL-00024855 (pink ribbon; carbon atoms in yellow) and NCL-00024855 (cream ribbon; carbon atoms in cyan). The $2mF_o-DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured at 1σ . 2D PoseView Diagram. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9 \text{ \AA} \pm 0.5 \text{ \AA}$ and the condition that the acceptor-hydrogen-donor angle must not fall $< 120^\circ$. Green dotted line: π -stacking, green smooth contour line: hydrophobic interactions.

In both structures, the piperidinyl-ethylamine moiety forms hydrogen bonds with Glu257 of the double E-loop and π -stacking interactions with Tyr261. However, only in NCL-00024855 a hydrogen bond is also formed between His414 of the THW-loop and the ring nitrogen atom. The middle linkers of both structures form hydrophobic interactions with adjacent residues. Additionally, in NCL-00024855, the linker oxygen forms a hydrogen bond with the hydroxyl-group of Tyr261.

Figure 6-33 shows the crystal structures and an overlay of NCL-00024861 and NCL-00024862. In both structures, the secondary amine group of the pyridine ring forms hydrogen bonds with His414 and the free amine group with Glu257. However, in NCL-00024861 an additional H-bond with Met259 is observed. The 3-pyridazinyl-ring forms π -stacking interactions with Tyr261 as seen in NCL-00024855 and NCL-00024856. In contrast, in NCL-00024862 the oxygen atom of the linker forms a hydrogen bond with Tyr153 and the 2-pyridine ring forms hydrophobic interactions with Tyr261. In the case of NCL-00024862 a large unassigned density in proximity to the pyridine ring was visible in the electron density map (Figure 6-33 C). Despite extensive efforts, additional ligand molecules or crystallisation buffer components such as PEG3350 could not be fitted into it.

An SAR study conducted by Shen *et al.* showed that switching the 4-benzyl group of NCL-00024056 at the piperidinyl ring to the 3-position increased potency significantly (Shen *et al.*, 2016). Thus, NCL-00024859 was additionally designed. Binding by SPR was detectable (Figure 6-31 F), but a K_d value could not be determined because saturation was not reached. A crystal structure could not be determined, most likely due to the low binding affinity of the ligand.

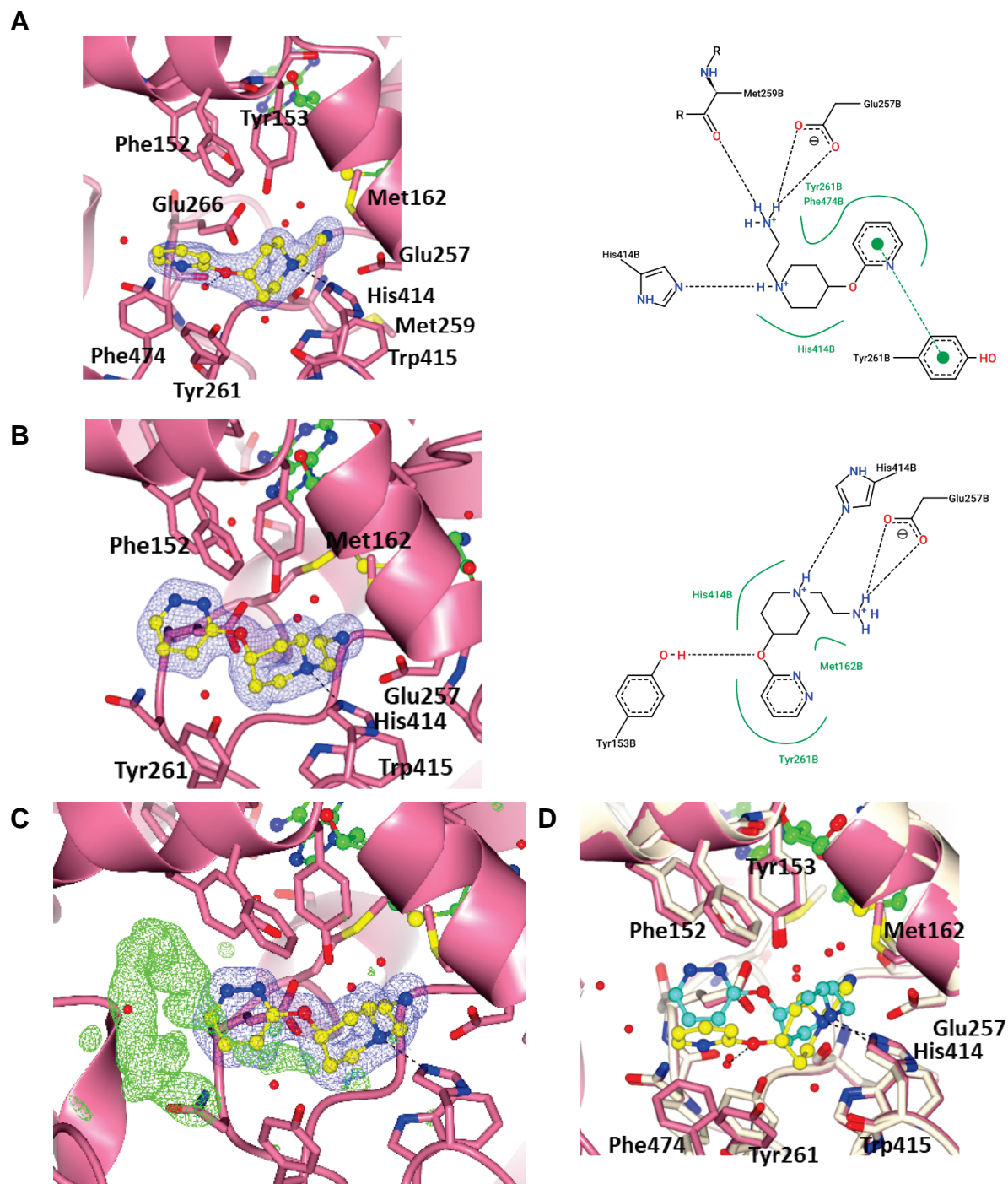


Figure 6-33 Characterisation of NCL-00024861 and NCL-00024862 Binding to CARM1.

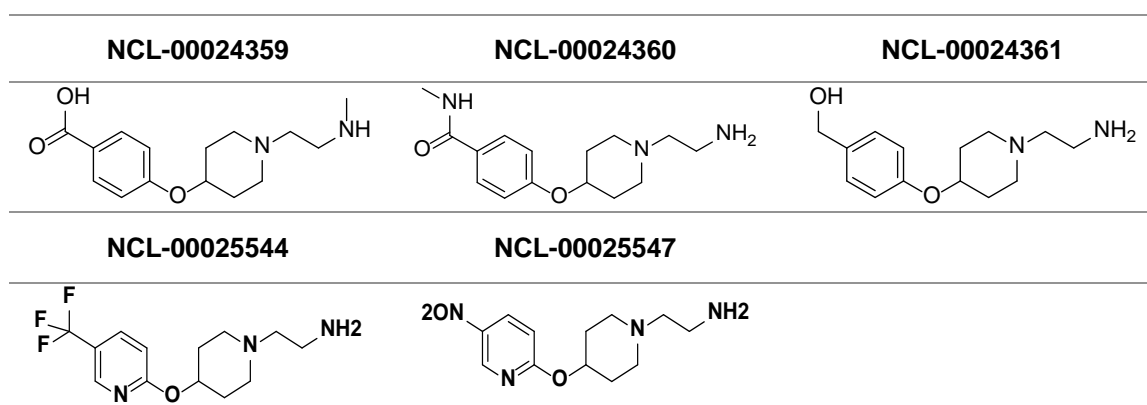
A) CARM1-NCL-00024862 structure. **B)** CARM1-NCL-00024861 structure. **C)** CARM1-NCL-00024861 structure with unassigned density. **D)** Superposition of NCL-00024861(cyan) and NCL-00024862 (yellow). The $2mF_o-DF_c$ map around the CARM1 ligand is shown as blue chicken wire and contoured to 1σ , difference map (green) is contoured at 3σ . 2D PoseView Diagram. Green dotted line: π -stacking, green smooth contour line: hydrophobic interactions. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor–hydrogen–donor angle must not fall $< 120^\circ$.

6.8.2 Substitution of the 2-Pyridine or Phenyl Ring

Three fragments with polar substitutions on the phenyl-ring were designed (Table 6-7). However, for all three fragments, (NCL-00024359/360/361), binding could not be observed by SPR and crystal structures could not be determined.

However, crystal structures were solved of CARM1 bound to NCL-00025544 and to NCL-00025547 (Figure 6-34). SPR analysis is currently ongoing. The structure of NCL-00025544 (Figure 6-34 A) shows that hydrogen bonds are formed between the free and secondary amine group of the ligand and Glu257, Met259 and His414. The trifluoromethyl-group reaches into a hydrophobic pocket formed by Phe152 and Tyr261. The latter also forms π -stacking interactions with the 2-pyridine ring. The electron density around the fluorine atoms is not visible, suggesting that they are flexible in the structure. The structure of NCL-00025547 (Figure 6-34 B) shows that the ligand adopts a different conformation (Figure 6-34 C). The 2-pyridine ring forms π -stacking interactions with Phe152, the free amine group forms a hydrogen bond with Tyr153. The piperidine ring is also in a different position and forms hydrophobic interactions with His414, Met162, Trp415 and Phe152. The polar nitro group does not interact with any neighbouring residues. However, it needs to be noted that the two structures are at lower resolution and higher-resolution and better-quality models will be needed in the future.

Table 6-7 2-Pyridine or Phenyl Ring Substitutions.



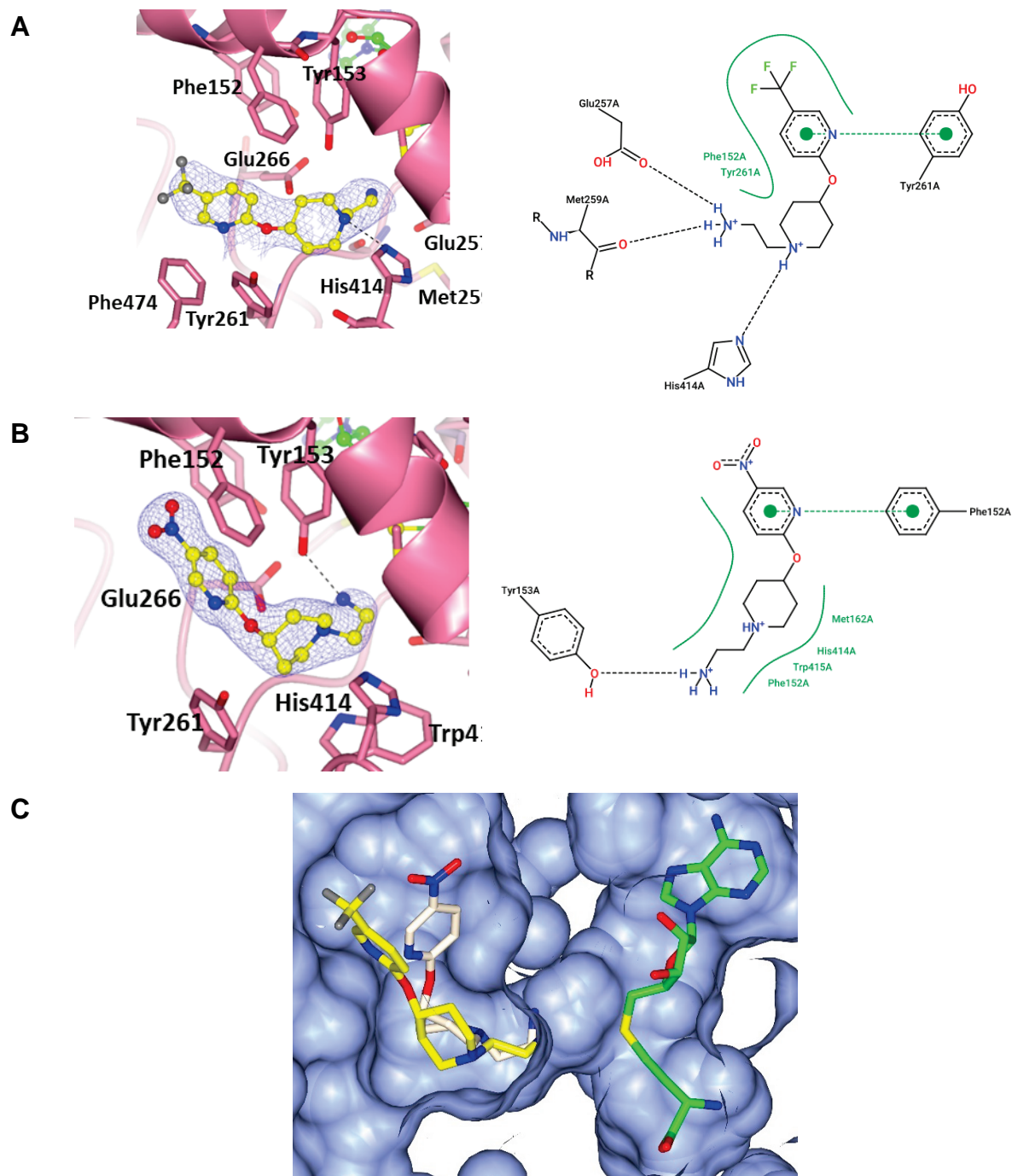


Figure 6-34 Structures of CARM1 in Complex with NCL-00025544 and NCL-00025547 Determined by Co-Crystallisation.

A) CARM1-NCL-00025544 **B)** CARM1-NCL-00025547. The $2mF_o-DF_c$ maps around the CARM1 ligands are shown as blue chicken wire and contoured at 1σ . 2D PoseView Diagram. Hydrogen bonds are shown as black dotted lines and follow the measures implemented by Desiraju and Steiner (2001), with an optimal hydrogen bond distance of $1.9\text{ \AA} \pm 0.5\text{ \AA}$ and the condition that the acceptor-hydrogen-donor angle must not fall $< 120^\circ$. Green dotted line: π -stacking, green smooth contour line: hydrophobic interactions. **C)** CARM1 binding pockets with NCL-00025544, NCL-00025547 and SAH are drawn with carbon atoms coloured yellow, cream and green, respectively.

In summary, the best fragments identified from the optimisation series, were ligand NCL-00024857 ($K_d = 2.4 \mu\text{M}$) and NCL-00024862 ($K_d = 2.9 \mu\text{M}$). Both fragments have a middle O-linker, and their pyridine ring is substituted with only one nitrogen atom at position 4 or 2, respectively. A 2-pyrimidinyl-group or 2-pyrazinyl group was not well tolerated (NCL-00024856, NCL-00024855). In the future, it would be useful to design a fragment where the pyridine ring is substituted at position 2 and 4 to see if ligand binding can be increased.

Further fragment optimisation was attempted by designing transition-state mimicking compounds that incorporate the here identified ligands but not only bind the substrate binding pocket but also reach into the SAM-binding site. The design and analysis of these bisubstrate fragments will be discussed in the next section.

6.9 Analysis of Bisubstrate Fragment Binding to CARM1

Different approaches exist for fragment optimisation, (Figure 6-35 (Lamoree and Hubbard, 2017)). A fragment can be optimised by adding chemical groups to the core fragment so that it grows into adjacent pockets (Fragment growing) (Ciulli and Abell, 2007).

Additionally, if two or more fragments exist that bind to separate but adjacent binding pockets, they can be linked to form one larger fragment that binds with higher affinity. Such fragment linking was first described by Shuker *et al.*, who used fragments identified by NMR (Shuker *et al.*, 1996). According to the concept of “super-additivity” the resulting fragment has a greater ΔG binding energy than the sum of the binding energies of the separate fragments that were linked together (Nazaré *et al.*, 2012). Third, fragments whose binding pockets overlap can be merged into one larger, more potent fragment (Ciulli and Abell, 2007). In the case of PRMTs, the substrate and co-factor binding pockets are located next to each other and are connected via the arginine binding tunnel. Thus, fragment optimisation via linking fragment hits to the SAM and substrate binding sites should be possible.

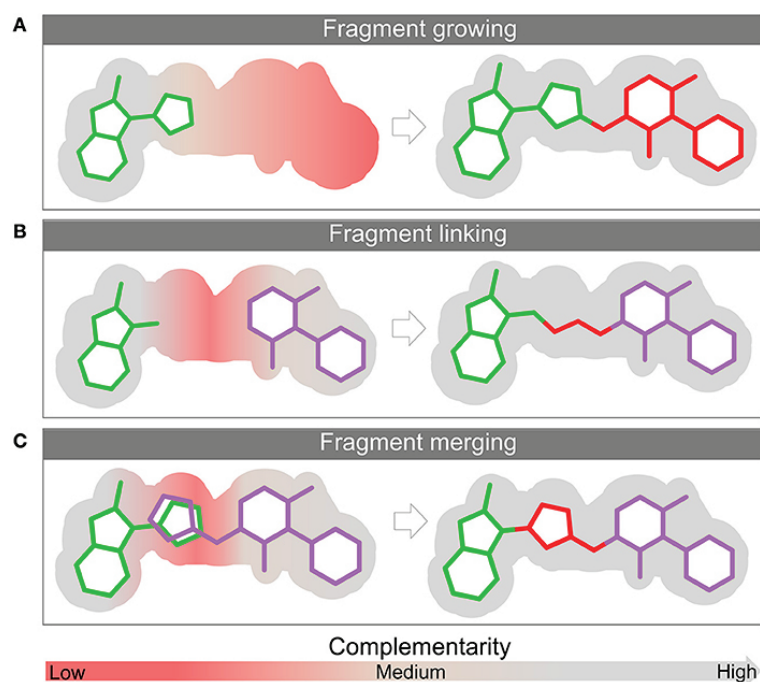


Figure 6-35 Approaches Used for Fragment Optimisation.

Fragments can be optimised by **A) fragment growing** **B) fragment linking** and **C) fragment merging**. The grey colour highlights pockets that have a high complementarity with the fragment, red areas with low complementarity. Figure taken, with permission, from de Souza Neto et al. (2020).

Figure 6-36 shows the two binding pockets of the CARM1-SAH co-crystal structure that includes one of the most potent newly identified fragments, NCL-00024062. However, the linker composition can potentially change the ligand orientations and cause a decrease in binding affinity (Ciulli and Abell, 2007).

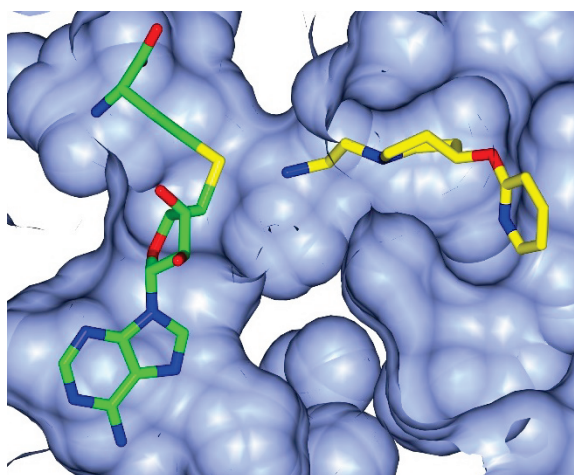


Figure 6-36 Location of the CARM1-SAH and NCL-00024062 Binding Pockets.

The binding pocket of SAH (carbon atoms in green) is connected to the substrate binding pocket, occupied by NCL-00024062 (yellow cylinders), via the arginine site channel.

The development of such bisubstrate inhibitors for PRMTs could help to overcome the selectivity issues seen with SAM mimics because the substrate binding site is more diverse, thus developing selective compounds is more likely. Additionally, they could also lead to the development of compounds with higher cell-permeability by decreasing the polarity of the ligand.

Different bisubstrate PMT inhibitors have been reported in the last couple of years, for example for the N-Terminal Methyltransferase 1 (Chen *et al.*, 2019) but also some inhibitors that showed activity against CARM1 (van Haren *et al.*, 2015). In the study by van Haren *et al.*, all PRMT inhibitors that showed activity against CARM1 contained the adenosine group of the cofactor which is connected to a guanidine moiety via a variable linker (van Haren *et al.*, 2015). Two and three atom spacers yielded the most potent CARM1 inhibitors, and the ligands showed no activity against the PKMT G9a (Figure 6-37). Ligand 1 and 3 inhibited CARM1 with IC_{50} values of $0.12 \pm 0.02 \mu\text{M}$ and $0.15 \pm 0.05 \mu\text{M}$ respectively (van Haren *et al.*, 2015). Another bisubstrate CARM1 inhibitor that showed 100x higher affinity against CARM1 than PRMT1 with an IC_{50} of $1.5 \mu\text{M}$ contained an amino-adenosine moiety linked to a cysteine moiety (Figure 6-37, ligand 4) (Halby *et al.*, 2018). In another study, van Haren *et al.* (2017) reported a more selective CARM1 inhibitor with an IC_{50} value of $0.09 \mu\text{M}$ for CARM1 and $26 \mu\text{M}$ for PRMT1.

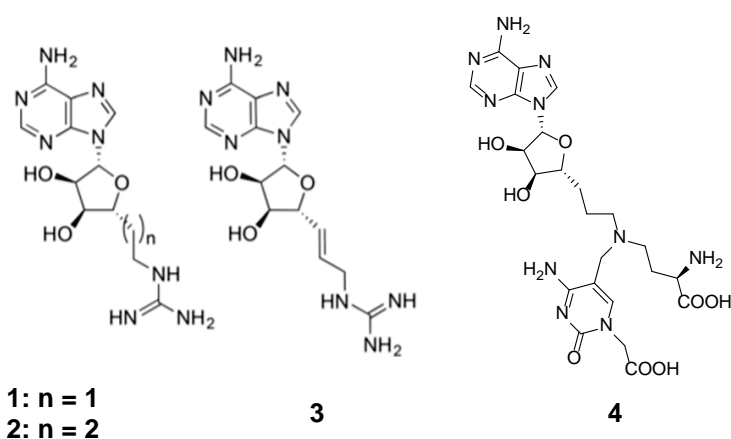


Figure 6-37 Bisubstrate PRMT Inhibitors.

However, one potential drawback of these inhibitors is susceptibility to proteolysis. A new set of non-peptide PRMT inhibitors that target both pockets has been reported very recently by Gunnell *et al.* (2020). The initial set of compounds contained amines and alkyl-guanidinium groups. A linker length of 1-2 spacer atoms gave the

highest inhibitory activity for CARM1. Further SAR studies showed that the selectivity of the ligands for CARM1 can be increased by replacing the guanidinium group with a 2-aminopyridine group. At the same time this also increases the polarity of the ligand. The most potent inhibitor to date was identified by Gunnell *et al.* (2020) has an IC₅₀ value of $1.1 \pm 0.132 \mu\text{M}$ for CARM1 and $25 \mu\text{M}$ for PRMT1 (Figure 6-38, ligand 10). Crystal structures of the catalytic domain of CARM1 in complex with the ligands taken together with mutation studies, showed that the high selectivity is probably mediated by an asparagine residue located in the arginine site channel (Asn265), which is a tyrosine in PRMT1 (Gunnell *et al.*, 2020). Sequence conservation at this site is low compared with other PRMT family members. Thus, these new fragments should be a good starting point for further optimisation and development of more selective and cell-permeable bisubstrate CARM1 inhibitors.

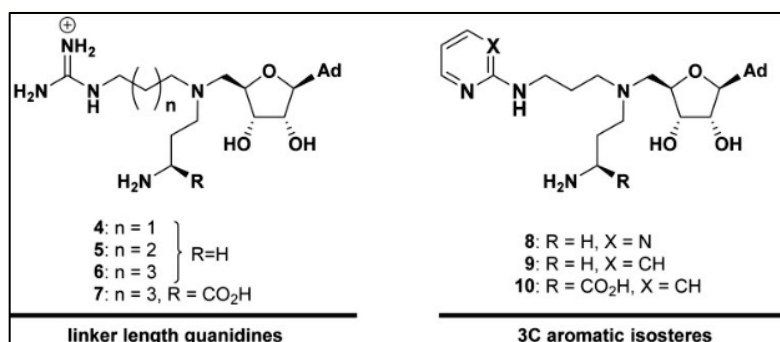


Figure 6-38 Newly Identified Bisubstrate Inhibitors of CARM1.

Figure taken, with permission from Gunnell *et al.* (2020).

In the current study a total of 13 different hybrid ligands were designed (Table 6-8). All contain a ribose moiety intended to bind into the SAM pocket connected to different previously designed substrate-competitive ligands via varying linkers.

Table 6-8 Chemical Structure of Designed Bisubstrate Ligands.

NCL-00025093		NCL-00024969	
NCL-00025094		NCL-00025005	
NCL-00025210		NCL-00025060	
NCL-00025211		NCL-00025061	
NCL-00025138		NCL-00025062	
NCL-00024965		NCL-00024972	
NCL-00024971			

Five of these hybrid ligands bound to CARM1 as measured by SPR (Figure 6-39). Experiments were performed with and without SAH in the SPR buffer. SPR analyses examples for ligands that did not show any binding can be found in Appendix L. The only fragment that almost reached saturation during the SPR analysis was NCL00025094 (Figure 6-39 B) with an estimated K_d value of 80 μM . Binding curves were only observed if SAH was present in the SPR buffer. Unfortunately, a crystal-structure could not be determined by CARM1-SAH crystal soaking experiments or co-crystallisation trials with and without SAH. The results indicate that ligand containing only the ribose moiety are not potent enough to replace the SAM/SAH molecule. NCL-0002471 and NCL-00025094 binding was also tested by ITC, using 20 μM CARM1 in the cell without SAH and 200 μM ligand in the syringe, but no binding was detected (Figure 6-40).

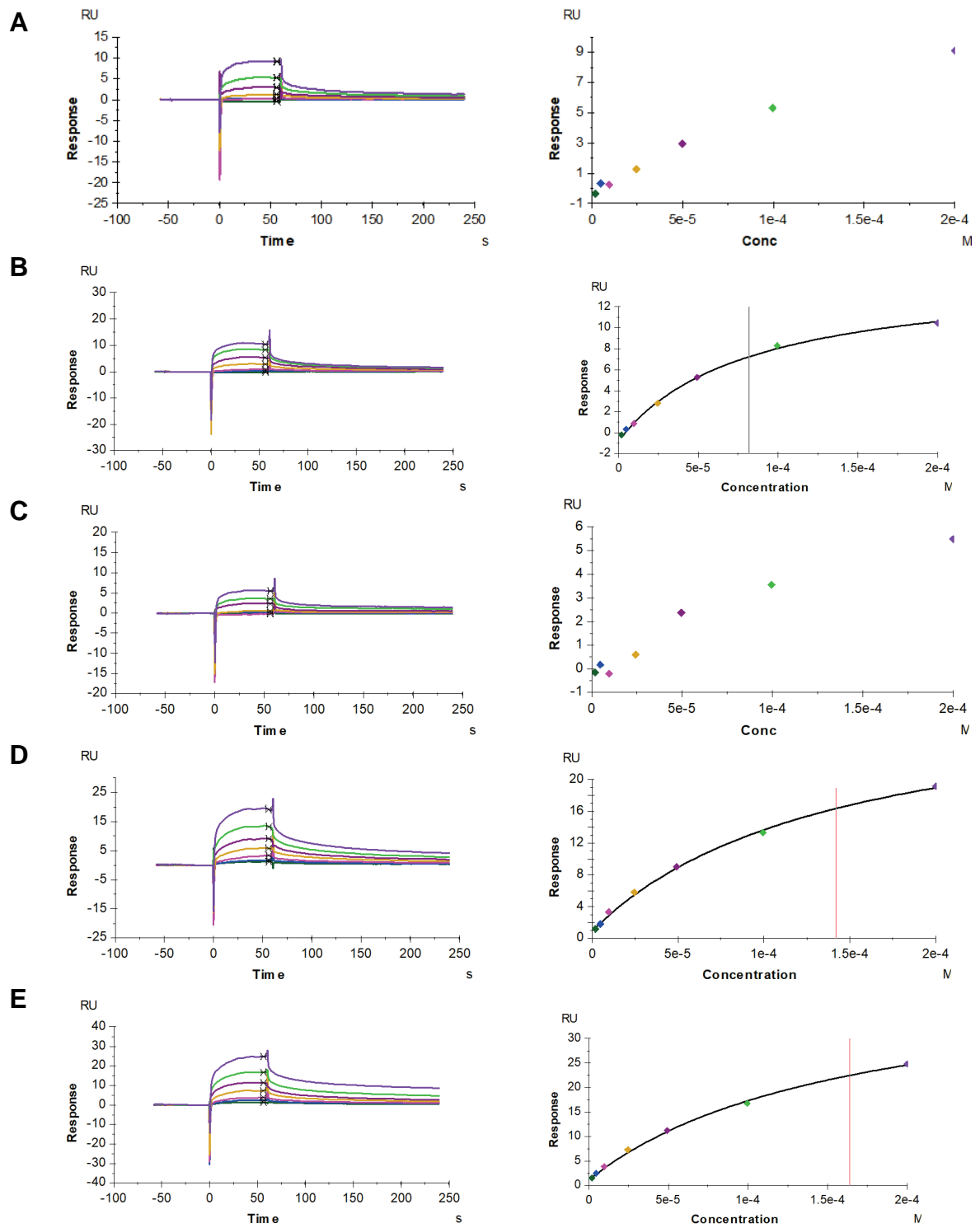


Figure 6-39 SPR Analysis of Bisubstrate Ligand Binding to CARM1.

A) NCL-00025093; **B)** NCL-00025094, **C)** NCL-00025318; **D)** NCL-00024972; **E)** NCL-0002471. The black vertical line in the righthanded plot indicates the value of the calculated K_d . Red vertical lines indicate that the reported K_d should not be trusted as they are higher than half the highest analyte concentration.

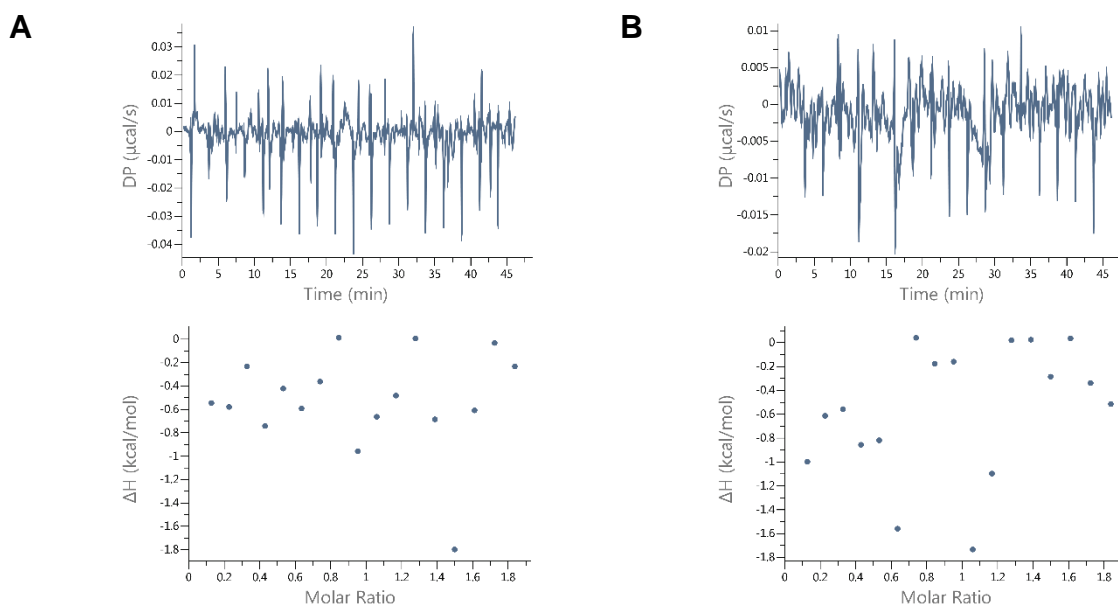


Figure 6-40 ITC Analysis of NCL-0002471 and NCL-00025094 and CARM1.

ITC titration experiment of the ligands against CARM1 were performed 20 °C with 20 μM CARM1 in the cell and 200 μM ligand in the syringe.

6.10 Analysis of FragLite Binding to CARM1

FragLites are small, halogenated compounds that have paired hydrogen-bonding motifs and can be used to identify new ligand binding sites on proteins. Bound FragLites are identified exploiting the anomalous scattering of the halogen substituent in an X-ray diffraction experiment. FragLites can not only be used to map ligand sites but can also be used as a starting point for fragment development via linking of different FragLite molecules, as has been demonstrated for cyclin-dependent kinase 2 (CDK2) (Wood *et al.*, 2019). The FragLite set of 31 compounds (Wood *et al.*, 2019) that was used for soaking experiments into existing SAH-CARM1 crystals is listed in Appendix N. Each fragment has a heavy halogen atom and a pharmacophore doublet, either a hydrogen bond acceptor and donor, or acceptor and acceptor, which are connected by linker atoms with varying length (1-5 atoms).

CARM1-SAH crystal soaks were conducted as previously described (Section 2.22.4) with a final FragLite concentration of 5 mM and 10 % DMSO. Unfortunately, fragment binding was not detected. One possible explanation for this result is that the CARM1 protein used for crystallisation experiments is the catalytic domain where the cofactor binding site is already occupied by SAH. Thus, identifying new binding sites

within the HsCARM1^{CAT} crystals is very unlikely. It would make sense to repeat the experiments with full-length CARM1.

6.11 Summary

Different PRMT inhibitors were designed that either target the substrate binding site of CARM1 or the cofactor pocket. Additionally, bisubstrate inhibitors were designed via fragment linking. Fragment binding was analysed by SPR and X-ray crystallography. Unfortunately, no binding was observed for the SAM-competitive fragments, which most likely bind with weak affinities and are not able to replace the bound cofactor.

Fragments designed to target the substrate binding site were more successful and different co-crystal structures in complex with the catalytic domain of CARM1 have been solved. Together with the SPR binding data, the most promising fragments were selected, and hybrid ligands were designed by fragment linking to the ribose moiety of SAM. Most of the fragments did not bind, but a handful compounds showed binding to CARM1 with high μM affinities. Optimisation of these fragments could lead to the development of more potent, selective, and cell-(permeable CARM1 inhibitors. The recently published CARM1 structures in complex with different bisubstrate inhibitors (Gunnell *et al.*, 2020) can help to guide fragment design, and the design of fragments that interact with the identified selectivity residues in CARM1 should increase selectivity of the ligands.

The FragLite screen conducted with the catalytic domain of CARM1 did not yield any hits. However, additional binding sites could potentially be identified if using the full-length CARM1 protein and other PRMT family members. Moreover, FragLites screen could be used to map protein-ligand or potentially also protein-protein interactions sites of the N-terminal domain.

Chapter 7 Conclusion and Future Direction

Dysregulation and overexpression of PRMTs has been observed in various diseases including cancer (Baldwin *et al.*, 2014), cardio-vascular (Franceschelli *et al.*, 2013), and viral diseases (Xie *et al.*, 2007), and pulmonary disorders (Zakrzewicz *et al.*, 2012, Sun *et al.*, 2015). As a result, interest in PRMT family members has grown and the number of drug discovery projects targeting PRMTs has increased. However, the function and biology of many family members is still not fully understood. Alternative PRMT splicing creates further complexity and has been associated with disease development (Baldwin *et al.*, 2014). Murine genetic studies showed that PRMTs play an important role in development and in the case of CARM1 are necessary for viability (Yadav *et al.*, 2003). Thus, possible side-effects or toxicity are a consideration when developing PRMT drugs.

The aim of this thesis was the development of selective chemical probes for PRMTs with a focus on PRMT2. Chemical probes can be used as tool compounds not only for target validation, but also as potential start points for drug discovery. The first step was the development of an enzyme activity assay. Currently a variety of mostly indirect assays are used for activity assessment. IC₅₀ values from different assay formats are difficult to compare and, in many studies, not enough counter screen assays against a wide panel of PMTs are performed. In some cases, suitable reagents/substrates for PRMTs are not available. In this thesis, RSF1, a newly identified PRMT2 substrate was purified and assayed, and a MALDI-ToF based MS assay was developed. This assay, used in this thesis to directly analyse the methyltransferase activity of CARM1 towards histone peptides, could be further improved and used to screen more PMTs. The assay is advantageous as labelling is not needed and it already exists in an HTS format. Moreover, it can be easily adapted to assay other PRMTs if they do not auto-methylate and a peptide substrate is available.

The development of selective inhibitors is hampered by the high degree of sequence conservation within the family. Nevertheless, despite the challenges for some PRMT family members, selective inhibitors have been already developed such as TP-064 for CARM1 (Nakayama *et al.*, 2018) or GSK3326595 and LLY-283 for PRMT5 (Bonday *et al.*, 2018, Watts *et al.*, 2019). The different PRMT inhibitors that have entered clinical trials, should answer the question whether selective inhibition without toxicity can be achieved. The published PRMT2 structure together with the here described

apo- and Sinefungin-bound structures can help to develop PRMT2 inhibitors. Optimisation of PRMT2 protein production would permit fragment-based screening.

Different fragments targeting the SAM domain of CARM1 were designed and binding analysed by surface plasmon resonance analysis combined with crystallography. Unfortunately, no hits could be identified. Recently reported SAM-competitive inhibitors often reach into the arginine channel and compete with the substrate. This approach is promising, but the high cellular cofactor concentration needs to be considered when designing *in vitro* assays (Ferreira de Freitas *et al.*, 2019).

Fragments targeting the CARM1 substrate binding site also gave very promising results, especially the high affinity fragments of the pyridine nitrogen substituted series. In the next step, hybrid ligands that link substrate-competitive fragments to a ribose moiety were designed and analysed. Some of the hybrid fragments did show weak binding to CARM1 when analysed by SPR. These results together with the recently published first co-crystal structure of CARM1 in complex with a new bisubstrate inhibitor (Gunnell *et al.*, 2020) could be used to develop more potent, selective, and cell-permeable inhibitors that target both binding pockets. In particular, fragments that interact with the suspected CARM1 selectivity driving asparagine residue (Asn265) should be designed and analysed. Unfortunately, a first FragLite screen to identify alternative fragment starting points did not yield any binding events. Repeating the screen with full-length CARM1 should be attempted to identify potential allosteric sites outside the catalytic domain.

How members of the PRMT family are regulated is still not well understood. The readers of the methyl-marks, which are currently limited to Tudor domain containing proteins, could also be targeted, as has been recently illustrated by the development of acetyl-lysine mimicking compounds towards bromo-domains (Dawson *et al.*, 2011, Mertz *et al.*, 2011). No specific demethylase has been reported to date, but it is likely that they exist. The future identification of accessory proteins, methyl-arginine effector molecules, and substrates may aid elucidation of the different signalling pathways in which PRMTs are involved and reveal additional potential drug targets.

References

- ABIDA, W. M., NIKOLAEV, A., ZHAO, W., ZHANG, W. & GU, W. 2007. FBXO11 promotes the Neddylation of p53 and inhibits its transcriptional activity. *Journal of Biological Chemistry*, 282, 1797-1804.
- ACKER, M. G. & AULD, D. S. 2014. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*, 1, 56-73.
- ADAMS-CIOABA, M. A. & MIN, J. 2009. Structure and function of histone methylation binding proteins. *Biochemistry and Cell Biology*, 87, 93-105.
- AITIO, O., HELLMAN, M., KAZLAUSKAS, A., VINGADASSALOM, D. F., LEONG, J. M., SAKSELA, K. & PERMI, P. 2010. Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly. *Proceedings of the National Academy of Sciences*, 107, 21743-21748.
- ALINARI, L., MAHASENAN, K. V., YAN, F., KARKHANIS, V., CHUNG, J.-H., SMITH, E. M., QUINION, C., SMITH, P. L., KIM, L. & PATTON, J. T. 2015. Selective inhibition of protein arginine methyltransferase 5 blocks initiation and maintenance of B-cell transformation. *Blood, The Journal of the American Society of Hematology*, 125, 2530-2543.
- ALLAN, J., HARTMAN, P., CRANE-ROBINSON, C. & AVILES, F. 1980. The structure of histone H1 and its location in chromatin. *Nature*, 288, 675-679.
- AN, W., KIM, J. & ROEDER, R. G. 2004. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell*, 117, 735-748.
- ANTONYSAMY, S., BONDAY, Z., CAMPBELL, R. M., DOYLE, B., DRUZINA, Z., GHEYI, T., HAN, B., JUNGHEIM, L. N., QIAN, Y. & RAUCH, C. 2012. Crystal structure of the human PRMT5: MEP50 complex. *Proceedings of the National Academy of Sciences*, 109, 17960-17965.
- BALDWIN, R. M., MORETTIN, A. & CÔTÉ, J. 2014. Role of PRMTs in cancer: Could minor isoforms be leaving a mark? *World Journal of Biological Chemistry*, 5, 115.
- BALDWIN, R. M., MORETTIN, A., PARIS, G., GOULET, I. & CÔTÉ, J. 2012. Alternatively spliced protein arginine methyltransferase 1 isoform PRMT1v2 promotes the survival and invasiveness of breast cancer cells. *Cell Cycle*, 11, 4597-4612.
- BALL, L. J., JARCHAU, T., OSCHKINAT, H. & WALTER, U. 2002. EVH1 domains: structure, function and interactions. *FEBS Letters*, 513, 45-52.
- BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21, 381-395.
- BARRALLO-GIMENO, A. & NIETO, M. A. 2005. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development*, 132, 3151-3161.
- BATLLE, E., SANCHO, E., FRANCÍ, C., DOMÍNGUEZ, D., MONFAR, M., BAULIDA, J. & DE HERREROS, A. G. 2000. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature Cell Biology*, 2, 84-89.
- BEDFORD, M. T. 2007. Arginine methylation at a glance. *Journal of Cell Science*, 120, 4243-4246.
- BEDFORD, M. T. & CLARKE, S. G. 2009. Protein arginine methylation in mammals: who, what, and why. *Molecular Cell*, 33, 1-13.

- BEDFORD, M. T., FRANKEL, A., YAFFE, M. B., CLARKE, S., LEDER, P. & RICHARD, S. 2000. Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *Journal of Biological Chemistry*, 275, 16030-16036.
- BEMBENEK, S. D., TOUNGE, B. A. & REYNOLDS, C. H. 2009. Ligand efficiency and fragment-based drug discovery. *Drug Discovery Today*, 14, 278-283.
- BERGER, I., FITZGERALD, D. J. & RICHMOND, T. J. 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnology*, 22, 1583.
- BHAL, S. K. 2007. LogP—Making Sense of the Value. *Application Note*. Toronto: Advanced Chemistry Development Inc.
- BHATT, V. S., ZENG, D., KRIEGER, I., SACCHETTINI, J. C. & CHO, J.-H. 2016. Binding mechanism of the N-terminal SH3 domain of CrklI and proline-rich motifs in cAbl. *Biophysical Journal*, 110, 2630-2641.
- BICKER, K. & THOMPSON, P. 2013. The protein arginine deiminases: Structure, function, inhibition, and disease. *Biopolymers*, 99, 155-163.
- BISSINGER, E.-M., HEINKE, R., SPANNHOFF, A., EBERLIN, A., METZGER, E., CURA, V., HASSENBOEHLER, P., CAVARELLI, J., SCHÜLE, R. & BEDFORD, M. T. 2011. Acyl derivatives of p-aminosulfonamides and dapsone as new inhibitors of the arginine methyltransferase hPRMT1. *Bioorganic & Medicinal Chemistry*, 19, 3717-3731.
- BISSWANGER, H. 2014. Enzyme Assays. *Perspectives in Science*, 1, 41-55.
- BJÖRKEGREN, C. & BARANELLO, L. 2018. DNA supercoiling, topoisomerases, and cohesin: partners in regulating chromatin architecture? *International Journal of Molecular Sciences*, 19, 884.
- BLUS, B. J., WIGGINS, K. & KHORASANIZADEH, S. 2011. Epigenetic virtues of chromodomains. *Critical Reviews in Biochemistry and Molecular Biology*, 46, 507-526.
- BLYTHE, S. A., CHA, S.-W., TADJUIDJE, E., HEASMAN, J. & KLEIN, P. S. 2010. β -Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Developmental Cell*, 19, 220-231.
- BOECKEL, J.-N., GUARANI, V., KOYANAGI, M., ROEXE, T., LENGELING, A., SCHERMULY, R. T., GELLERT, P., BRAUN, T., ZEIHNER, A. & DIMMELER, S. 2011. Jumonji domain-containing protein 6 (Jmjd6) is required for angiogenic sprouting and regulates splicing of VEGF-receptor 1. *Proceedings of the National Academy of Sciences*, 108, 3276-3281.
- BONDAY, Z. Q., CORTEZ, G. S., GROGAN, M. J., ANTONYSAMY, S., WEICHERT, K., BOCCHINFUSO, W. P., LI, F., KENNEDY, S., LI, B. & MADER, M. M. 2018. LLY-283, a potent and selective inhibitor of arginine methyltransferase 5, PRMT5, with antitumor activity. *ACS Medicinal Chemistry Letters*, 9, 612-617.
- BORCHARDT, R. T., EIDEN, L. E., WU, B. & RUTLEDGE, C. O. 1979. Sinefungin, a potent inhibitor of S-adenosylmethionine: protein O-methyltransferase. *Biochemical and Biophysical Research Communications*, 89, 919-924.
- BORIACK-SJODIN, P. A., JIN, L., JACQUES, S. L., DREW, A., SNEERINGER, C., SCOTT, M. P., MOYER, M. P., RIBICH, S., MORADEI, O. & COPELAND, R. A. 2015. Structural insights into ternary complex formation of human CARM1 with various substrates. *ACS Chemical Biology*, 11, 763-771.
- BORIACK-SJODIN, P. A. & SWINGER, K. K. 2016. Protein methyltransferases: a distinct, diverse, and dynamic family of enzymes. *Biochemistry*, 55, 1557-1569.

- BÖTTGER, A., ISLAM, M. S., CHOWDHURY, R., SCHOFIELD, C. J. & WOLF, A. 2015. The oxygenase Jmjd6—a case study in conflicting assignments. *Biochemical Journal*, 468, 191-202.
- BOTUYAN, M. V. & MER, G. 2016. Tudor Domains as Methyl-Lysine and Methyl-Arginine Readers. *Chromatin Signaling and Diseases*. Cambridge, MA: Elsevier.
- BOULANGER, M.-C., LIANG, C., RUSSELL, R. S., LIN, R., BEDFORD, M. T., WAINBERG, M. A. & RICHARD, S. 2005. Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. *Journal of Virology*, 79, 124-131.
- BRADLEY, T., COOK, M. E. & BLANCHETTE, M. 2015. SR proteins control a complex network of RNA-processing events. *RNA*, 21, 75-92.
- BRAND, S. F., PICHOFF, S., NOSELLI, S. & BOURBON, H.-M. 1995. Novel *Drosophila melanogaster* genes encoding RRM-type RNA-binding proteins identified by a degenerate PCR strategy. *Gene*, 154, 187-192.
- BRANSCOMBE, T. L., FRANKEL, A., LEE, J.-H., COOK, J. R., YANG, Z.-H., PESTKA, S. & CLARKE, S. 2001. PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. *Journal of Biological Chemistry*, 276, 32971-32976.
- BROWN, N. P., LEROY, C. & SANDER, C. 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics (Oxford, England)*, 14, 380-381.
- BUYSE, I. M., SHAO, G. & HUANG, S. 1995. The retinoblastoma protein binds to RIZ, a zinc-finger protein that shares an epitope with the adenovirus E1A protein. *Proceedings of the National Academy of Sciences*, 92, 4467-4471.
- CAI, X.-C., ZHANG, T., KIM, E.-J., JIANG, M., WANG, K., WANG, J., CHEN, S., ZHANG, N., WU, H. & LI, F. 2019. A chemical probe of CARM1 alters epigenetic plasticity against breast cancer cell invasion. *eLife*, 2019, 1-42.
- CAMPAGNA-SLATER, V., MOK, M. W., NGUYEN, K. T., FEHER, M., NAJMANOVICH, R. & SCHAPIRA, M. 2011. Structural chemistry of the histone methyltransferases cofactor binding site. *Journal of Chemical Information and Modeling*, 51, 612-623.
- CARASCOSSA, S., DUDEK, P., CENNI, B., BRIAND, P.-A. & PICARD, D. 2010. CARM1 mediates the ligand-independent and tamoxifen-resistant activation of the estrogen receptor α by cAMP. *Genes & Development*, 24, 708-719.
- CARRIO, M. & VILLAVERDE, A. 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*, 96, 3-12.
- CASADIO, F., LU, X., POLLOCK, S. B., LEROY, G., GARCIA, B. A., MUIR, T. W., ROEDER, R. G. & ALLIS, C. D. 2013. H3R42me2a is a histone modification with positive transcriptional effects. *Proceedings of the National Academy of Sciences*, 110, 14894-14899.
- CHAN-PENEBRE, E., KUPLAST, K. G., MAJER, C. R., BORIACK-SJODIN, P. A., WIGLE, T. J., JOHNSTON, L. D., RIOUX, N., MUNCHHOF, M. J., JIN, L. & JACQUES, S. L. 2015. A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nature Chemical Biology*, 11, 432.
- CHAN, B., LANYI, A., SONG, H. K., GRIESBACH, J., SIMARRO-GRANDE, M., POY, F., HOWIE, D., SUMEGI, J., TERHORST, C. & ECK, M. J. 2003. SAP couples Fyn to SLAM immune receptors. *Nature Cell Biology*, 5, 155-160.

- CHANG, B., CHEN, Y., ZHAO, Y. & BRUICK, R. K. 2007. JMJD6 is a histone arginine demethylase. *Science*, 318, 444-447.
- CHANG, N. C., SINCENNES, M.-C., CHEVALIER, F. P., BRUN, C. E., LACARIA, M., SEGALÉS, J., MUÑOZ-CÁNOVES, P., MING, H. & RUDNICKI, M. A. 2018. The dystrophin glycoprotein complex regulates the epigenetic activation of muscle stem cell commitment. *Cell Stem Cell*, 22, 755-768. e6.
- CHANG, X. & HAN, J. 2006. Expression of peptidylarginine deiminase type 4 (PAD4) in various tumors. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, 45, 183-196.
- CHAROENSUKSAI, P., KUHN, P., WANG, L., SHERER, N. & XU, W. 2015. O-GlcNAcylation of co-activator-associated arginine methyltransferase 1 regulates its protein substrate specificity. *Biochemical Journal*, 466, 587-599.
- CHEN, C., NOTT, T. J., JIN, J. & PAWSON, T. 2011. Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews Molecular Cell Biology*, 12, 629.
- CHEN, D., DONG, G., NOINAJ, N. & HUANG, R. 2019. Discovery of Bisubstrate Inhibitors for Protein N-Terminal Methyltransferase 1. *Journal of Medicinal Chemistry*, 62, 3773-3779.
- CHEN, D., MA, H., HONG, H., KOH, S. S., HUANG, S.-M., SCHURTER, B. T., ASWAD, D. W. & STALLCUP, M. R. 1999. Regulation of transcription by a protein methyltransferase. *Science*, 284, 2174-2177.
- CHEN, H., LORTON, B., GUPTA, V. & SHECHTER, D. 2017. A TGF β -PRMT5-MEP50 axis regulates cancer cell invasion through histone H3 and H4 arginine methylation coupled transcriptional activation and repression. *Oncogene*, 36, 373-386.
- CHEN, Z., ZANG, J., KAPPLER, J., HONG, X., CRAWFORD, F., WANG, Q., LAN, F., JIANG, C., WHETSTINE, J. & DAI, S. 2007. Structural basis of the recognition of a methylated histone tail by JMJD2A. *Proceedings of the National Academy of Sciences*, 104, 10818-10823.
- CHENG, D., CÔTÉ, J., SHAABAN, S. & BEDFORD, M. T. 2007. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Molecular Cell*, 25, 71-83.
- CHENG, D., YADAV, N., KING, R. W., SWANSON, M. S., WEINSTEIN, E. J. & BEDFORD, M. T. 2004. Small molecule regulators of protein arginine methyltransferases. *Journal of Biological Chemistry*, 279, 23892-23899.
- CHENG, X., COLLINS, R. E. & ZHANG, X. 2005. Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.*, 34, 267-294.
- CHEUNG, W. D., SAKABE, K., HOUSLEY, M. P., DIAS, W. B. & HART, G. W. 2008. O-linked β -N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *Journal of Biological Chemistry*, 283, 33935-33941.
- CHI, P., ALLIS, C. D. & WANG, G. G. 2010. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer*, 10, 457.
- CHUMANOV, R. S., KUHN, P. A., XU, W. & BURGESS, R. R. 2011. Expression and purification of full-length mouse CARM1 from transiently transfected HEK293T cells using HaloTag technology. *Protein Expression and Purification*, 76, 145-153.

- CIULLI, A. & ABELL, C. 2007. Fragment-based approaches to enzyme inhibition. *Current Opinion in Biotechnology*, 18, 489-496.
- CLEMENTS, A., POUX, A. N., LO, W.-S., PILLUS, L., BERGER, S. L. & MARMORSTEIN, R. 2003. Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. *Molecular Cell*, 12, 461-473.
- CONGREVE, M., CARR, R., MURRAY, C. & JHOTI, H. 2003. A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today*, 19, 876-877.
- CONGREVE, M., CHESSARI, G., TISI, D. & WOODHEAD, A. J. 2008. Recent developments in fragment-based drug discovery. *Journal of medicinal chemistry*, 51, 3661-3680.
- COOK, J. R., LEE, J.-H., YANG, Z.-H., KRAUSE, C. D., HERTH, N., HOFFMANN, R. & PESTKA, S. 2006. FBXO11/PRMT9, a new protein arginine methyltransferase, symmetrically dimethylates arginine residues. *Biochemical and Biophysical Research Communications*, 342, 472-481.
- COPELAND, R. A. 2018. Protein methyltransferase inhibitors as precision cancer therapeutics: a decade of discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170080.
- COPELAND, R. A., SOLOMON, M. E. & RICHON, V. M. 2009. Protein methyltransferases as a target class for drug discovery. *Nature Reviews Drug discovery*, 8, 724-732.
- CORLESS, S. & GILBERT, N. 2016. Effects of DNA supercoiling on chromatin architecture. *Biophysical Reviews*, 8, 245-258.
- CÔTÉ, J. & RICHARD, S. 2005. Tudor domains bind symmetrical dimethylated arginines. *Journal of Biological Chemistry*, 280, 28476-28483.
- CURA, V., MARECHAL, N., TROFFER-CHARLIER, N., STRUB, J. M., VAN HAREN, M. J., MARTIN, N. I., CIANFÉRANI, S., BONNEFOND, L. & CAVARELLI, J. 2017. Structural studies of protein arginine methyltransferase 2 reveal its interactions with potential substrates and inhibitors. *The FEBS Journal*, 284, 77-96.
- CURA, V., TROFFER-CHARLIER, N., WURTZ, J.-M., 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 Crystallographica Section D: Biological Crystallography*, 70, 2401-2412.
- CUTHBERT, G. L., DAUJAT, S., SNOWDEN, A. W., ERDJUMENT-BROMAGE, H., HAGIWARA, T., YAMADA, M., SCHNEIDER, R., GREGORY, P. D., TEMPST, P. & BANNISTER, A. J. 2004. Histone deimination antagonizes arginine methylation. *Cell*, 118, 545-553.
- DAIGLE, S. R., OLHAVA, E. J., THERKELSEN, C. A., BASAVAPATHRUNI, A., JIN, L., BORIACK-SJODIN, P. A., ALLAIN, C. J., KLAUS, C. R., RAIMONDI, A. & SCOTT, M. P. 2013. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood, The Journal of the American Society of Hematology*, 122, 1017-1025.
- DAWSON, M. A., PRINJHA, R. K., DITTMANN, A., GIOTOPOULOS, G., BANTSCHIEFF, M., CHAN, W.-I., ROBSON, S. C., CHUNG, C.-W., HOPF, C. & SAVITSKI, M. M. 2011. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, 478, 529-533.
- DE SOUZA NETO, L. R., MOREIRA-FILHO, J. T., NEVES, B. J., MAIDANA, R. L. B. R., GUIMARÃES, A. C. R., FURNHAM, N., ANDRADE, C. H. & SILVA JR, F.

- P. 2020. In silico strategies to support fragment-to-lead optimization in drug discovery. *Frontiers in Chemistry*, 8, 1-18.
- DELANO, W. L. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 82-92.
- DELLER, M. C., KONG, L. & RUPP, B. 2016. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F: Structural Biology Communications*, 72, 72-95.
- DESIRAJU, G. R. & STEINER, T. 2001. *The weak hydrogen bond: in structural chemistry and biology*, International Union of Crystal.
- DI LORENZO, A. & BEDFORD, M. T. 2011. Histone arginine methylation. *FEBS Letters*, 585, 2024-2031.
- DILLON, M. B., RUST, H. L., THOMPSON, P. R. & MOWEN, K. A. 2013. Automethylation of protein arginine methyltransferase 8 (PRMT8) regulates activity by impeding S-adenosylmethionine sensitivity. *Journal of Biological Chemistry*, 288, 27872-27880.
- DILLON, S. C., ZHANG, X., TRIEVEL, R. C. & CHENG, X. 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology*, 6, 227.
- DLAKIC, M. 2001. Chromatin silencing protein and pachytene checkpoint regulator Dot1p has a methyltransferase fold. *Trends in Biochemical Sciences*, 26, 405-407.
- DONALDSON, L. W., GISH, G., PAWSON, T., KAY, L. E. & FORMAN-KAY, J. D. 2002. Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. *Proceedings of the National Academy of Sciences*, 99, 14053-14058.
- DONG, F., LI, Q., YANG, C., HUO, D., WANG, X., AI, C., KONG, Y., SUN, X., WANG, W. & ZHOU, Y. 2018. PRMT2 links histone H3R8 asymmetric dimethylation to oncogenic activation and tumorigenesis of glioblastoma. *Nature Communications*, 9, 1-14.
- DOUANGAMATH, A., FILIPP, F. V., KLEIN, A. T., BARNETT, P., ZOU, P., VOORN-BROUWER, T., VEGA, M. C., MAYANS, O. M., SATTLER, M. & DISTEL, B. 2002. Topography for independent binding of α -helical and PPII-helical ligands to a peroxisomal SH3 domain. *Molecular Cell*, 10, 1007-1017.
- DOWDEN, J., HONG, W., PARRY, R. V., PIKE, R. A. & WARD, S. G. 2010. Toward the development of potent and selective bisubstrate inhibitors of protein arginine methyltransferases. *Bioorganic & Medicinal Chemistry Letters*, 20, 2103-2105.
- DREW, A. E., MORADEI, O., JACQUES, S. L., RIOUX, N., BORIACK-SJODIN, A. P., ALLAIN, C., SCOTT, M. P., JIN, L., RAIMONDI, A. & HANDLER, J. L. 2017. Identification of a CARM1 inhibitor with potent in vitro and in vivo activity in preclinical models of multiple myeloma. *Scientific Reports*, 7, 1-13.
- EL MESSAOUDI, S., FABBRIZIO, E., RODRIGUEZ, C., CHUCHANA, P., FAUQUIER, L., CHENG, D., THEILLET, C., VANDEL, L., BEDFORD, M. T. & SARDET, C. 2006. Coactivator-associated arginine methyltransferase 1 (CARM1) is a positive regulator of the Cyclin E1 gene. *Proceedings of the National Academy of Sciences*, 103, 13351-13356.
- ELAKOUM, R., GAUCHOTTE, G., OUSSALAH, A., WISSLER, M.-P., CLÉMENT-DUCHÊNE, C., VIGNAUD, J.-M., GUÉANT, J.-L. & NAMOUR, F. 2014. CARM1 and PRMT1 are dysregulated in lung cancer without hierarchical features. *Biochimie*, 97, 210-218.

- ELLOUL, S., SILINS, I., TROPÉ, C. G., BENSHUSHAN, A., DAVIDSON, B. & REICH, R. 2006. Expression of E-cadherin transcriptional regulators in ovarian carcinoma. *Virchows Archiv*, 449, 520-528.
- EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60, 2126-2132.
- ERAM, M. S., BUSTOS, S. P., LIMA-FERNANDES, E., SIARHEYEVA, A., SENISTERRA, G., HAJIAN, T., CHAU, I., DUAN, S., WU, H. & DOMBROVSKI, L. 2014. Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *Journal of Biological Chemistry*, 289, 12177-12188.
- ERAM, M. S., SHEN, Y., SZEWCZYK, M. M., WU, H., SENISTERRA, G., LI, F., BUTLER, K. V., KANISKAN, H. U. M., SPEED, B. A. & DELA SEÑA, C. 2016. A potent, selective, and cell-active inhibitor of human type I protein arginine methyltransferases. *ACS Chemical Biology*, 11, 772-781.
- ERLANSON, D. A., FESIK, S. W., HUBBARD, R. E., JAHNKE, W. & JHOTI, H. 2016. Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery*, 15, 605.
- ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R. & COYNE, M. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473, 43.
- ESPOSITO, D. & CHATTERJEE, D. K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*, 17, 353-358.
- ESSE, R., LEANDRO, P., RIVERA, I., DE ALMEIDA, I. T., BLOM, H. J. & CASTRO, R. 2012. Deciphering Protein Arginine Methylation in Mammals. In: DRICU, A. (ed.) *Methylation - From DNA, RNA and Histones to Diseases and Treatment*. Rijeka: IntechOpen.
- EVANKOVICH, J., LEAR, T., MCKELVEY, A., DUNN, S., LONDINO, J., LIU, Y., CHEN, B. B. & MALLAMPALLI, R. K. 2017. Receptor for advanced glycation end products is targeted by FBXO10 for ubiquitination and degradation. *The FASEB Journal*, 31, 3894-3903.
- FAUQUIER, L., DUBOÉ, C., JORÉ, C., TROUCHE, D. & VANDEL, L. 2008. Dual role of the arginine methyltransferase CARM1 in the regulation of c-Fos target genes. *The FASEB Journal*, 22, 3337-3347.
- FEDORIW, A., RAJAPURKAR, S. R., O'BRIEN, S., GERHART, S. V., MITCHELL, L. H., ADAMS, N. D., RIOUX, N., LINGARAJ, T., RIBICH, S. A. & PAPPALARDI, M. B. 2019. Anti-tumor activity of the type I PRMT inhibitor, GSK3368715, synergizes with PRMT5 inhibition through MTAP loss. *Cancer Cell*, 36, 100-114. e25.
- FENG, Q., HE, B., JUNG, S.-Y., SONG, Y., QIN, J., TSAI, S. Y., TSAI, M.-J. & O'MALLEY, B. W. 2009. Biochemical control of CARM1 enzymatic activity by phosphorylation. *Journal of Biological Chemistry*, 284, 36167-36174.
- FENG, Q., YI, P., WONG, J. & O'MALLEY, B. W. 2006. Signaling within a coactivator complex: methylation of SRC-3/AIB1 is a molecular switch for complex disassembly. *Molecular and Cellular Biology*, 26, 7846-7857.
- FENG, S., CHEN, J. K., YU, H., SIMON, J. A. & SCHREIBER, S. L. 1994. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*, 266, 1241-1247.
- FENG, Y., HADJIKYRIACOU, A. & CLARKE, S. G. 2014. Substrate specificity of human protein arginine methyltransferase 7 (PRMT7) the importance of acidic

- residues in the double E loop. *Journal of Biological Chemistry*, 289, 32604-32616.
- FENG, Y., LI, M., WANG, B. & ZHENG, Y. G. 2010. Discovery and mechanistic study of a class of protein arginine methylation inhibitors. *Journal of Medicinal Chemistry*, 53, 6028-6039.
- FENG, Y., MAITY, R., WHITELEGGE, J. P., HADJIKYRIACOU, A., LI, Z., ZURITA-LOPEZ, C., AL-HADID, Q., CLARK, A. T., BEDFORD, M. T. & MASSON, J.-Y. 2013. Mammalian protein arginine methyltransferase 7 (PRMT7) specifically targets RXR sites in lysine-and arginine-rich regions. *Journal of Biological Chemistry*, 288, 37010-37025.
- FENG, Y., XIE, N., JIN, M., STAHLEY, M. R., STIVERS, J. T. & ZHENG, Y. G. 2011. A transient kinetic analysis of PRMT1 catalysis. *Biochemistry*, 50, 7033-7044.
- FERREIRA DE FREITAS, R., ERAM, M. S., SZEWCZYK, M. M., STEUBER, H., SMIL, D., WU, H., LI, F., SENISTERRA, G., DONG, A. & BROWN, P. J. 2016. Discovery of a potent class I protein arginine methyltransferase fragment inhibitor. *Journal of Medicinal Chemistry*, 59, 1176-1183.
- FERREIRA DE FREITAS, R., IVANOCHKO, D. & SCHAPIRA, M. 2019. Methyltransferase Inhibitors: Competing with, or Exploiting the Bound Cofactor. *Molecules*, 24, 4492.
- FIELENBACH, N., GUARDAVACCARO, D., NEUBERT, K., CHAN, T., LI, D., FENG, Q., HUTTER, H., PAGANO, M. & ANTEBI, A. 2007. DRE-1: an evolutionarily conserved F box protein that regulates *C. elegans* developmental age. *Developmental Cell*, 12, 443-455.
- FISCHER, M. J. 2010. Amine Coupling through EDC/NHS: A Practical Approach. *Surface Plasmon Resonance*. London: Springer.
- FONTANA, A., DE LAURETO, P. P., SPOLAORE, B., FRARE, E., PICOTTI, P. & ZAMBONIN, M. 2004. Probing protein structure by limited proteolysis. *ACTA Acta Biochimica Polonica - English Edition-*, 51, 299-322.
- FONTANA, A., FASSINA, G., VITA, C., DALZOPPO, D., ZAMAI, M. & ZAMBONIN, M. 1986. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry*, 25, 1847-1851.
- FRANCESCHELLI, S., FERRONE, A., PESCE, M., RICCIONI, G. & SPERANZA, L. 2013. Biological functional relevance of asymmetric dimethylarginine (ADMA) in cardiovascular disease. *International Journal of Molecular Sciences*, 14, 24412-24421.
- FRANKEL, A., YADAV, N., LEE, J., BRANSCOMBE, T. L., CLARKE, S. & BEDFORD, M. T. 2002. The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity. *Journal of Biological Chemistry*, 277, 3537-3543.
- FRIESEN, W. J., MASSENET, S., PAUSHKIN, S., WYCE, A. & DREYFUSS, G. 2001. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Molecular Cell*, 7, 1111-1117.
- FRIETZE, S., LUPIEN, M., SILVER, P. A. & BROWN, M. 2008. CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. *Cancer Research*, 68, 301-306.
- FUHRMANN, J., CLANCY, K. W. & THOMPSON, P. R. 2015. Chemical biology of protein arginine modifications in epigenetic regulation. *Chemical Reviews*, 115, 5413-5461.

- FUHRMANN, J. & THOMPSON, P. R. 2016. Protein arginine methylation and citrullination in epigenetic regulation. *ACS Chemical Biology*, 11, 654-668.
- GAN, B., CHEN, S., LIU, H., MIN, J. & LIU, K. 2019. Structure and function of eTudor domain containing TDRD proteins. *Critical Reviews in Biochemistry and Molecular Biology*, 54, 119-132.
- GANESH, L., YOSHIMOTO, T., MOORTHY, N. C., AKAHATA, W., BOEHM, M., NABEL, E. G. & NABEL, G. J. 2006. Protein methyltransferase 2 inhibits NF- κ B function and promotes apoptosis. *Molecular and Cellular Biology*, 26, 3864-3874.
- GAO, W.-W., XIAO, R.-Q., PENG, B.-L., XU, H.-T., SHEN, H.-F., HUANG, M.-F., SHI, T.-T., YI, J., ZHANG, W.-J. & WU, X.-N. 2015. Arginine methylation of HSP70 regulates retinoid acid-mediated RAR β 2 gene activation. *Proceedings of the National Academy of Sciences*, 112, E3327-E3336.
- GAO, Y.-G., YAN, X.-Z., SONG, A.-X., CHANG, Y.-G., GAO, X.-C., JIANG, N., ZHANG, Q. & HU, H.-Y. 2006. Structural insights into the specific binding of huntingtin proline-rich region with the SH3 and WW domains. *Structure*, 14, 1755-1765.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: WALKER, J. M. (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press.
- GAYATRI, S. & BEDFORD, M. T. 2014. Readers of histone methylarginine marks. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1839, 702-710.
- GENG, P., ZHANG, Y., LIU, X., ZHANG, N., LIU, Y., LIU, X., LIN, C., YAN, X., LI, Z. & WANG, G. 2017. Automethylation of protein arginine methyltransferase 7 and its impact on breast cancer progression. *The FASEB Journal*, 31, 2287-2300.
- GERHART, S. V., KELLNER, W. A., THOMPSON, C., PAPPALARDI, M. B., ZHANG, X.-P., DE OCA, R. M., PENEBRE, E., DUNCAN, K., BORIACK-SJODIN, A. & LE, B. 2018. Activation of the p53-MDM4 regulatory axis defines the anti-tumour response to PRMT5 inhibition through its role in regulating cellular splicing. *Scientific Reports*, 8, 1-15.
- GOUET, P., COURCELLE, E., STUART, D. I. & M[√]© TOZ, F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, 15, 305-308.
- GOULET, I., GAUVIN, G., BOISVENUE, S. & CÔTÉ, J. 2007. Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization. *Journal of Biological Chemistry*, 282, 33009-33021.
- GREER, E. L. & SHI, Y. 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, 13, 343-357.
- GROS, L., RENODON-CORNIÈRE, A., DE SAINT VINCENT, B. R., FEDER, M., BUJNICKI, J. M. & JACQUEMIN-SABLON, A. 2006. Characterization of prmt7 α and β isozymes from Chinese hamster cells sensitive and resistant to topoisomerase II inhibitors. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1760, 1646-1656.
- GUCCIONE, E. & RICHARD, S. 2019. The regulation, functions and clinical relevance of arginine methylation. *Nat. Rev. Mol. Cell Biol*, 20, 642-657.
- GUITOT, K., DRUJON, T., BURLINA, F., SAGAN, S., BEAUPIERRE, S., PAMLARD, O., DODD, R. H., GUILLOU, C., BOLBACH, G. & SACHON, E. 2017. A direct label-free MALDI-TOF mass spectrometry based assay for the characterization

- of inhibitors of protein lysine methyltransferases. *Analytical and Bioanalytical Chemistry*, 409, 3767-3777.
- GUITOT, K., SCARABELLI, S., DRUJON, T., BOLBACH, G., AMOURA, M., BURLINA, F., JELTSCH, A., SAGAN, S. & GUIANVARC'H, D. 2014. Label-free measurement of histone lysine methyltransferases activity by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical Biochemistry*, 456, 25-31.
- GUNNELL, E. A., AL-NOORI, A., MUHSEN, U., DAVIES, C. C., DOWDEN, J. & DREVENY, I. 2020. Structural and biochemical evaluation of bisubstrate inhibitors of protein arginine N-methyltransferases PRMT1 and CARM1 (PRMT4). *Biochemical Journal*, 477, 787-800.
- HAJDUK, P. J. & GREER, J. 2007. A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug discovery*, 6, 211-219.
- HALBY, L., MARECHAL, N., PECHALRIEU, D., CURA, V., FRANCHINI, D.-M., FAUX, C., ALBY, F., TROFFER-CHARLIER, N., KUDITHIPUDI, S. & JELTSCH, A. 2018. Hijacking DNA methyltransferase transition state analogues to produce chemical scaffolds for PRMT inhibitors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170072.
- HANUKOGLU, I. 2015. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education*, 43, 206-209.
- HARD, R., LI, N., HE, W., ROSS, B., MO, G. C., PENG, Q., STEIN, R. S., KOMIVES, E., WANG, Y. & ZHANG, J. 2018. Deciphering and engineering chromodomain-methyllysine peptide recognition. *Science Advances*, 4, eaau1447.
- HATANAKA, Y., TSUSAKA, T., SHIMIZU, N., MORITA, K., SUZUKI, T., MACHIDA, S., SATOH, M., HONDA, A., HIROSE, M. & KAMIMURA, S. 2017. Histone H3 methylated at arginine 17 is essential for reprogramming the paternal genome in zygotes. *Cell Reports*, 20, 2756-2765.
- HEAP, R. E., HOPE, A. G., PEARSON, L.-A., REYSKENS, K. M., MCELROY, S. P., HASTIE, C. J., PORTER, D. W., ARTHUR, J. S. C., GRAY, D. W. & TROST, M. 2017. Identifying inhibitors of inflammation: a novel high-throughput MALDI-TOF screening assay for salt-inducible kinases (SIKs). *SLAS DISCOVERY: Advancing Life Sciences R&D*, 22, 1193-1202.
- HERRMANN, F. & FACKELMAYER, F. O. 2009. Nucleo-cytoplasmic shuttling of protein arginine methyltransferase 1 (PRMT1) requires enzymatic activity. *Genes to Cells*, 14, 309-317.
- HERRMANN, F., LEE, J., BEDFORD, M. T. & FACKELMAYER, F. O. 2005. Dynamics of human protein arginine methyltransferase 1 (PRMT1) in vivo. *Journal of Biological Chemistry*, 280, 38005-38010.
- HERZ, H.-M., GARRUSS, A. & SHILATIFARD, A. 2013. SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends in Biochemical Sciences*, 38, 621-639.
- HEVEL, J. M. & PRICE, O. M. 2020. Rapid and direct measurement of methyltransferase activity in about 30 min. *Methods*, 175, 3-9.
- HIDAKA, Y., HAGIWARA, T. & YAMADA, M. 2005. Methylation of the guanidino group of arginine residues prevents citrullination by peptidylarginine deiminase IV. *FEBS Letters*, 579, 4088-4092.
- HIGASHIMOTO, K., KUHN, P., DESAI, D., CHENG, X. & XU, W. 2007. Phosphorylation-mediated inactivation of coactivator-associated arginine

- methyltransferase 1. *Proceedings of the National Academy of Sciences*, 104, 12318-12323.
- HO, M.-C., WILCZEK, C., BONANNO, J. B., XING, L., SEZNEC, J., MATSUI, T., CARTER, L. G., ONIKUBO, T., KUMAR, P. R. & CHAN, M. K. 2013. Structure of the arginine methyltransferase PRMT5-MEP50 reveals a mechanism for substrate specificity. *PLoS One*, 8, 1-16.
- HOELZ, A., JANZ, J. M., LAWRIE, S. D., CORWIN, B., LEE, A. & SAKMAR, T. P. 2006. Crystal structure of the SH3 domain of β PIX in complex with a high affinity peptide from PAK2. *Journal of Molecular Biology*, 358, 509-522.
- HOLBERT, M. A. & MARMORSTEIN, R. 2005. Structure and activity of enzymes that remove histone modifications. *Current Opinion in Structural Biology*, 15, 673-680.
- HOLCOMB, J., SPELLMON, N., ZHANG, Y., DOUGHAN, M., LI, C. & YANG, Z. 2017. Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophysics*, 4, 557.
- HOPKINS, A. L., GROOM, C. R. & ALEX, A. 2004. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today*, 9, 430-431.
- HORN, P. J. & PETERSON, C. L. 2002. Chromatin higher order folding--wrapping up transcription. *Science*, 297, 1824-1827.
- HORWICH, A. L., FARR, G. W. & FENTON, W. A. 2006. GroEL– GroES-mediated protein folding. *Chemical Reviews*, 106, 1917-1930.
- HOU, W., NEMITZ, S., SCHOPPER, S., NIELSEN, M. L., KESSELS, M. M. & QUALMANN, B. 2018. Arginine methylation by PRMT2 controls the functions of the actin nucleator Cobl. *Developmental Cell*, 45, 262-275. e8.
- HSIAO, K., ZEGZOUTI, H. & GOUELI, S. A. 2016. Methyltransferase-Glo: a universal, bioluminescent and homogenous assay for monitoring all classes of methyltransferases. *Epigenomics*, 8, 321-339.
- HU, H., QIAN, K., HO, M.-C. & ZHENG, Y. G. 2016. Small molecule inhibitors of protein arginine methyltransferases. *Expert Opinion on Investigational Drugs*, 25, 335-358.
- HYLLUS, D., STEIN, C., SCHNABEL, K., SCHILTZ, E., IMHOF, A., DOU, Y., HSIEH, J. & BAUER, U.-M. 2007. PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation. *Genes & Development*, 21, 3369-3380.
- IBERG, A. N., ESPEJO, A., CHENG, D., KIM, D., MICHAUD-LEVESQUE, J., RICHARD, S. & BEDFORD, M. T. 2008. Arginine methylation of the histone H3 tail impedes effector binding. *Journal of Biological Chemistry*, 283, 3006-3010.
- ITALIANO, A., SORIA, J.-C., TOULMONDE, M., MICHOT, J.-M., LUCCHESI, C., VARGA, A., COINDRE, J.-M., BLAKEMORE, S. J., CLAWSON, A. & SUTTLE, B. 2018. Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. *The Lancet Oncology*, 19, 649-659.
- IWASAKI, H., KOVACIC, J. C., OLIVE, M., BEERS, J. K., YOSHIMOTO, T., CROOK, M. F., TONELLI, L. H. & NABEL, E. G. 2010. Disruption of protein arginine N-methyltransferase 2 regulates leptin signaling and produces leanness in vivo through loss of STAT3 methylation. *Circulation Research*, 107, 992-1001.
- IWASAKI, W., MIYA, Y., HORIKOSHI, N., OSAKABE, A., TAGUCHI, H., TACHIWANA, H., SHIBATA, T., KAGAWA, W. & KURUMIZAKA, H. 2013. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio*, 3, 363-369.

- JACOBS, S. A., HARP, J. M., DEVARAKONDA, S., KIM, Y., RASTINEJAD, F. & KHORASANIZADEH, S. 2002. The active site of the SET domain is constructed on a knot. *Nature Structural Biology*, 9, 833-838.
- JACQUES, S. L., AQUINO, K. P., GUREASKO, J., BORIACK-SJODIN, P. A., PORTER SCOTT, M., COPELAND, R. A. & RIERA, T. V. 2016. CARM1 preferentially methylates H3R17 over H3R26 through a random kinetic mechanism. *Biochemistry*, 55, 1635-1644.
- JAIN, K. & CLARKE, S. G. 2019. PRMT7 as a unique member of the protein arginine methyltransferase family: A review. *Archives of Biochemistry and Biophysics*, 2019, 36-45.
- JANECKI, D. M., SAJEK, M., SMIALEK, M. J., KOTECKI, M., GINTER-MATUSZEWSKA, B., KUCZYNSKA, B., SPIK, A., KOLANOWSKI, T., KITAZAWA, R. & KURPISZ, M. 2018. SPIN1 is a proto-oncogene and SPIN3 is a tumor suppressor in human seminoma. *Oncotarget*, 9, 32466.
- JANSON, G., ZHANG, C., PRADO, M. G. & PAIARDINI, A. 2016. PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics*, 33, 444-446.
- JARROLD, J. & DAVIES, C. C. 2019. PRMTs and Arginine Methylation: Cancer's Best-Kept Secret? *Trends in Molecular Medicine*, 25, 993-1009.
- JASON-MOLLER, L., MURPHY, M. & BRUNO, J. 2006. Overview of Biacore systems and their applications. *Current Protocols in Protein Science*, 45, 19.13. 1-19.13. 14.
- JIN, J., XIE, X., CHEN, C., PARK, J. G., STARK, C., JAMES, D. A., OLHOVSKY, M., LINDING, R., MAO, Y. & PAWSON, T. 2009. Eukaryotic protein domains as functional units of cellular evolution. *Science Signaling*, 2, ra76-ra76.
- KABSCH, W. 2010. Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66, 125-132.
- KANEKO, T., LI, L. & LI, S. 2008. The SH3 domain—a family of versatile peptide-and protein-recognition module. *Frontiers in Bioscience*, 13, 4938-4952.
- KANG, H., FREUND, C., DUKE-COHAN, J. S., MUSACCHIO, A., WAGNER, G. & RUDD, C. E. 2000. SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55. *The EMBO Journal*, 19, 2889-2899.
- KANG, M., MEHRAZARIN, S., PARK, N. H. & WANG, C. Y. 2017. Epigenetic gene regulation by histone demethylases: emerging role in oncogenesis and inflammation. *Oral Diseases*, 23, 709-720.
- KÄRKKÄINEN, S., HIIPAKKA, M., WANG, J. H., KLEINO, I., VÄHÄ-JAAKKOLA, M., RENKEMA, G. H., LISS, M., WAGNER, R. & SAKSELA, K. 2006. Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. *EMBO Reports*, 7, 186-191.
- KATSANIS, N., YASPO, M.-L. & FISHER, E. M. 1997. Identification and mapping of a novel human gene, HRMT1L1, homologous to the rat protein arginine N-methyltransferase 1 (PRMT1) gene. *Mammalian Genome*, 8, 526-529.
- KEARNEY, P. L., BHATIA, M., JONES, N. G., YUAN, L., GLASCOCK, M. C., CATCHINGS, K. L., YAMADA, M. & THOMPSON, P. R. 2005. Kinetic characterization of protein arginine deiminase 4: a transcriptional corepressor implicated in the onset and progression of rheumatoid arthritis. *Biochemistry*, 44, 10570-10582.

- KEBEDE, A. F., SCHNEIDER, R. & DAUJAT, S. 2015. Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *The FEBS Journal*, 282, 1658-1674.
- KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845.
- KIM, D., LEE, J., CHENG, D., LI, J., CARTER, C., RICHIE, E. & BEDFORD, M. T. 2010a. Enzymatic activity is required for the in vivo functions of CARM1. *Journal of Biological Chemistry*, 285, 1147-1152.
- KIM, D., LIM, S., PARK, M., CHOI, J., KIM, J., HAN, H., YOON, K., KIM, K., LIM, J. & PARK, S. 2014. Ubiquitination-dependent CARM1 degradation facilitates Notch1-mediated podocyte apoptosis in diabetic nephropathy. *Cellular Signalling*, 26, 1774-1782.
- KIM, Y.-R., LEE, B. K., PARK, R.-Y., NGUYEN, N. T. X., BAE, J. A., KWON, D. D. & JUNG, C. 2010b. Differential CARM1 expression in prostate and colorectal cancers. *BMC Cancer*, 10, 197.
- KLOSE, R. J., GARDNER, K. E., LIANG, G., ERDJUMENT-BROMAGE, H., TEMPST, P. & ZHANG, Y. 2007. Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in *Saccharomyces cerevisiae*? *Molecular and Cellular Biology*, 27, 3951-3961.
- KNUTSON, S. K., KAWANO, S., MINOSHIMA, Y., WARHOLIC, N. M., HUANG, K.-C., XIAO, Y., KADOWAKI, T., UESUGI, M., KUZNETSOV, G. & KUMAR, N. 2014. Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Molecular Cancer Therapeutics*, 13, 842-854.
- KOCH, C. A., ANDERSON, D., MORAN, M. F., ELLIS, C. & PAWSON, T. 1991. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science*, 252, 668-674.
- KOH, S. S., LI, H., LEE, Y.-H., WIDELITZ, R. B., CHUONG, C.-M. & STALLCUP, M. R. 2002. Synergistic coactivator function by coactivator-associated arginine methyltransferase (CARM) 1 and β -catenin with two different classes of DNA-binding transcriptional activators. *Journal of Biological Chemistry*, 277, 26031-26035.
- KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell*, 128, 693-705.
- KRAUSE, C. D., YANG, Z.-H., KIM, Y.-S., LEE, J.-H., COOK, J. R. & PESTKA, S. 2007. Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacology & Therapeutics*, 113, 50-87.
- KUHN, P., CHUMANOV, R., WANG, Y., GE, Y., BURGESS, R. R. & XU, W. 2010. Automethylation of CARM1 allows coupling of transcription and mRNA splicing. *Nucleic Acids Research*, 39, 2717-2726.
- KUHN, P., CHUMANOV, R., WANG, Y., GE, Y., BURGESS, R. R. & XU, W. 2011. Automethylation of CARM1 allows coupling of transcription and mRNA splicing. *Nucleic acids research*, 39, 2717-2726.
- KUNTZ, I., CHEN, K., SHARP, K. & KOLLMAN, P. 1999. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences*, 96, 9997-10002.
- KUROCHKINA, N. & GUHA, U. 2013. SH3 domains: modules of protein-protein interactions. *Biophysical Reviews*, 5, 29-39.

- KZHYSHKOWSKA, J., SCHÜTT, H., MICHAEL, L., KREMMER, E., STAUBER, R., HANS, W. & DOBNER, T. 2001. Heterogeneous nuclear ribonucleoprotein E1B-AP5 is methylated in its Arg-Gly-Gly (RGG) box and interacts with human arginine methyltransferase HRMT1L1. *Biochemical Journal*, 358, 305-314.
- 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 & Development*, 13, 740-753.
- LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680.
- LAI, Y., LI, J., LI, X. & ZOU, C. 2017. Lipopolysaccharide modulates p300 and Sirt1 to promote PRMT1 stability via an SCFFbx17-recognized acetyldegron. *Journal of Cell Science*, 130, 3578-3587.
- LAKOWSKI, T. & FRANKEL, A. 2009. Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl-and asymmetric dimethyl-arginine residues on histone H4. *Biochemical Journal*, 421, 253.
- LAKOWSKI, T. M. & FRANKEL, A. 2010. Sources of S-adenosyl-L-homocysteine background in measuring protein arginine N-methyltransferase activity using tandem mass spectrometry. *Analytical Biochemistry*, 396, 158-160.
- LAMOREE, B. & HUBBARD, R. E. 2017. Current perspectives in fragment-based lead discovery (FBLD). *Essays in Biochemistry*, 61, 453.
- LAU, D. H., HOGSETH, M., PHILLIPS, E. C., O'NEILL, M. J., POOLER, A. M., NOBLE, W. & HANGER, D. P. 2016. Critical residues involved in tau binding to fyn: implications for tau phosphorylation in Alzheimer's disease. *Acta Neuropathologica Communications*, 4, 49.
- LAWRENCE, M., DAUJAT, S. & SCHNEIDER, R. 2016. Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*, 32, 42-56.
- LEE, G. 2005. Tau and src family tyrosine kinases. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1739, 323-330.
- LEE, J.-H., COOK, J. R., YANG, Z.-H., MIROCHNITCHENKO, O., GUNDERSON, S. I., FELIX, A. M., HERTH, N., HOFFMANN, R. & PESTKA, S. 2005. PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *Journal of Biological Chemistry*, 280, 3656-3664.
- LEE, J. & BEDFORD, M. T. 2002. PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Reports*, 3, 268-273.
- LEE, Y.-H., KOH, S. S., ZHANG, X., CHENG, X. & STALLCUP, M. R. 2002. Synergy among nuclear receptor coactivators: selective requirement for protein methyltransferase and acetyltransferase activities. *Molecular and Cellular Biology*, 22, 3621-3632.
- LEMMON, M. A. Pleckstrin homology (PH) domains and phosphoinositides. *Biochemical Society Symposia*, 2007. Portland Press Ltd., 81-93.
- LEMMON, M. A., FERGUSON, K. M. & SCHLESSINGER, J. 1996. PH domains: diverse sequences with a common fold recruit signaling molecules to the cell surface. *Cell*, 85, 621-624.
- LEVINSON, N. M., VISPERAS, P. R. & KURIYAN, J. 2009. The tyrosine kinase Csk dimerizes through its SH3 domain. *PLoS One*, 4, 1-5.
- LI, X., HU, X., PATEL, B., ZHOU, Z., LIANG, S., YBARRA, R., QIU, Y., FELSENFELD, G., BUNGERT, J. & HUANG, S. 2010. H4R3 methylation facilitates β -globin

- transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood, The Journal of the American Society of Hematology*, 115, 2028-2037.
- LI, X., LAI, Y., LI, J., ZOU, M. & ZOU, C. 2017. Oxidative stress destabilizes protein arginine methyltransferase 4 via glycogen synthase kinase 3 β to impede lung epithelial cell migration. *American Journal of Physiology-Cell Physiology*, 313, C285-C294.
- LI, X., WANG, C., JIANG, H. & LUO, C. 2019. A patent review of arginine methyltransferase inhibitors (2010–2018). *Expert Opinion on Therapeutic Patents*, 29, 97-114.
- LIM, W. A., RICHARDS, F. M. & FOX, R. O. 1994. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*, 372, 375-379.
- LIN, H. & LUENGO, J. I. 2019. Nucleoside protein arginine methyltransferase 5 (PRMT5) inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 29, 1264-1269.
- LIPINSKI, C. A. 2004. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies*, 1, 337-341.
- LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23, 3-25.
- LIU, F., LI, F., MA, A., DOBROVETSKY, E., DONG, A., GAO, C., KORBOUKH, I., LIU, J., SMIL, D. & BROWN, P. J. 2013. Exploiting an allosteric binding site of PRMT3 yields potent and selective inhibitors. *Journal of Medicinal Chemistry*, 56, 2110-2124.
- LIU, K., CHEN, C., GUO, Y., LAM, R., BIAN, C., XU, C., ZHAO, D. Y., JIN, J., MACKENZIE, F. & PAWSON, T. 2010. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proceedings of the National Academy of Sciences*, 107, 18398-18403.
- LIU, K., GUO, Y., LIU, H., BIAN, C., LAM, R., LIU, Y., MACKENZIE, F., ROJAS, L. A., REINBERG, D. & BEDFORD, M. T. 2012a. Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One*, 7.
- LIU, K., GUO, Y., LIU, H., BIAN, C., LAM, R., LIU, Y., MACKENZIE, F., ROJAS, L. A., REINBERG, D. & BEDFORD, M. T. 2012b. Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One*, 7, 1-8.
- LIU, Q., BERRY, D., NASH, P., PAWSON, T., MCGLADE, C. J. & LI, S. S.-C. 2003. Structural basis for specific binding of the Gads SH3 domain to an RxxK motif-containing SLP-76 peptide: a novel mode of peptide recognition. *Molecular Cell*, 11, 471-481.
- LONG, J. C. & CACERES, J. F. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, 417, 15-27.
- LORTON, B. M. & SHECHTER, D. 2019. Cellular consequences of arginine methylation. *Cellular and Molecular Life Sciences*, 1-24.
- LU, Y.-F., CAI, X.-L., LI, Z.-Z., LV, J., XIANG, Y.-A., CHEN, J.-J., CHEN, W.-J., SUN, W.-Y., LIU, X.-M. & CHEN, J.-B. 2018. LncRNA SNHG16 functions as an oncogene by sponging MiR-4518 and up-regulating PRMT5 expression in glioma. *Cellular Physiology and Biochemistry*, 45, 1975-1985.

- LUGER, K., MÄDER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389, 251-260.
- LUO, M. 2012. Current chemical biology approaches to interrogate protein methyltransferases. *ACS Chemical Biology*, 7, 443-463.
- LUO, M. 2015. Inhibitors of protein methyltransferases as chemical tools. *Epigenomics*, 7, 1327-1338.
- MAI, A., CHENG, D., BEDFORD, M. T., VALENTE, S., NEBBIOSO, A., PERRONE, A., BROSCHE, G., SBARDELLA, G., DE BELLIS, F. & MICELI, M. 2008. Epigenetic multiple ligands: mixed histone/protein methyltransferase, acetyltransferase, and class III deacetylase (sirtuin) inhibitors. *Journal of Medicinal Chemistry*, 51, 2279-2290.
- MAI, A., VALENTE, S., CHENG, D., PERRONE, A., RAGNO, R., SIMEONI, S., SBARDELLA, G., BROSCHE, G., NEBBIOSO, A. & CONTE, M. 2007. Synthesis and biological validation of novel synthetic histone/protein methyltransferase inhibitors. *ChemMedChem: Chemistry Enabling Drug Discovery*, 2, 987-991.
- MANNHOLD, R., KUBINYI, H. & FOLKERS, G. 2015. *Fragment-based drug discovery: lessons and outlook*, John Wiley & Sons.
- MAO, R., SHAO, J., ZHU, K., ZHANG, Y., DING, H., ZHANG, C., SHI, Z., JIANG, H., SUN, D. & DUAN, W. 2017. Potent, selective, and cell active protein arginine methyltransferase 5 (PRMT5) inhibitor developed by structure-based virtual screening and hit optimization. *Journal of Medicinal Chemistry*, 60, 6289-6304.
- MARTIN, C. & ZHANG, Y. 2005. The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology*, 6, 838-849.
- MARTIN, J. L. & MCMILLAN, F. M. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology*, 12, 783-793.
- MAURER-STROH, S., DICKENS, N. J., HUGHES-DAVIES, L., KOUZARIDES, T., EISENHABER, F. & PONTING, C. P. 2003. The Tudor domain 'Royal Family': tudor, plant agenet, chromo, PWWP and MBT domains. *Trends in Biochemical Sciences*, 28, 69-74.
- MAYER, B. J. 2001. SH3 domains: complexity in moderation. *Journal of Cell Science*, 114, 1253-1263.
- MAYER, B. J., HAMAGUCHI, M. & HANAFUSA, H. 1988. A novel viral oncogene with structural similarity to phospholipase C. *Nature*, 332, 272.
- MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658-674.
- MCGUFFIN, L. J., BRYSON, K. & JONES, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 16, 404-405.
- MCNICHOLAS, S., POTTERTON, E., WILSON, K. & NOBLE, M. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, 67, 386-394.
- MEANWELL, N. A. 2015. *Tactics in Contemporary Drug Design*, Heidelberg, Springer Verlag.
- MERTZ, J. A., CONERY, A. R., BRYANT, B. M., SANDY, P., BALASUBRAMANIAN, S., MELE, D. A., BERGERON, L. & SIMS, R. J. 2011. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proceedings of the National Academy of Sciences*, 108, 16669-16674.

- MEYER, R., WOLF, S. S. & OBENDORF, M. 2007. PRMT2, a member of the protein arginine methyltransferase family, is a coactivator of the androgen receptor. *The Journal of Steroid Biochemistry and Molecular Biology*, 107, 1-14.
- MIAO, F., LI, S., CHAVEZ, V., LANTING, L. & NATARAJAN, R. 2006. Coactivator-associated arginine methyltransferase-1 enhances nuclear factor- κ B-mediated gene transcription through methylation of histone H3 at arginine 17. *Molecular Endocrinology*, 20, 1562-1573.
- MIGLIORI, V., MAPELLI, M. & GUCCIONE, E. 2012. On WD40 proteins: propelling our knowledge of transcriptional control? *Epigenetics*, 7, 815-822.
- MIN, J., FENG, Q., LI, Z., ZHANG, Y. & XU, R.-M. 2003. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell*, 112, 711-723.
- MIRANDA, T. B., MIRANDA, M., FRANKEL, A. & CLARKE, S. 2004. PRMT7 is a member of the protein arginine methyltransferase family with a distinct substrate specificity. *Journal of Biological Chemistry*, 279, 22902-22907.
- MITCHELL, L. H., DREW, A. E., RIBICH, S. A., RIOUX, N., SWINGER, K. K., JACQUES, S. L., LINGARAJ, T., BORIACK-SJODIN, P. A., WATERS, N. J. & WIGLE, T. J. 2015. Aryl pyrazoles as potent inhibitors of arginine methyltransferases: identification of the first PRMT6 tool compound. *ACS Medicinal Chemistry Letters*, 6, 655-659.
- MOHAMMAD, H. P., BARBASH, O. & CREASY, C. L. 2019. Targeting epigenetic modifications in cancer therapy: erasing the roadmap to cancer. *Nature Medicine*, 25, 403-418.
- MONGIOVÍ, A. M., ROMANO, P. R., PANNI, S., MENDOZA, M., WONG, W. T., MUSACCHIO, A., CESARENI, G. & DI FIORE, P. P. 1999. A novel peptide-SH3 interaction. *The EMBO Journal*, 18, 5300-5309.
- MORALES, Y., CÁCERES, T., MAY, K. & HEVEL, J. M. 2016. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Archives of Biochemistry and Biophysics*, 590, 138-152.
- MORETTIN, A., BALDWIN, R. M. & CÔTÉ, J. 2015. Arginine methyltransferases as novel therapeutic targets for breast cancer. *Mutagenesis*, 30, 177-189.
- MORRIS, J. 2001. Genes, genetics, and epigenetics: a correspondence. *Science*, 293, 1103-1105.
- MURSHUDOV, G. N., SKUBÁK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67, 355-367.
- MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography*, 53, 240-255.
- MYSZKA, D. G. 1997. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Current Opinion in Biotechnology*, 8, 50-57.
- NAKAYAMA, K., SZEWCZYK, M. M., DELA SENA, C., WU, H., DONG, A., ZENG, H., LI, F., DE FREITAS, R. F., ERAM, M. S. & SCHAPIRA, M. 2018. TP-064, a potent and selective small molecule inhibitor of PRMT4 for multiple myeloma. *Oncotarget*, 9, 18480.
- NAZARÉ, M., MATTER, H., WILL, D. W., WAGNER, M., URMANN, M., CZECH, J., SCHREUDER, H., BAUER, A., RITTER, K. & WEHNER, V. 2012. Fragment deconstruction of small, potent factor Xa inhibitors: exploring the superadditivity

- energetics of fragment linking in protein–ligand complexes. *Angewandte Chemie International Edition*, 51, 905-911.
- NIE, M., WANG, Y., GUO, C., LI, X., WANG, Y., DENG, Y., YAO, B., GUI, T., MA, C. & LIU, M. 2018. CARM1-mediated methylation of protein arginine methyltransferase 5 represses human γ -globin gene expression in erythroleukemia cells. *Journal of Biological Chemistry*, 293, 17454-17463.
- NISHIDA, M., NAGATA, K., HACHIMORI, Y., HORIUCHI, M., OGURA, K., MANDIYAN, V., SCHLESSINGER, J. & INAGAKI, F. 2001. Novel recognition mode between Vav and Grb2 SH3 domains. *The EMBO Journal*, 20, 2995-3007.
- NOBLE, M., MUSACCHIO, A., SARASTE, M., COURTNEIDGE, S. A. & WIERENGA, R. K. 1993. Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *The EMBO Journal*, 12, 2617-2624.
- OBIANYO, O., OSBORNE, T. C. & THOMPSON, P. R. 2008. Kinetic mechanism of protein arginine methyltransferase 1. *Biochemistry*, 47, 10420-10427.
- OBIANYO, O. & THOMPSON, P. R. 2012. Kinetic mechanism of protein arginine methyltransferase 6 (PRMT6). *Journal of Biological Chemistry*, 287, 6062-6071.
- OLSEN, I., SINGHRAO, S. K. & POTEPA, J. 2018. Citrullination as a plausible link to periodontitis, rheumatoid arthritis, atherosclerosis and Alzheimer's disease. *Journal of Oral Microbiology*, 10, 1487742.
- OSBORNE, T. C., OBIANYO, O., ZHANG, X., CHENG, X. & THOMPSON, P. R. 2007. Protein arginine methyltransferase 1: positively charged residues in substrate peptides distal to the site of methylation are important for substrate binding and catalysis. *Biochemistry*, 46, 13370-13381.
- PAK, M. L., LAKOWSKI, T. M., THOMAS, D., VHUIYAN, M. I., HÜSECKEN, K. & FRANKEL, A. 2011. A protein arginine N-methyltransferase 1 (PRMT1) and 2 heteromeric interaction increases PRMT1 enzymatic activity. *Biochemistry*, 50, 8226-8240.
- PAWSON, T. 1995. Protein modules and signalling networks. *Nature*, 373, 573.
- PFISTER, S. X. & ASHWORTH, A. 2017. Marked for death: targeting epigenetic changes in cancer. *Nature Reviews Drug Discovery*, 16, 241.
- PISABARRO, M. T., SERRANO, L. & WILMANN, M. 1998. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *Journal of Molecular Biology*, 281, 513-521.
- POTTERTON, L., AGIRRE, J., BALLARD, C., COWTAN, K., DODSON, E., EVANS, P. R., JENKINS, H. T., KEEGAN, R., KRISINEL, E. & STEVENSON, K. 2018. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Structural Biology*, 74, 68-84.
- POULARD, C., CORBO, L. & LE ROMANCER, M. 2016. Protein arginine methylation/demethylation and cancer. *Oncotarget*, 7, 67532.
- POULARD, C., RAMBAUD, J., HUSSEIN, N., CORBO, L. & LE ROMANCER, M. 2014. JMJD6 regulates ER α methylation on arginine. *PLoS One*, 9, 1-9.
- PRISM, G. 2003. GraphPad Prism Version 4.00 for Windows. *GraphPad Software, San Diego: CA, USA*.
- PURANDARE, A. V., CHEN, Z., HUYNH, T., PANG, S., GENG, J., VACCARO, W., POSS, M. A., OCONNELL, J., NOWAK, K. & JAYARAMAN, L. 2008. Pyrazole

- inhibitors of coactivator associated arginine methyltransferase 1 (CARM1). *Bioorganic & Medicinal Chemistry Letters*, 18, 4438-4441.
- QI, C., CHANG, J., ZHU, Y., YELDANDI, A. V., RAO, S. M. & ZHU, Y.-J. 2002. Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor α . *Journal of Biological Chemistry*, 277, 28624-28630.
- RAIJMAKERS, R., ZENDMAN, A. J., EGBERTS, W. V., VOSSENAAR, E. R., RAATS, J., SOEDE-HUIJBREGTS, C., RUTJES, F. P., VAN VEELLEN, P. A., DRIJFHOUT, J. W. & PRUIJN, G. J. 2007. Methylation of arginine residues interferes with citrullination by peptidylarginine deiminases in vitro. *Journal of Molecular Biology*, 367, 1118-1129.
- RAKOW, S., PULLAMSETTI, S. S., BAUER, U.-M. & BOUCHARD, C. 2020. Assaying epigenome functions of PRMTs and their substrates. *Methods*, 175, 53-65.
- RATHERT, P., CHENG, X. & JELTSCH, A. 2007. Continuous enzymatic assay for histone lysine methyltransferases. *Biotechniques*, 43, 602-608.
- READ, R. J. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica Section A: Foundations of Crystallography*, 42, 140-149.
- REES, D. 2016. Fragment-based Drug Discovery: Lessons and Outlook. Edited by Daniel A. Erlanson and Wolfgang Jahnke; Series Editors: Raimund Mannhold, Hugo Kubinyi, and Gerd Folkers. *ChemMedChem*, 11, 1667-1667.
- REYNOLDS, C. H., BEMBENEK, S. D. & TOUNGE, B. A. 2007. The role of molecular size in ligand efficiency. *Bioorganic & Medicinal Chemistry Letters*, 17, 4258-4261.
- RHEIN, V. F., CARROLL, J., DING, S., FEARNLEY, I. M. & WALKER, J. E. 2013. NDUFAF7 methylates arginine 85 in the NDUF52 subunit of human complex I. *Journal of Biological Chemistry*, 288, 33016-33026.
- RITORTO, M. S., EWAN, R., PEREZ-OLIVA, A. B., KNEBEL, A., BUHRLAGE, S. J., WIGHTMAN, M., KELLY, S. M., WOOD, N. T., VIRDEE, S. & GRAY, N. S. 2014. Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nature Communications*, 5, 1-11.
- ROBIN-LESPINASSE, Y., SENTIS, S., KOLYTCHEFF, C., ROSTAN, M.-C., CORBO, L. & LE ROMANCER, M. 2007. hCAF1, a new regulator of PRMT1-dependent arginine methylation. *Journal of Cell Science*, 120, 638-647.
- ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172.
- ROUX, K. J., KIM, D. I. & BURKE, B. 2013. BioID: a screen for protein-protein interactions. *Current Protocols in Protein Science*, 74, 19.23. 1-19.23. 14.
- RUSH, T., ROTH, J. R., THOMPSON, S. J., ALDAHER, A. R., COCHRAN, J. N. & ROBERSON, E. D. 2020. A peptide inhibitor of Tau-SH3 interactions ameliorates amyloid- β toxicity. *Neurobiology of Disease*, 134, 104668.
- RUST, H. L., ZURITA-LOPEZ, C. I., CLARKE, S. & THOMPSON, P. R. 2011. Mechanistic studies on transcriptional coactivator protein arginine methyltransferase 1. *Biochemistry*, 50, 3332-3345.
- SABRA, M., TEXIER, P., EL MAALOUF, J. & LOMONTE, P. 2013. The Tudor protein survival motor neuron (SMN) is a chromatin-binding protein that interacts with methylated lysine 79 of histone H3. *Journal of Cell Science*, 126, 3664-3677.
- SACK, J. S., THIEFFINE, S., BANDIERA, T., FASOLINI, M., DUKE, G. J., JAYARAMAN, L., KISH, K. F., KLEI, H. E., PURANDARE, A. V. & ROSETTANI,

- P. 2011. Structural basis for CARM1 inhibition by indole and pyrazole inhibitors. *Biochemical Journal*, 436, 331-339.
- SAHDEV, S., KHATTAR, S. K. & SAINI, K. S. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, 307, 249-264.
- SAKSELA, K. & PERMI, P. 2012. SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Letters*, 586, 2609-2614.
- SAYEGH, J., WEBB, K., CHENG, D., BEDFORD, M. T. & CLARKE, S. G. 2007. Regulation of protein arginine methyltransferase 8 (PRMT8) activity by its N-terminal domain. *Journal of Biological Chemistry*, 282, 36444-36453.
- SCAPIN, G. 2006. Structural biology and drug discovery. *Current Pharmaceutical Design*, 12, 2087-2097.
- SCHAPIRA, M. 2011. Suppl 1: Structural Chemistry of Human SET Domain Protein Methyltransferases. *Current Chemical Genomics*, 5, 85.
- SCHAPIRA, M. & DE FREITAS, R. F. 2014. Structural biology and chemistry of protein arginine methyltransferases. *MedChemComm*, 5, 1779-1788.
- SCHEER, S., ACKLOO, S., MEDINA, T. S., SCHAPIRA, M., LI, F., WARD, J. A., LEWIS, A. M., NORTHROP, J. P., RICHARDSON, P. L. & KANISKAN, H. Ü. 2019. A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nature Communications*, 10, 1-14.
- SCHLUCKEBIER, G., O'GARA, M., SAENGER, W. & CHENG, X. 1995a. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *Journal of Molecular Biology*, 1995, 16-20.
- SCHLUCKEBIER, G., O'GARA, M., SAENGER, W. & CHENG, X. 1995b. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *Journal of Molecular Biology*, 1995, 16-20.
- SCHNEEMANN, A. & YOUNG, M. J. 2003. Viral assembly using heterologous expression systems and cell extracts. *Advances in Protein Chemistry*, 64, 1-36.
- SCHUBERT, H. L., BLUMENTHAL, R. M. & CHENG, X. 2003. Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences*, 28, 329-335.
- SCHURTER, B. T., KOH, S. S., CHEN, D., BUNICK, G. J., HARP, J. M., HANSON, B. L., HENSCHEN-EDMAN, A., MACKAY, D. R., STALLCUP, M. R. & ASWAD, D. W. 2001. Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry*, 40, 5747-5756.
- SCORILAS, A., BLACK, M. H., TALIERI, M. & DIAMANDIS, E. P. 2000. Genomic organization, physical mapping, and expression analysis of the human protein arginine methyltransferase 1 gene. *Biochemical and Biophysical Research Communications*, 278, 349-359.
- SCOTT, H. S., ANTONARAKIS, S. E., LALIOTI, M. D., ROSSIER, C., SILVER, P. A. & HENRY, M. F. 1998. Identification and characterization of two putative human arginine methyltransferases (HRMT1L1 and HRMT1L2). *Genomics*, 48, 330-340.
- SELENKO, P., SPRANGERS, R., STIER, G., BÜHLER, D., FISCHER, U. & SATTLER, M. 2001. SMN tudor domain structure and its interaction with the Sm proteins. *Nature Structural & Molecular Biology*, 8, 27.
- SEZONOV, G., JOSELEAU-PETIT, D. & D'ARI, R. 2007. Escherichia coli Physiology in Luria-Bertani broth. *Journal of Bacteriology*, 189, 8746-8749.

- SHEN, Y., SZEWCZYK, M. M., ERAM, M. S., SMIL, D., KANISKAN, H. U., FERREIRA DE FREITAS, R., SENISTERRA, G., LI, F., SCHAPIRA, M. & BROWN, P. J. 2016. Discovery of a potent, selective, and cell-active dual inhibitor of protein arginine methyltransferase 4 and protein arginine methyltransferase 6. *Journal of Medicinal Chemistry*, 59, 9124-9139.
- SHI, Y. 2007. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nature Reviews Genetics*, 8, 829-833.
- SHIN, H.-J. R., KIM, H., OH, S., LEE, J.-G., KEE, M., KO, H.-J., KWEON, M.-N., WON, K.-J. & BAEK, S. H. 2016. AMPK–SKP2–CARM1 signalling cascade in transcriptional regulation of autophagy. *Nature*, 534, 553-557.
- SHISHKOVA, E., ZENG, H., LIU, F., KWIECIEN, N. W., HEBERT, A. S., COON, J. J. & XU, W. 2017. Global mapping of CARM1 substrates defines enzyme specificity and substrate recognition. *Nature Communications*, 8, 15571.
- SHLENSKY, D., MIRRIELEES, J. A., ZHAO, Z., WANG, L., MAHAJAN, A., YU, M., SHERER, N. M., WILKE, L. G. & XU, W. 2015. Differential CARM1 isoform expression in subcellular compartments and among malignant and benign breast tumors. *PLoS One*, 10, 1-14.
- SHOGREN-KNAAK, M., ISHII, H., SUN, J.-M., PAZIN, M. J., DAVIE, J. R. & PETERSON, C. L. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, 311, 844-847.
- SHOICHET, B. K. 2004. Virtual screening of chemical libraries. *Nature*, 432, 862-865.
- SHUKER, S. B., HAJDUK, P. J., MEADOWS, R. P. & FESIK, S. W. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, 274, 1531-1534.
- SIARHEYEVA, A., SENISTERRA, G., ALLALI-HASSANI, A., DONG, A., DOBROVETSKY, E., WASNEY, G. A., CHAU, I., MARCELLUS, R., HAJIAN, T. & LIU, F. 2012. An allosteric inhibitor of protein arginine methyltransferase 3. *Structure*, 20, 1425-1435.
- SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M. & SÖDING, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 1-6.
- SIMS, R. J., ROJAS, L. A., BECK, D., BONASIO, R., SCHÜLLER, R., DRURY, W. J., EICK, D. & REINBERG, D. 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science*, 332, 99-103.
- SINGHROY, D. N., MESPLÈDE, T., SABBAH, A., QUASHIE, P. K., FALGUEYRET, J.-P. & WAINBERG, M. A. 2013. Automethylation of protein arginine methyltransferase 6 (PRMT6) regulates its stability and its anti-HIV-1 activity. *Retrovirology*, 10, 73.
- SMITH, E., ZHOU, W., SHINDIAPINA, P., SIF, S., LI, C. & BAIOCCHI, R. A. 2018. Recent advances in targeting protein arginine methyltransferase enzymes in cancer therapy. *Expert Opinion on Therapeutic Targets*, 22, 527-545.
- SØRENSEN, H. P. & MORTENSEN, K. K. 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, 115, 113-128.
- SPANNHOFF, A., HEINKE, R., BAUER, I., TROJER, P., METZGER, E., GUST, R., SCHÜLE, R., BROSCHE, G., SIPPL, W. & JUNG, M. 2007a. Target-based approach to inhibitors of histone arginine methyltransferases. *Journal of Medicinal Chemistry*, 50, 2319-2325.

- SPANNHOFF, A., MACHMUR, R., HEINKE, R., TROJER, P., BAUER, I., BROSCH, G., SCHÜLE, R., HANEFELD, W., SIPPL, W. & JUNG, M. 2007b. A novel arginine methyltransferase inhibitor with cellular activity. *Bioorganic & Medicinal Chemistry Letters*, 17, 4150-4153.
- STEIN, E. M., GARCIA-MANERO, G., RIZZIERI, D. A., TIBES, R., BERDEJA, J. G., SAVONA, M. R., JONGEN-LAVRENIC, M., ALTMAN, J. K., THOMSON, B. & BLAKEMORE, S. J. 2018. The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood, The Journal of the American Society of Hematology*, 131, 2661-2669.
- STIERAND, K. & RAREY, M. 2010a. Drawing the PDB: protein– ligand complexes in two dimensions. *ACS Medicinal Chemistry Letters*, 1, 540-545.
- STIERAND, K. & RAREY, M. 2010b. PoseView--molecular interaction patterns at a glance. *Journal of Cheminformatics*, 2, P50.
- STOPA, N., KREBS, J. E. & SHECHTER, D. 2015. The PRMT5 arginine methyltransferase: many roles in development, cancer and beyond. *Cellular and Molecular Life Sciences*, 72, 2041-2059.
- STRAHL, B. D., BRIGGS, S. D., BRAME, C. J., CALDWELL, J. A., KOH, S. S., MA, H., COOK, R. G., SHABANOWITZ, J., HUNT, D. F. & STALLCUP, M. R. 2001. Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. *Current Biology*, 11, 996-1000.
- STUDIER, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, 219, 37-44.
- STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41, 207-234.
- STUDIER, F. W. 2014. Stable Expression Clones and Auto-Induction for Protein Production in *E. coli*. *Methods in Molecular Biology*, 17-32.
- SU, X., ZHU, G., DING, X., LEE, S. Y., DOU, Y., ZHU, B., WU, W. & LI, H. 2014. Molecular basis underlying histone H3 lysine–arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. *Genes & Development*, 28, 622-636.
- SUN, Q., LIU, L., ROTH, M., TIAN, J., HE, Q., ZHONG, B., BAO, R., LAN, X., JIANG, C. & SUN, J. 2015. PRMT1 upregulated by epithelial proinflammatory cytokines participates in COX2 expression in fibroblasts and chronic antigen-induced pulmonary inflammation. *The Journal of Immunology*, 195, 298-306.
- SWIERCZ, R., PERSON, M. D. & BEDFORD, M. T. 2005. Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochemical Journal*, 386, 85-91.
- SZEWCZYK, M. M., ISHIKAWA, Y., ORGAN, S., SAKAI, N., LI, F., ACKLOO, S., ERAM, M. S., DILWORTH, D., FUKUSHI, H. & HARDING, R. 2019. Pharmacological inhibition of PRMT7 links arginine monomethylation to the cellular stress response. *BioRxiv*, 503136.
- TALBOT, K., MIGUEL-ALIAGA, I., MOHAGHEGH, P., PONTING, C. P. & DAVIES, K. E. 1998. Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. *Human Molecular Genetics*, 7, 2149-2156.
- TANG, J., FRANKEL, A., COOK, R. J., KIM, S., PAIK, W. K., WILLIAMS, K. R., CLARKE, S. & HERSCHMAN, H. R. 2000. PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *Journal of Biological Chemistry*, 275, 7723-7730.

- TANG, J., GARY, J. D., CLARKE, S. & HERSCHMAN, H. R. 1998. PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *Journal of Biological Chemistry*, 273, 16935-16945.
- TANG, Y., ZENG, X. & LIANG, J. 2010. Surface plasmon resonance: an introduction to a surface spectroscopy technique. *Journal of Chemical Education*, 87, 742-746.
- TARAKHOVSKY, A. 2010. Tools and landscapes of epigenetics. *Nature Immunology*, 11, 565-568.
- TEWARY, S. K., ZHENG, Y. G. & HO, M.-C. 2019. Protein arginine methyltransferases: insights into the enzyme structure and mechanism at the atomic level. *Cellular and Molecular Life Sciences*, 1-16.
- TEYRA, J., HUANG, H., JAIN, S., GUAN, X., DONG, A., LIU, Y., TEMPEL, W., MIN, J., TONG, Y. & KIM, P. M. 2017. Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities. *Structure*, 25, 1598-1610.
- TEYSSIER, C., CHEN, D. & STALLCUP, M. R. 2002. Requirement for multiple domains of the protein arginine methyltransferase CARM1 in its transcriptional coactivator function. *Journal of Biological Chemistry*, 277, 46066-46072.
- TEYSSIER, C., OU, C.-Y., KHETCHOUMIAN, K., LOSSON, R. & STALLCUP, M. R. 2006a. TIF1 α mediates physical interaction and functional synergy between the CARM1 and GRIP1 nuclear receptor coactivators. *Molecular Endocrinology*, 20, 1276.
- TEYSSIER, C., OU, C.-Y., KHETCHOUMIAN, K., LOSSON, R. & STALLCUP, M. R. 2006b. Transcriptional intermediary factor 1 α mediates physical interaction and functional synergy between the coactivator-associated arginine methyltransferase 1 and glucocorticoid receptor-interacting protein 1 nuclear receptor coactivators. *Molecular Endocrinology*, 20, 1276-1286.
- THANDAPANI, P., O'CONNOR, T. R., BAILEY, T. L. & RICHARD, S. 2013. Defining the RGG/RG motif. *Molecular Cell*, 50, 613-623.
- THOMA, F., KOLLER, T. & KLUG, A. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, 83, 403-427.
- THOMPSON, P. R. & FAST, W. 2006. Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? *ACS Chemical Biology*, 1, 433-441.
- THORING, L., ZEMELLA, A., WÜSTENHAGEN, D. & KUBICK, S. 2019. Accelerating the production of druggable targets: Eukaryotic cell-free systems come into focus. *Methods and Protocols*, 2, 30.
- TIKHANOVICH, I., KURAVI, S., ARTIGUES, A., VILLAR, M. T., DORKO, K., NAWABI, A., ROBERTS, B. & WEINMAN, S. A. 2015. Dynamic arginine methylation of tumor necrosis factor (TNF) receptor-associated factor 6 regulates Toll-like receptor signaling. *Journal of Biological Chemistry*, 290, 22236-22249.
- TOMA-FUKAI, S., KIM, J.-D., PARK, K.-E., KUWABARA, N., SHIMIZU, N., KRAYUKHINA, E., UCHIYAMA, S., FUKAMIZU, A. & SHIMIZU, T. 2016. Novel helical assembly in arginine methyltransferase 8. *Journal of Molecular Biology*, 428, 1197-1208.
- TOSHIKI WATANABE, M. Y., OSAMU KANNO, JUN WATANABE, NOBUAKI ADACHI, DAISUKE HONMA, YOSHITO HAMADA. 2016. *Therapeutic and/or prophylactic agent for adult t cell leukemia/lymphoma*. EP3329917A1.

- TRIPSIANES, K., MADL, T., MACHYNA, M., FESSAS, D., ENGLBRECHT, C., FISCHER, U., NEUGEBAUER, K. M. & SATTLE, M. 2011. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nature Structural & Molecular Biology*, 18, 1414.
- TROFFER-CHARLIER, N., CURA, V., HASSENBOEHLER, P., MORAS, D. & CAVARELLI, J. 2007. Functional insights from structures of coactivator-associated arginine methyltransferase 1 domains. *The EMBO Journal*, 26, 4391-4401.
- TROPBERGER, P. & SCHNEIDER, R. 2013. Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nature Structural & Molecular Biology*, 20, 657.
- TSUKADA, Y.-I., FANG, J., ERDJUMENT-BROMAGE, H., WARREN, M. E., BORCHERS, C. H., TEMPST, P. & ZHANG, Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, 439, 811-816.
- VAGENENDE, V., YAP, M. G. & TROUT, B. L. 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*, 48, 11084-11096.
- VAN HAREN, M., VAN UFFORD, L. Q., MORET, E. E. & MARTIN, N. I. 2015. Synthesis and evaluation of protein arginine N-methyltransferase inhibitors designed to simultaneously occupy both substrate binding sites. *Organic & Biomolecular Chemistry*, 13, 549-560.
- VAN HAREN, M. J., MARECHAL, N., TROFFER-CHARLIER, N., CIANCIULLI, A., SBARDELLA, G., CAVARELLI, J. & MARTIN, N. I. 2017. Transition state mimics are valuable mechanistic probes for structural studies with the arginine methyltransferase CARM1. *Proceedings of the National Academy of Sciences*, 114, 3625-3630.
- VAN HOLDE, K. & ZLATANOVA, J. 1995. Chromatin higher order structure: chasing a mirage? *Journal of Biological Chemistry*, 270, 8373-8376.
- VHUIYAN, M. I., PAK, M. L., PARK, M. A., THOMAS, D., LAKOWSKI, T. M., CHALFANT, C. E. & FRANKEL, A. 2017. PRMT2 interacts with splicing factors and regulates the alternative splicing of BCL-X. *The Journal of Biochemistry*, 162, 17-25.
- VONRHEIN, C., FLENSBURG, C., KELLER, P., SHARFF, A., SMART, O., PACIOREK, W., WOMACK, T. & BRICOGNE, G. 2011. Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography*, 67, 293-302.
- VU, L. P., PERNA, F., WANG, L., VOZA, F., FIGUEROA, M. E., TEMPST, P., ERDJUMENT-BROMAGE, H., GAO, R., CHEN, S. & PAIETTA, E. 2013. PRMT4 blocks myeloid differentiation by assembling a methyl-RUNX1-dependent repressor complex. *Cell Reports*, 5, 1625-1638.
- WALPORT, L. J., HOPKINSON, R. J., CHOWDHURY, R., SCHILLER, R., GE, W., KAWAMURA, A. & SCHOFIELD, C. J. 2016. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nature Communications*, 7, 11974.
- WANG, C., ZHU, Y., CACERES, T. B., LIU, L., PENG, J., WANG, J., CHEN, J., CHEN, X., ZHANG, Z. & ZUO, X. 2014. Structural determinants for the strict monomethylation activity by trypanosoma brucei protein arginine methyltransferase 7. *Structure*, 22, 756-768.

- WANG, L., CHAROENSUKSAI, P., WATSON, N. J., WANG, X., ZHAO, Z., CORIANO, C. G., KERR, L. R. & XU, W. 2013a. CARM1 automethylation is controlled at the level of alternative splicing. *Nucleic Acids Research*, 41, 6870-6880.
- WANG, M., XU, R.-M. & THOMPSON, P. R. 2013b. Substrate specificity, processivity, and kinetic mechanism of protein arginine methyltransferase 5. *Biochemistry*, 52, 5430-5440.
- WANG, Y., WYSOCKA, J., SAYEGH, J., LEE, Y.-H., PERLIN, J. R., LEONELLI, L., SONBUCHNER, L. S., MCDONALD, C. H., COOK, R. G. & DOU, Y. 2004. Human PAD4 regulates histone arginine methylation levels via demethyliminination. *Science*, 306, 279-283.
- WARD, J. J., MCGUFFIN, L. J., BRYSON, K., BUXTON, B. F. & JONES, D. T. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, 20, 2138-2139.
- WATTS, J. M., BRADLEY, T. J., THOMASSEN, A., BRUNNER, A. M., MINDEN, M. D., PAPADANTONAKIS, N., ABEDIN, S., BAINES, A. J., BARBASH, O. & GORMAN, S. 2019. A Phase I/II Study to Investigate the Safety and Clinical Activity of the Protein Arginine Methyltransferase 5 Inhibitor GSK3326595 in Subjects with Myelodysplastic Syndrome and Acute Myeloid Leukemia. *Blood, The Journal of the American Society of Hematology*, 134, 2656.
- WEBBY, C. J., WOLF, A., GROMAK, N., DREGER, M., KRAMER, H., KESSLER, B., NIELSEN, M. L., SCHMITZ, C., BUTLER, D. S. & YATES, J. R. 2009. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science*, 325, 90-93.
- WEI, H.-H., FAN, X.-J., HU, Y., GUO, M., FANG, Z.-Y., WU, P., TIAN, X.-X., GAO, S.-X., PENG, C. & YANG, Y. 2019. A systematic survey of PRMT interactomes reveals the key roles of arginine methylation in the global control of RNA splicing and translation. *BioRxiv*, 746529.
- WEISS, V. H., MCBRIDE, A. E., SORIANO, M. A., FILMAN, D. J., SILVER, P. A. & HOGLE, J. M. 2000. The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nature Structural Biology*, 7, 1165-1171.
- WINTER, M., BRETSCHEIDER, T., KLEINER, C., RIES, R., HEHN, J. P., REDEMANN, N., LUIPPOLD, A. H., BISCHOFF, D. & BÜTTNER, F. H. 2018. Establishing MALDI-TOF as versatile drug discovery readout to dissect the PTP1B enzymatic reaction. *SLAS DISCOVERY: Advancing Life Sciences R&D*, 23, 561-573.
- WOOD, D. J., LOPEZ-FERNANDEZ, J. D., KNIGHT, L. E., AL-KHAWALDEH, I., GAI, C., LIN, S., MARTIN, M. P., MILLER, D. C., CANO, C. & ENDICOTT, J. A. 2019. FragLites—Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *Journal of Medicinal Chemistry*, 62, 3741-3752.
- WU, H., ZHENG, W., ERAM, M. S., VHUIYAN, M., DONG, A., ZENG, H., HE, H., BROWN, P., FRANKEL, A. & VEDADI, M. 2016. Structural basis of arginine asymmetrical dimethylation by PRMT6. *Biochemical Journal*, 473, 3049-3063.
- WU, J., XIE, N., FENG, Y. & ZHENG, Y. G. 2012. Scintillation proximity assay of arginine methylation. *Journal of Biomolecular Screening*, 17, 237-244.
- WU, T., BREHMER, D., BEKE, L., BOECKX, A., DIELS, G. S. M., LAWSON, E. C., MEERPOEL, L., PANDE, V., PARADÉ, M. C. B. C. & SCHEPENS, W. B. G. 2019. Novel monocyclic and bicyclic ring system substituted carbanucleoside analogues for use as prmt5 inhibitors. US patent application.

- WU, T., MILLAR, H., GAFFNEY, D., BEKE, L., MANNENS, G., VINKEN, P., SOMMERS, I., THURING, J.-W., SUN, W. & MOY, C. JNJ-64619178, a selective and pseudo-irreversible PRMT5 inhibitor with potent in vitro and in vivo activity, demonstrated in several lung cancer models. AACR Annual Meeting, 2018a. AACR.
- WU, T., MILLAR, H., GAFFNEY, D., BEKE, L., MANNENS, G., VINKEN, P., SOMMERS, I., THURING, J.-W., SUN, W. & MOY, C. 2018b. JNJ-64619178, a selective and pseudo-irreversible PRMT5 inhibitor with potent in vitro and in vivo activity, demonstrated in several lung cancer models. *Experimental and Molecular Therapeutics*, 78, 2414-2428.
- WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T. & BADENHORST, P. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442, 86-90.
- XIE, B., INVERNIZZI, C. F., RICHARD, S. & WAINBERG, M. A. 2007. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat Interactions with both cyclin T1 and the Tat transactivation region. *Journal of Virology*, 81, 4226-4234.
- XIE, Y., ZHOU, R., LIAN, F., LIU, Y., CHEN, L., SHI, Z., ZHANG, N., ZHENG, M., SHEN, B. & JIANG, H. 2014. Virtual screening and biological evaluation of novel small molecular inhibitors against protein arginine methyltransferase 1 (PRMT1). *Organic & Biomolecular Chemistry*, 12, 9665-9673.
- XU, W., CHEN, H., DU, K., ASAHARA, H., TINI, M., EMERSON, B. M., MONTMINY, M. & EVANS, R. M. 2001. A transcriptional switch mediated by cofactor methylation. *Science*, 294, 2507-2511.
- XU, W., CHO, H., KADAM, S., BANAYO, E. M., ANDERSON, S., YATES, J. R., EMERSON, B. M. & EVANS, R. M. 2004. A methylation-mediator complex in hormone signaling. *Genes & Development*, 18, 144-156.
- XU, Z., JIANG, J., XU, C., WANG, Y., SUN, L., GUO, X. & LIU, H. 2013. MicroRNA-181 regulates CARM1 and histone arginine methylation to promote differentiation of human embryonic stem cells. *PLoS One*, 8, e53146.
- YADAV, N., LEE, J., KIM, J., SHEN, J., HU, M. C.-T., ALDAZ, C. M. & BEDFORD, M. T. 2003. Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. *Proceedings of the National Academy of Sciences*, 100, 6464-6468.
- YANG, Y. & BEDFORD, M. T. 2013. Protein arginine methyltransferases and cancer. *Nature Reviews Cancer*, 13, 37.
- YANG, Y., HADJIKYRIACOU, A., XIA, Z., GAYATRI, S., KIM, D., ZURITA-LOPEZ, C., KELLY, R., GUO, A., LI, W. & CLARKE, S. G. 2015. PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145. *Nature Communications*, 6, 1-12.
- YANG, Y., LU, Y., ESPEJO, A., WU, J., XU, W., LIANG, S. & BEDFORD, M. T. 2010. TDRD3 is an effector molecule for arginine-methylated histone marks. *Molecular Cell*, 40, 1016-1023.
- YANG, Y., MCBRIDE, K. M., HENSLEY, S., LU, Y., CHEDIN, F. & BEDFORD, M. T. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Molecular Cell*, 53, 484-497.
- YANG, Z. R., THOMSON, R., MCNEIL, P. & ESNOUF, R. M. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21, 3369-3376.

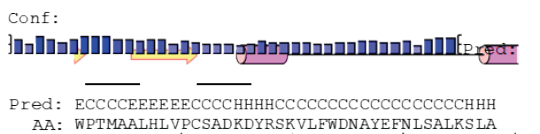
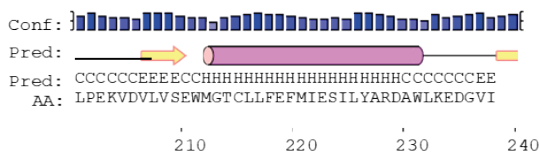
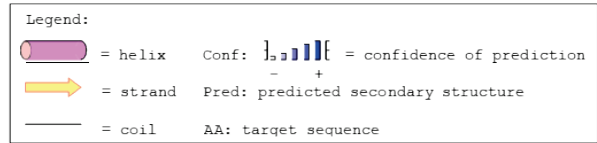
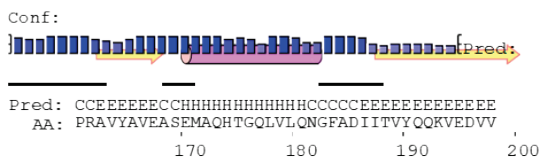
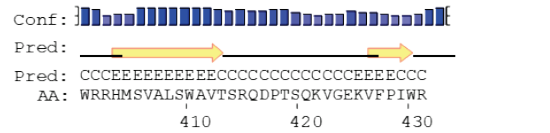
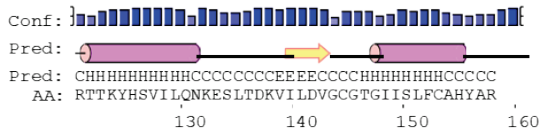
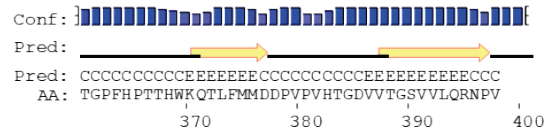
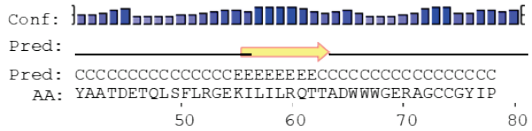
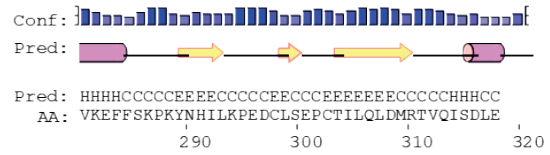
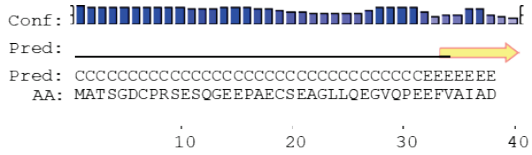
- YOSHIMOTO, T., BOEHM, M., OLIVE, M., CROOK, M. F., SAN, H., LANGENICKEL, T. & NABEL, E. G. 2006. The arginine methyltransferase PRMT2 binds RB and regulates E2F function. *Experimental Cell Research*, 312, 2040-2053.
- YU, H., CHEN, J. K., FENG, S., DALGARNO, D. C., BRAUER, A. W. & SCHRELBER, S. L. 1994. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*, 76, 933-945.
- YU, Y., NIE, Y., FENG, Q., QU, J., WANG, R., BIAN, L. & XIA, J. 2017. Targeted covalent inhibition of Grb2–Sos1 interaction through proximity-induced conjugation in breast cancer cells. *Molecular Pharmaceutics*, 14, 1548-1557.
- YUE, W. W., HASSLER, M., ROE, S. M., THOMPSON-VALE, V. & PEARL, L. H. 2007. Insights into histone code syntax from structural and biochemical studies of CARM1 methyltransferase. *The EMBO Journal*, 26, 4402-4412.
- ZAKRZEWICZ, D., ZAKRZEWICZ, A., PREISSNER, K. T., MARKART, P. & WYGRECKA, M. 2012. Protein arginine methyltransferases (PRMTs): promising targets for the treatment of pulmonary disorders. *International Journal of Molecular Sciences*, 13, 12383-12400.
- ZHANG, H.-T., ZENG, L.-F., HE, Q.-Y., TAO, W. A., ZHA, Z.-G. & HU, C.-D. 2016. The E3 ubiquitin ligase CHIP mediates ubiquitination and proteasomal degradation of PRMT5. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1863, 335-346.
- ZHANG, H., LIU, K., IZUMI, N., HUANG, H., DING, D., NI, Z., SIDHU, S. S., CHEN, C., TOMARI, Y. & MIN, J. 2017a. Structural basis for arginine methylation-independent recognition of PIWIL1 by TDRD2. *Proceedings of the National Academy of Sciences*, 114, 12483-12488.
- ZHANG, J., JING, L., LI, M., HE, L. & GUO, Z. 2019. Regulation of histone arginine methylation/demethylation by methylase and demethylase. *Molecular Medicine Reports*, 19, 3963-3971.
- ZHANG, J. & ZHENG, Y. G. 2015. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chemical Biology*, 11, 583-597.
- ZHANG, J. & ZHENG, Y. G. 2016. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chemical Biology*, 11, 583-597.
- ZHANG, R., LI, X., LIANG, Z., ZHU, K., LU, J., KONG, X., OUYANG, S., LI, L., ZHENG, Y. G. & LUO, C. 2013. Theoretical insights into catalytic mechanism of protein arginine methyltransferase 1. *PLoS One*, 8, e72424.
- 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.
- ZHANG, X., ZHOU, L. & CHENG, X. 2000. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *The EMBO Journal*, 19, 3509-3519.
- ZHANG, Z., NIKOLAI, B. C., GATES, L. A., JUNG, S. Y., SIWAK, E. B., HE, B., RICE, A. P., O'MALLEY, B. W. & FENG, Q. 2017b. Crosstalk between histone modifications indicates that inhibition of arginine methyltransferase CARM1 activity reverses HIV latency. *Nucleic Acids Research*, 45, 9348-9360.
- ZHAO, Q., QIN, L., JIANG, F., WU, B., YUE, W., XU, F., RONG, Z., YUAN, H., XIE, X. & GAO, Y. 2007. Structure of human Spindlin1 Tandem tudor-like domains for cell cycle regulation. *Journal of Biological Chemistry*, 282, 647-656.
- ZHAO, Q., RANK, G., TAN, Y. T., LI, H., MORITZ, R. L., SIMPSON, R. J., CERRUTI, L., CURTIS, D. J., PATEL, D. J. & ALLIS, C. D. 2009. PRMT5-mediated

- methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature Structural & Molecular Biology*, 16, 304.
- ZHAO, Z. & SHILATIFARD, A. 2019. Epigenetic modifications of histones in cancer. *Genome Biology*, 20, 1-16.
- ZHONG, J., CAO, R.-X., HONG, T., YANG, J., ZU, X.-Y., XIAO, X.-H., LIU, J.-H. & WEN, G.-B. 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.
- ZHONG, J., CAO, R., LIU, J., LIU, Y., WANG, J., LIU, L., CHEN, Y., YANG, J., ZHANG, Q. & WU, Y. 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.
- ZHONG, J., CAO, R. X., ZU, X. Y., HONG, T., YANG, J., LIU, L., XIAO, X. H., DING, W. J., ZHAO, Q. & LIU, J. H. 2012. Identification and characterization of novel spliced variants of PRMT2 in breast carcinoma. *The FEBS Journal*, 279, 316-335.
- ZHONG, J., CHEN, Y.-J., CHEN, L., SHEN, Y.-Y., ZHANG, Q.-H., YANG, J., CAO, R.-X., ZU, X.-Y. & WEN, G.-B. 2017. PRMT2 β , a C-terminal splice variant of PRMT2, inhibits the growth of breast cancer cells. *Oncology Reports*, 38, 1303-1311.
- ZHOU, K., GAULLIER, G. & LUGER, K. 2019. Nucleosome structure and dynamics are coming of age. *Nature Structural & Molecular Biology*, 26, 3-13.

Appendix A.

Secondary Structure Prediction of HsPRMT2.

The secondary structure of PRMT2 was predicted using the structure prediction server PSI-PRED (McGuffin et al., 2000).



Appendix B.

E. coli Strains and Their Properties.

<i>E. Coli</i> Strain	Properties
BL21(DE3) (Promega)	BL21(DE3) can be used for the expression of target genes which are under the control of the T7 RNA Polymerase promoter. The genes which encode for the T7 RNA Polymerase which are under the control of the lacUV5 promoter are integrated into the bacterial chromosome. Moreover, the cells are deficient in the proteases Lon and OmpT reducing degradation of the recombinant proteins.
BL21(DE3)Gold (Agilent Technologies)	The BL21(DE3) derivate has an increased transformation efficiency and the genes that encode endonucleases which degrades plasmids are inactivated. They are mostly used for the expression of non-toxic proteins. BL21(DE3)Gold cells are resistant to tetracycline.
BL21(DE3)pLysS (Promega)	This strain expresses the T7 lysozyme, which inhibits the T7 RNA Polymerase prior to expression induction. It can be used to enhance the expression of toxic proteins by decreasing basal expression levels. The T7 lysozyme gene is carried on a pLysS plasmid that renders the strain resistant to Chloramphenicol.
Rosetta (DE3)pLysS (Promega)	This BL21(DE3) derivate is designed for expression of proteins that contain rare <i>E. coli</i> codons. It supplies tRNAs for the rare codons on a chloramphenicol-resistant plasmid. Moreover, they express T7 lysozyme which reduces basal expression levels of the target protein prior to induction.
ArcticExpress (Agilent Technologies)	The ArcticExpress strain expresses the two chaperonins Cpn10 and Cpn60, which show high activity even at low temperatures. Often recombinant protein expression in <i>E. coli</i> results in the formation of protein aggregates and inclusion bodies. Lower temperature can help to decrease incorrect folding of proteins; however, the <i>E. coli</i> chaperonins show reduced activity at lower temperatures. Using ArcticExpress cells potentially increases the amount of properly folded soluble protein expression at reduced expression temperatures due to the presence of cold-adapted chaperonins. The plasmid that carries the chaperone genes also carries gentamycin-resistance genes.
DH5a	For DNA amplification the bacterial strain DH5a was used. The cells allow for high-efficiency transformations and are less prone to unwanted recombination. In cases where the protein seemed to be toxic for the cells (Mouse PRMT2 constructs), Stellar competent cells were used instead.
Stellar Competent Cells (Clontech)	The Stellar <i>E. coli</i> strain has a very high transformation efficiency and was used for InFusion cloning.

Appendix C.

Buffer Screen Layout and Composition.

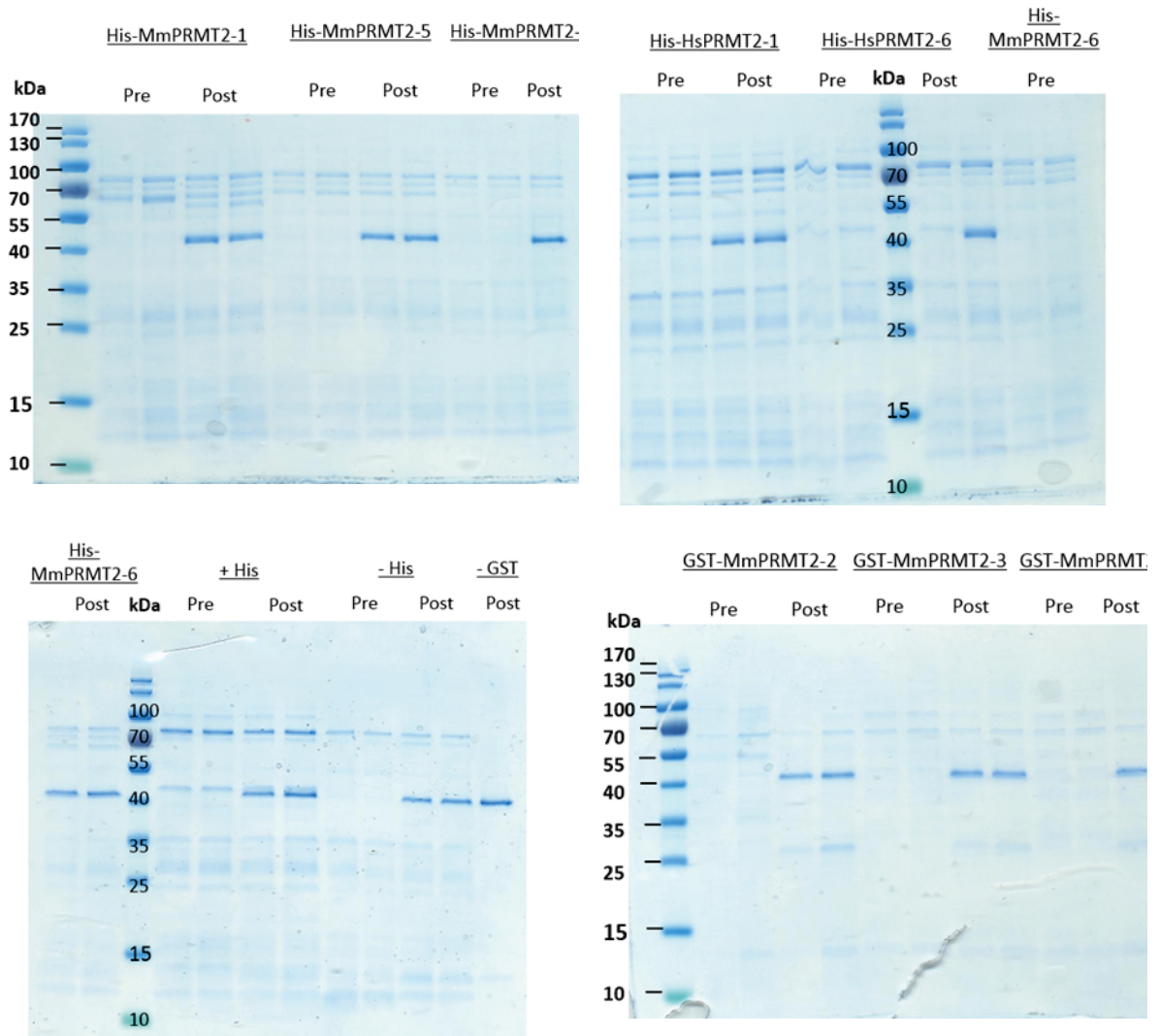
	1	2	3	4	5	6	7	8	9	10	11	12	
A,B,C	NaAcetate, pH4.8, 0 mM NaCl	NaAcetate, pH4.8, 100 mM NaCl	NaAcetate, pH4.8, 300 mM NaCl	NaCitrate, pH5.5, 0mM NaCl	NaCitrate, pH5.5, 100mM NaCl	NaCitrate, pH5.5, 300mM NaCl	MES, pH6.1, 0 mM NaCl	MES, pH6.1, 100 mM NaCl	MES, pH6.1, 300 mM NaCl	NaPhos, pH6.1, 0 mM NaCl	NaPhos, pH6.1, 100 mM NaCl	NaPhos, pH6.1, 300 mM NaCl	
D,E,F	Bis-Tris, pH 6.5, 0 mM NaCl	Bis-Tris, pH 6.5, 100 mM NaCl	Bis-Tris, pH 6.5, 300mM NaCl	PIPES, pH6.8, 0 mM NaCl	PIPES, pH6.8, 10 mM NaCl	PIPES, pH6.8, 300 mM NaCl	NaPhos, pH 6.8, 0 mM NaCl	NaPhos, pH 6.8, 100 mM NaCl	NaPhos, pH 6.8, 300 mM NaCl	BES, pH 7.0, 0 mM NaCl	BES, pH 7.0, 100 mM NaCl	BES, pH 7.0, 300 mM NaCl	
G,H,I	MOPS, pH7.2, 0mM NaCl	MOPS, pH7.2, 100 mM NaCl	MOPS, pH7.2, 300 mM NaCl	NaPhos, pH7.5, 0 mM NaCl	NaPhos, pH7.5, 100 mM NaCl	NaPhos, pH7.5, 300 mM NaCl	HEPES, pH7.5, 0 mM NaCl	HEPES, pH7.5, 100 mM NaCl	HEPES, pH7.5, 300 mM NaCl	Tris, pH 8.1, 0 mM NaCl	Tris, pH 8.1, 100 mM NaCl	Tris, pH 8.1, 300 mM NaCl	
J,K,L	BICINE, pH 8.4, 0 mM NaCl	BICINE, pH 8.4, 100 mM NaCl	BICINE, pH 8.4, 300 mM NaCl	Bis-Tris Propane, pH8.9, 0 mM NaCl	CHES, pH 9.5, 100 mM NaCl	CHES, pH 9.5, 300 mM NaCl	CHES, pH 9.5, 0 mM NaCl	CHES, pH 9.5, 100 mM NaCl	CHES, pH 9.5, 300 mM NaCl	Control buffer	Control buffer	Control buffer	
M,N,O	50 mM Tris pH 7.5, 50 mM NaCl	50 mM Tris pH 7.5, 120 mM NaCl	50 mM Tris pH 7.5, 140 mM NaCl	50 mM Tris pH 7.5, 160 mM NaCl	50 mM Tris pH 7.5, 180 mM NaCl	50 mM Tris pH 7.5, 500 mM NaCl, 10 % Glycerol	50 mM Tris pH 7.5, 300 mM NaCl, 10 % Glycerol	50 mM Tris pH 7.5, 10 % Glycerol		x	x	x	x

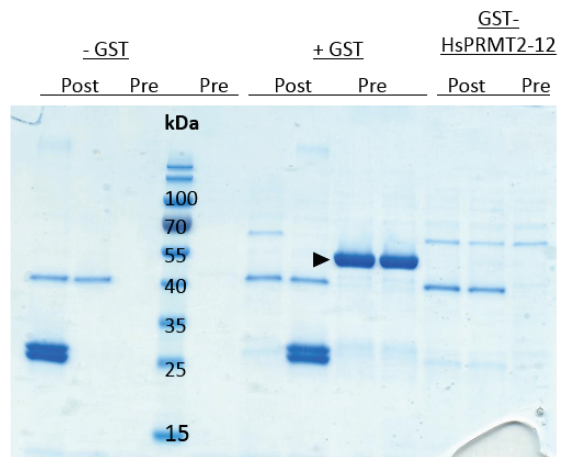
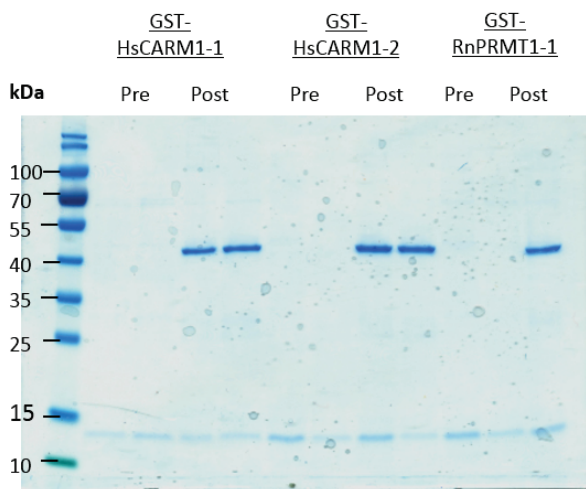
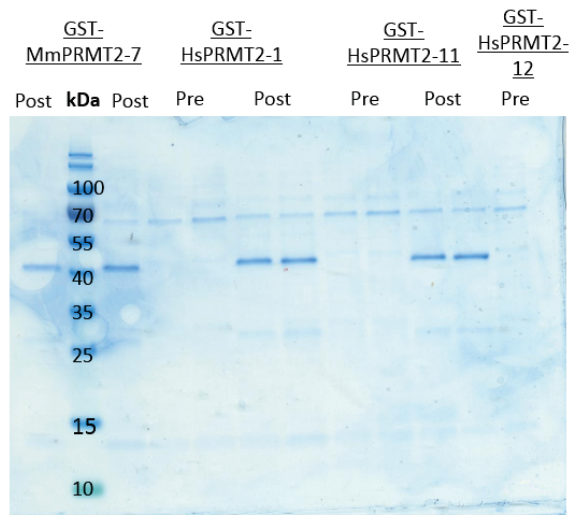
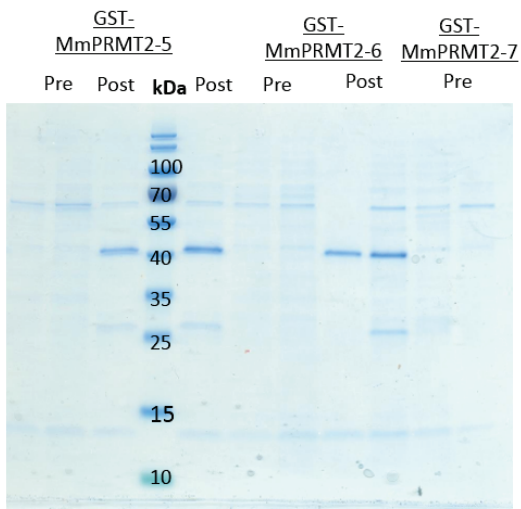
*The control buffer is the previously used buffer (40 mM HEPES pH 8.0, 600 mM NaCl, 0.5 mM TCEP, 5 % (v/v) glycerol).

Appendix D.

SDS-PAGE Analysis of Expression Tests in *E. coli* BL21(DE3)Gold.

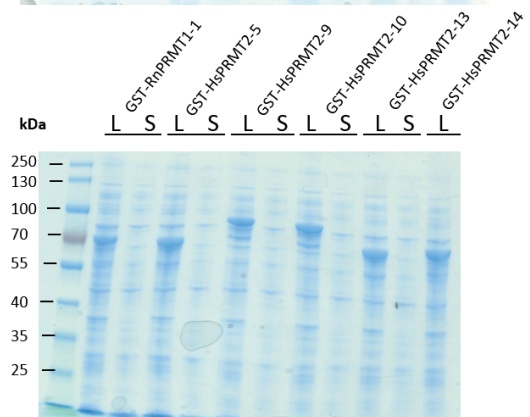
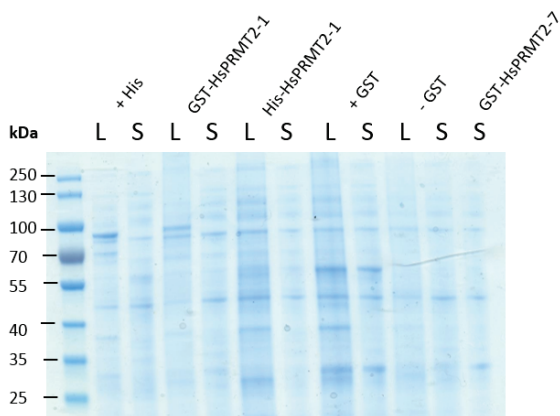
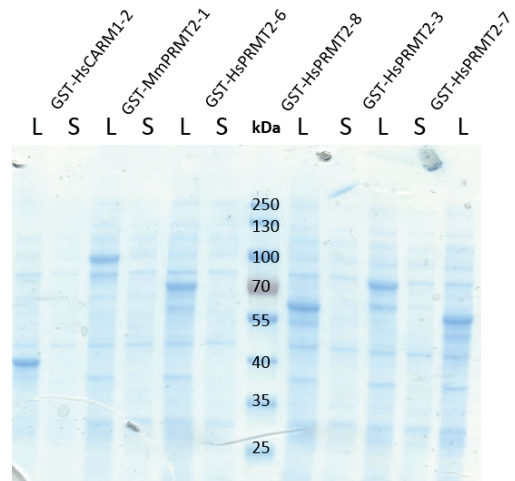
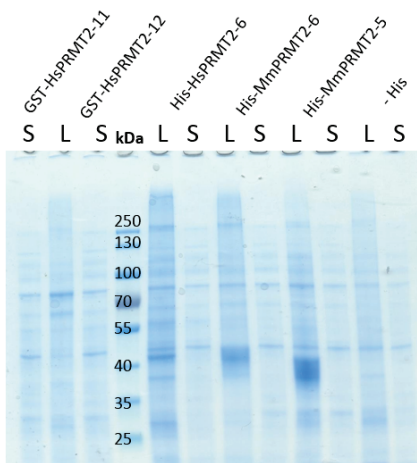
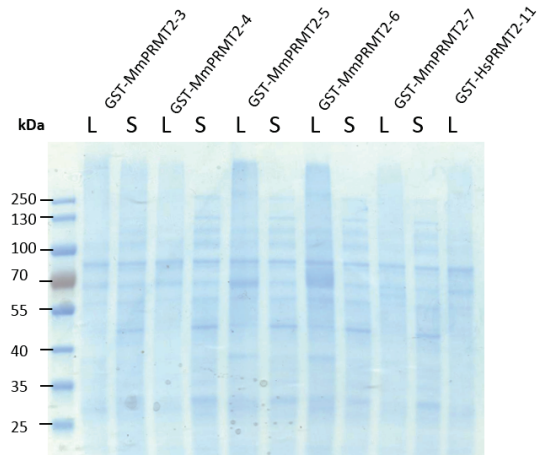
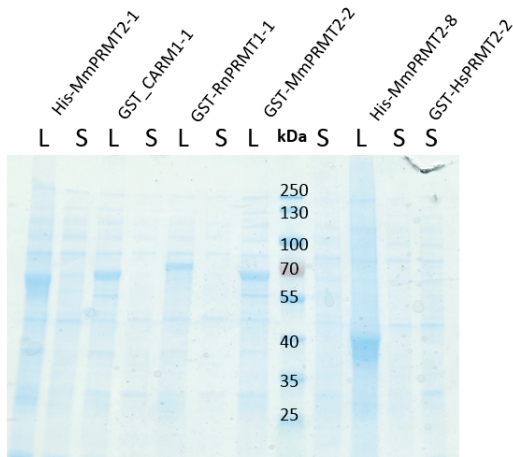
The cells were grown in 3 mL AIM with the appropriate antibiotic at 37 °C, 220 rpm for 5-6 h and then shifted to 18 °C overnight. The cells were harvested by centrifugation and lysed via lysozyme and freeze and thaw cycles. After batch purification one sample was taken for SDS-PAGE (Pre), another sample was cleaved with 3C-protease overnight and also run on the gel (Post). Protein ladder: PAGERuler Prestained (ThermoFisher).

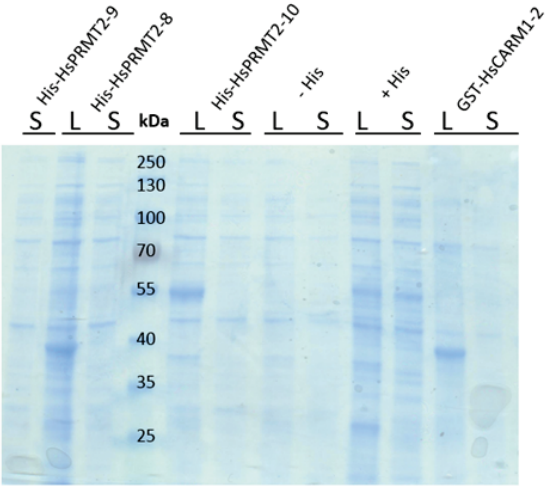
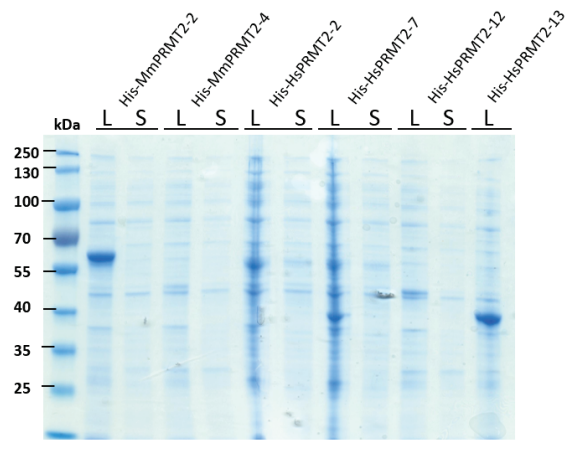
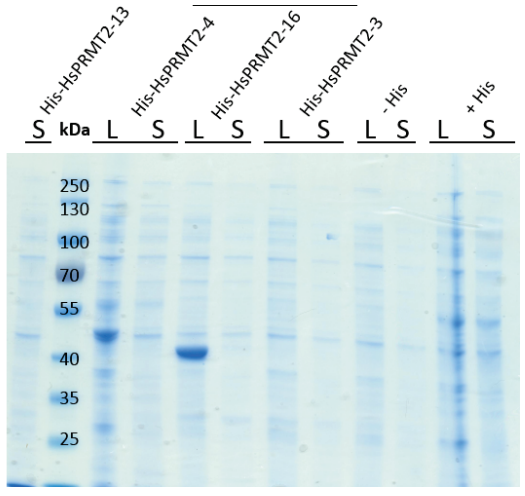
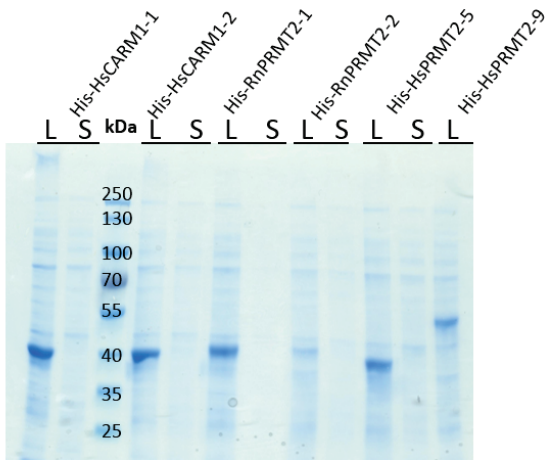
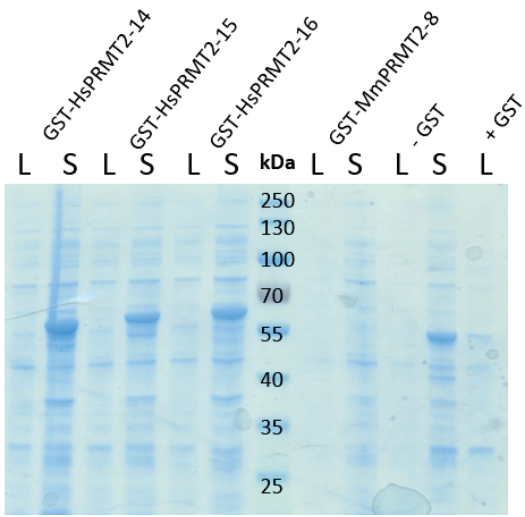




SDS-PAGE Analysis of Expression Tests in *E. coli* BL21(DE3)Gold.

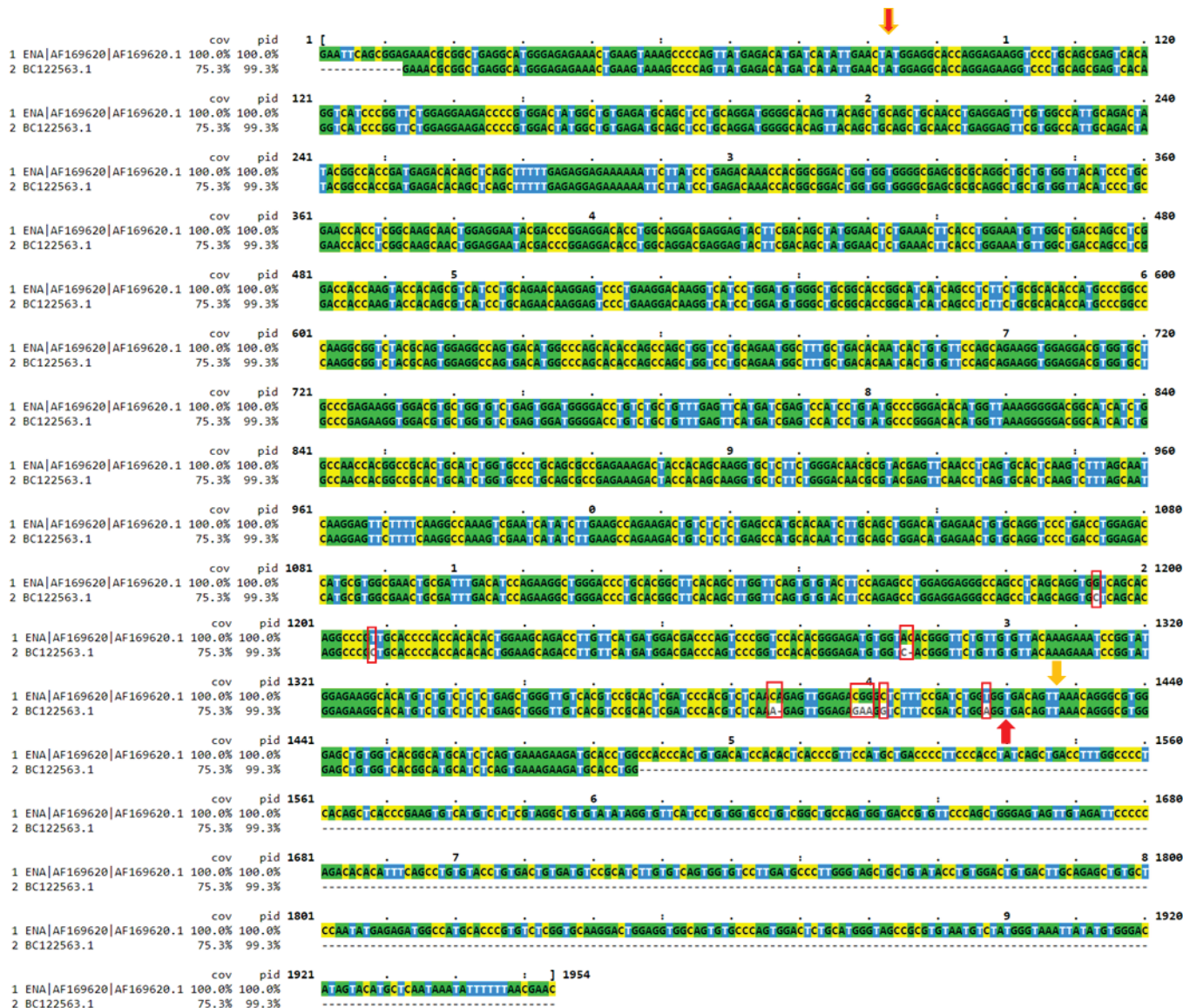
The cells were grown in 3 mL AIM with the appropriate antibiotic at 37 °C, 220 rpm for 5-6 h and then shifted to 18 °C overnight. The cells were harvested by centrifugation and lysed via lysozyme and freeze and thaw cycles. A lysate and a soluble fraction sample were analysed on a 10 % SDS-PAGE gel to determine the solubility of the expressed protein. Protein ladder: PAGERuler Prestained (ThermoFisher).





Appendix E.

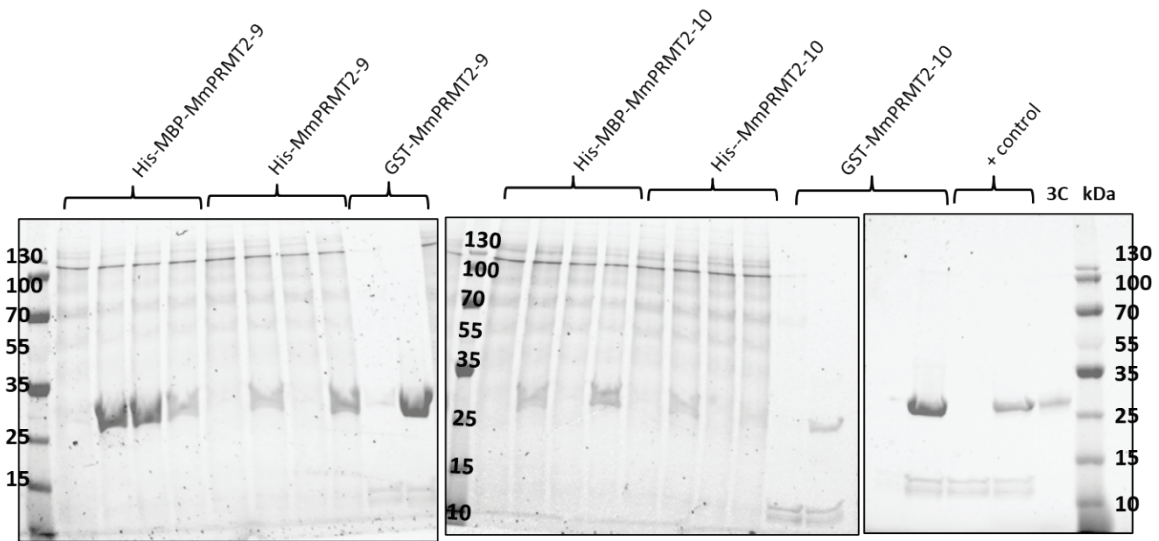
MmPRMT2 DNA Sequence Alignment.



The two different MmPRMT2 DNA sequences (GenBank AF169620.1 and BC122563.1) were aligned. Differences are highlighted with red boxes. The start and stop codons are indicated with red and orange arrows. Figure was prepared using MView 1.63 (Brown et al., 1998).

Appendix F.

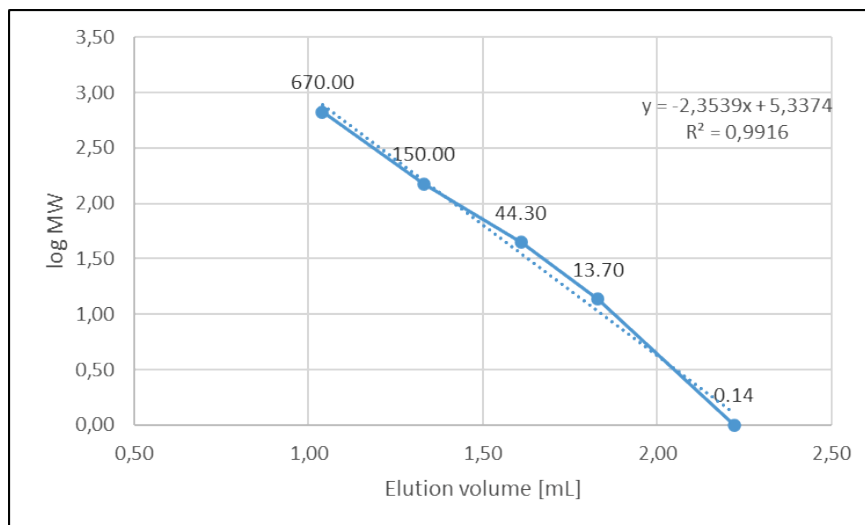
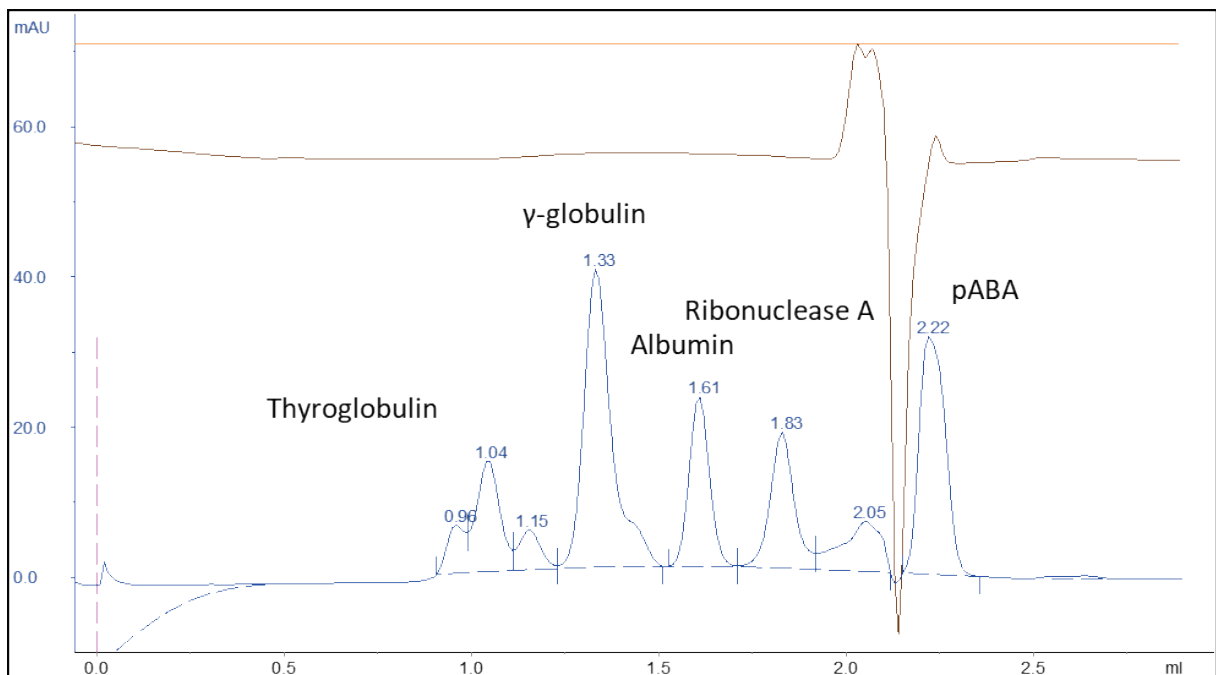
Small-Scale Expression Test of MmPRMT2-9 and MmPRMT2-10.



Elution fractions were incubated with His-3C-protease for 1 h at RT, 20 μ L of the uncleaved and cleaved sample were run on a 10 % SDS Gel in duplicate. The band above the 25 kDa weight marker is His-3C-protease. Protein ladder: PAGERuler Prestained (ThermoFisher).

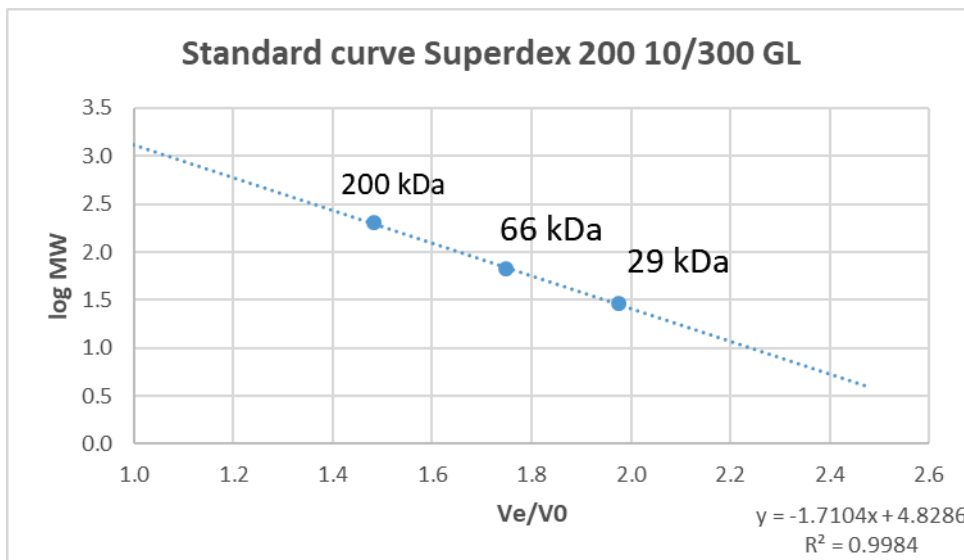
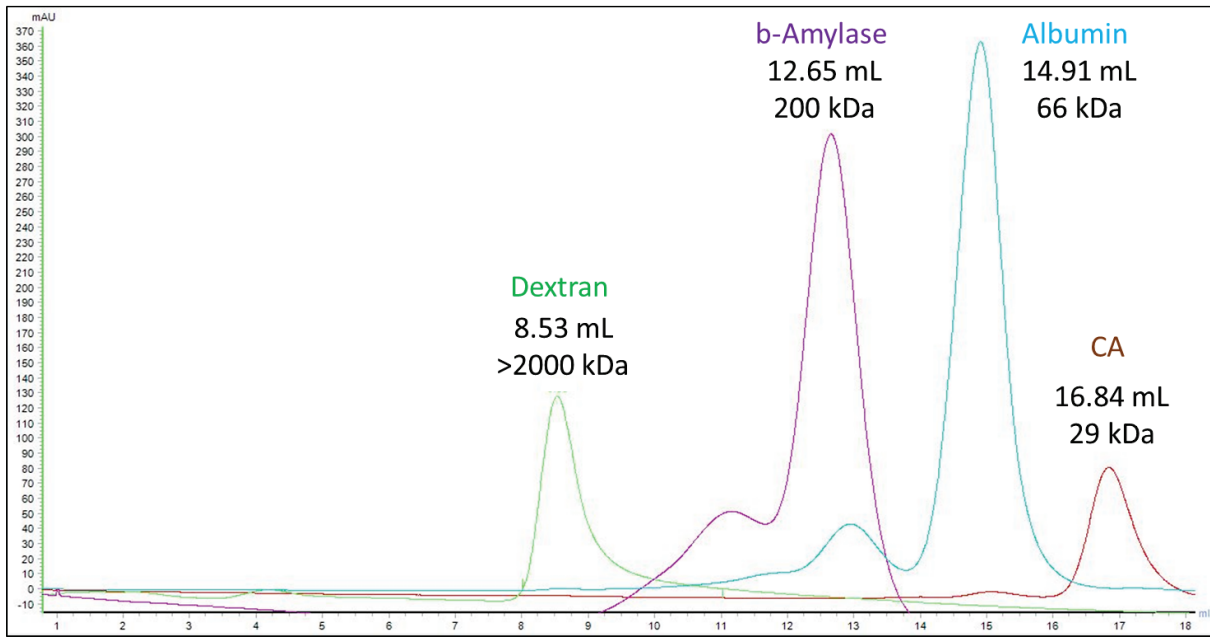
Appendix G.

Superdex 75 10/300 GL Calibration Curve.



The logarithm of the molecular weight of each marker was plotted against the elution volume in mL to obtain the calibration curve. The molecular weight of each marker protein is shown above each point.

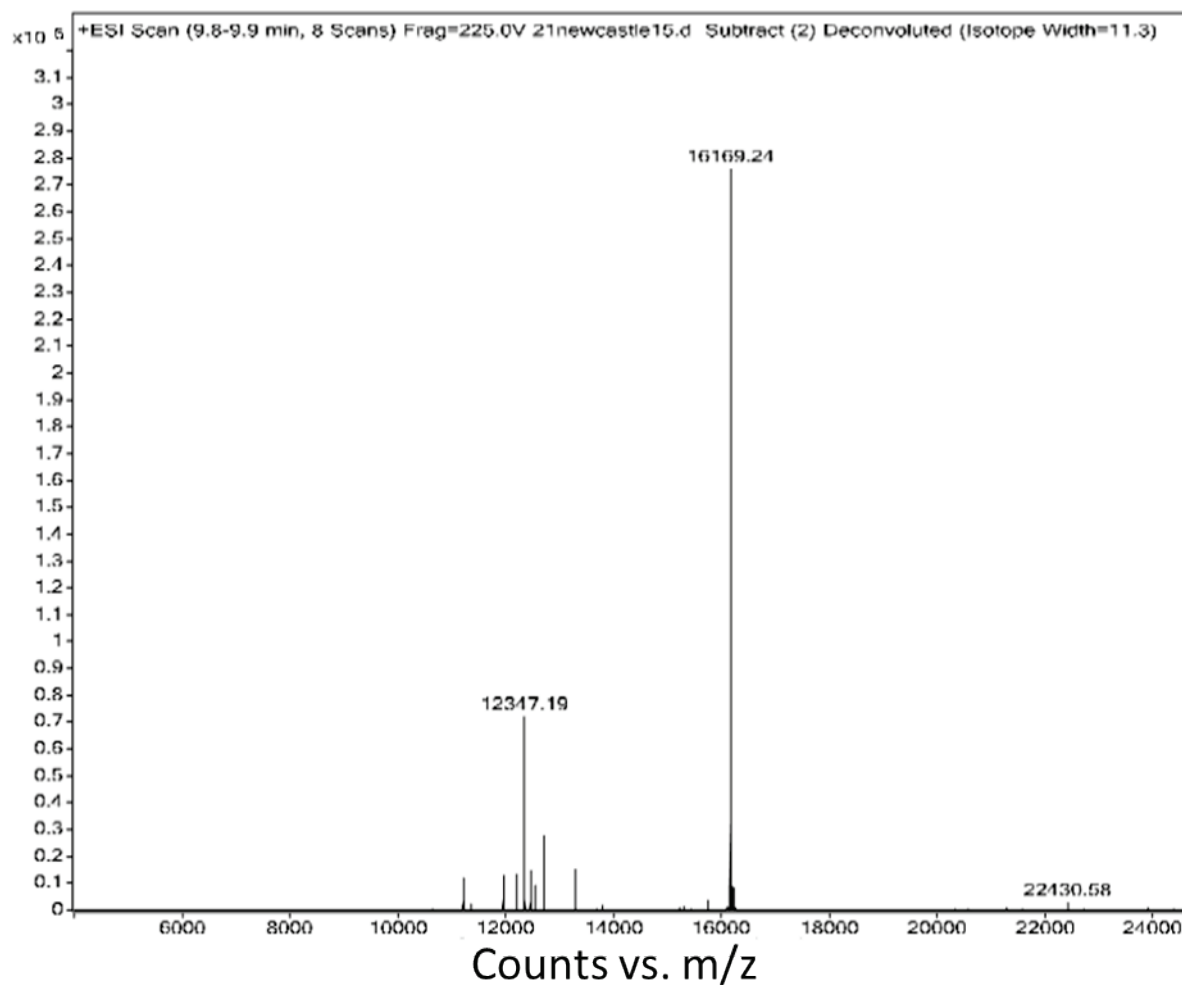
Superdex 200 10/300 GL Calibration Curve.



The logarithm of the molecular weight of each marker was plotted against the elution volume/void volume in mL to obtain the calibration curve. The molecular weight of each marker protein is shown above each point.

Appendix H.

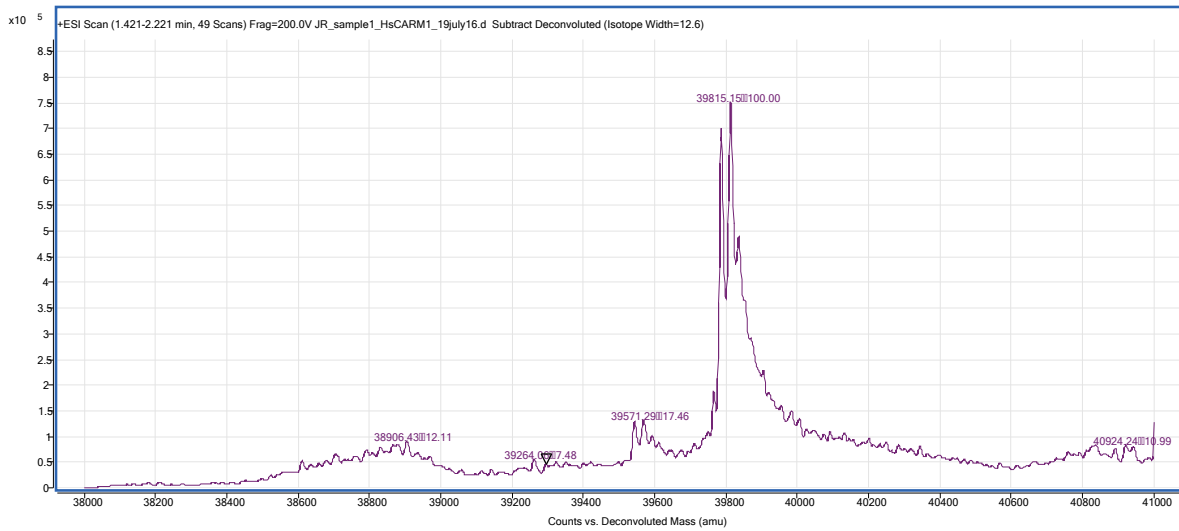
Analysis of RSF1 Sample by Mass Spectrometry.



The deconvoluted spectrum reveals a dominant mass species at 16169.24 Da corresponding to RSF1, which has an expected molecular weight of 16111.80 Da. Another unknown mass species at 12347.19 Da was also detected.

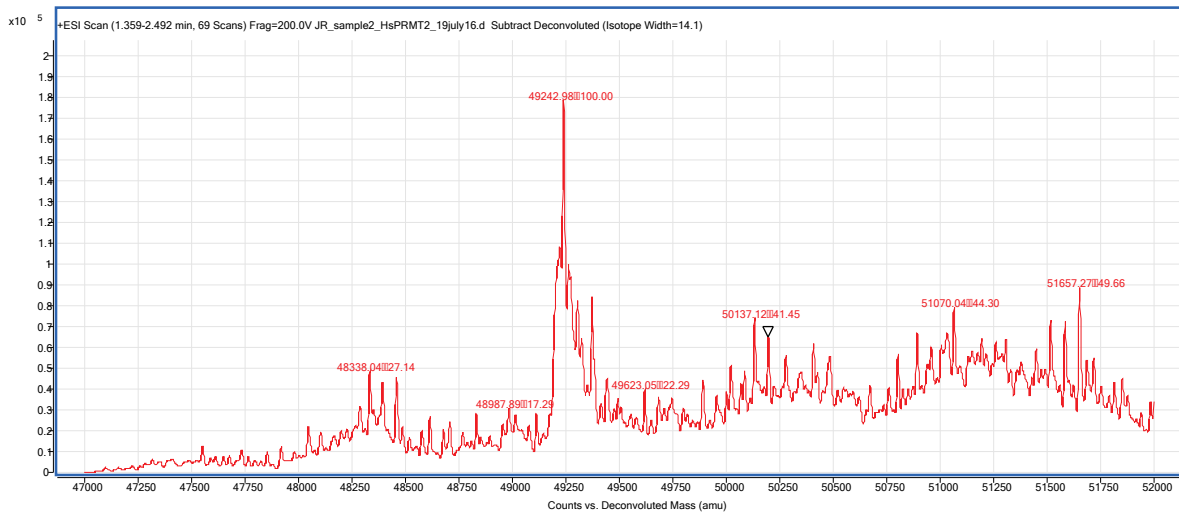
Appendix I.

Analysis of HsCARM1^{CAT} Sample by Mass Spectrometry.



The deconvoluted spectrum reveals a dominant mass species at 39915.15 Da corresponding to HsCARM1^{CAT}, which has an expected molecular weight of 39782.49 Da.

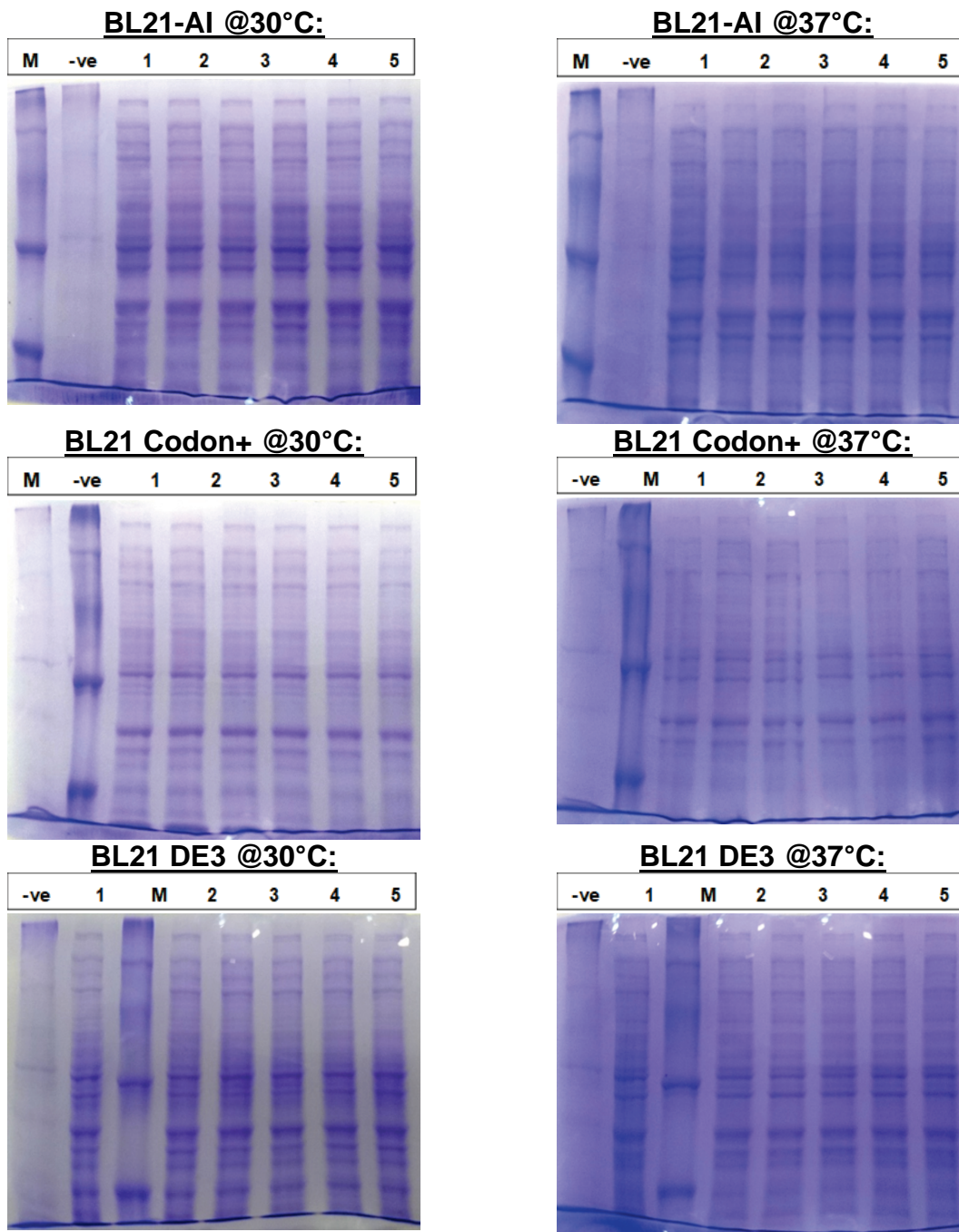
Analysis of HsCARM1^{CAT} Sample by Mass Spectrometry.



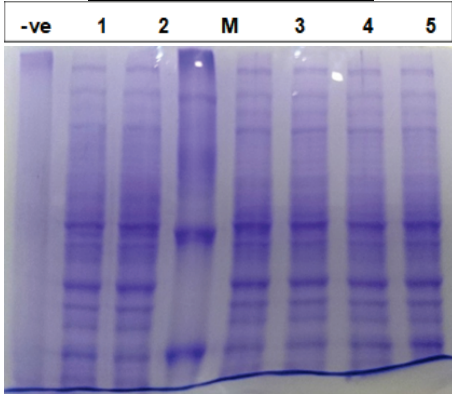
The deconvoluted spectrum reveals a dominant mass species at 49242.98 Da corresponding to HsPRMT2-1, which has an expected molecular weight of 49195.90 Da.

Appendix J.

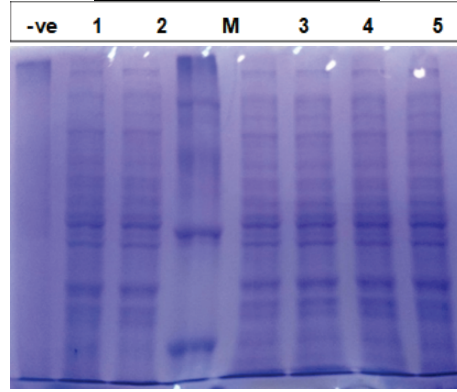
Small Scale Expression Test Analysis of GST-GAR in different E. coli Strains Conducted by NUPPA (Newcastle, UK).



BL21 plys^s @30°C:

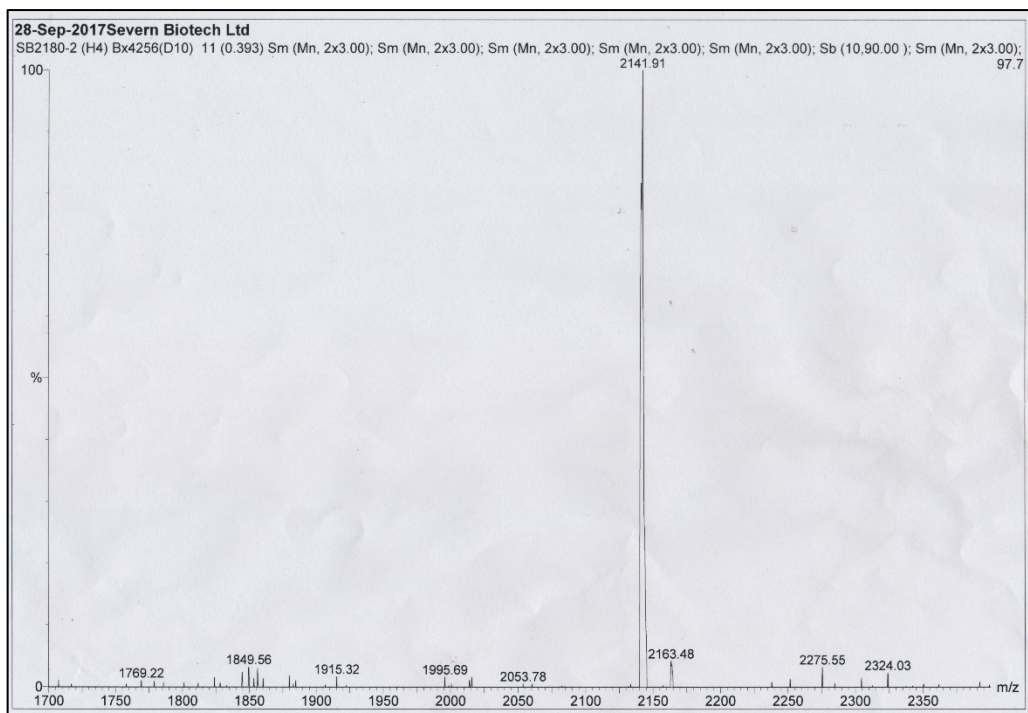
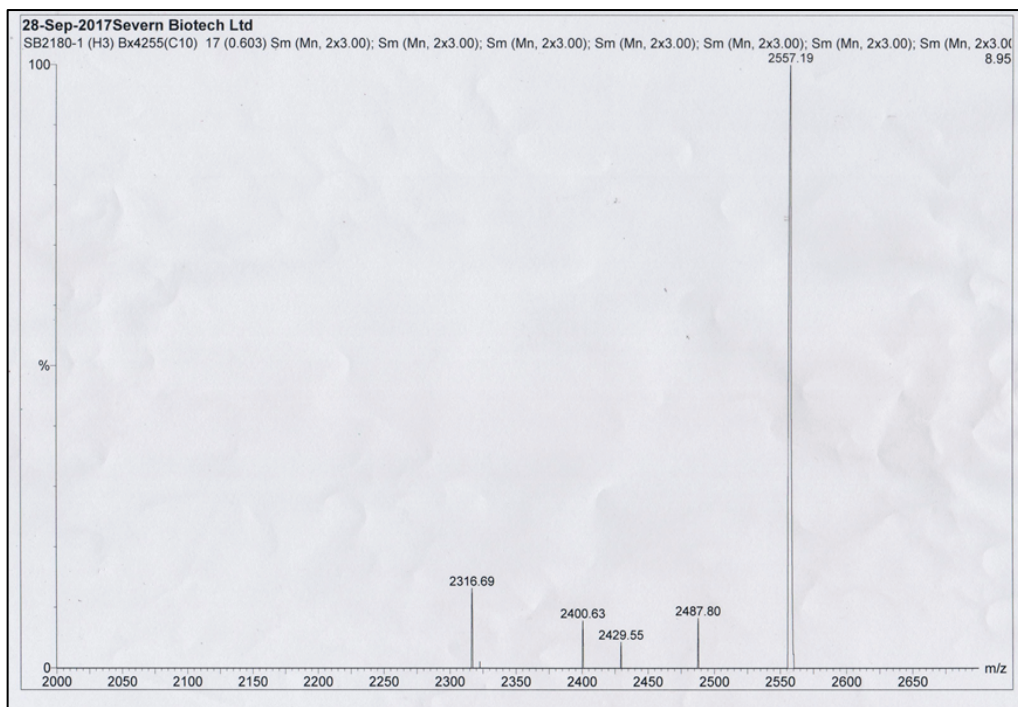


BL21 plys^s @37°C:



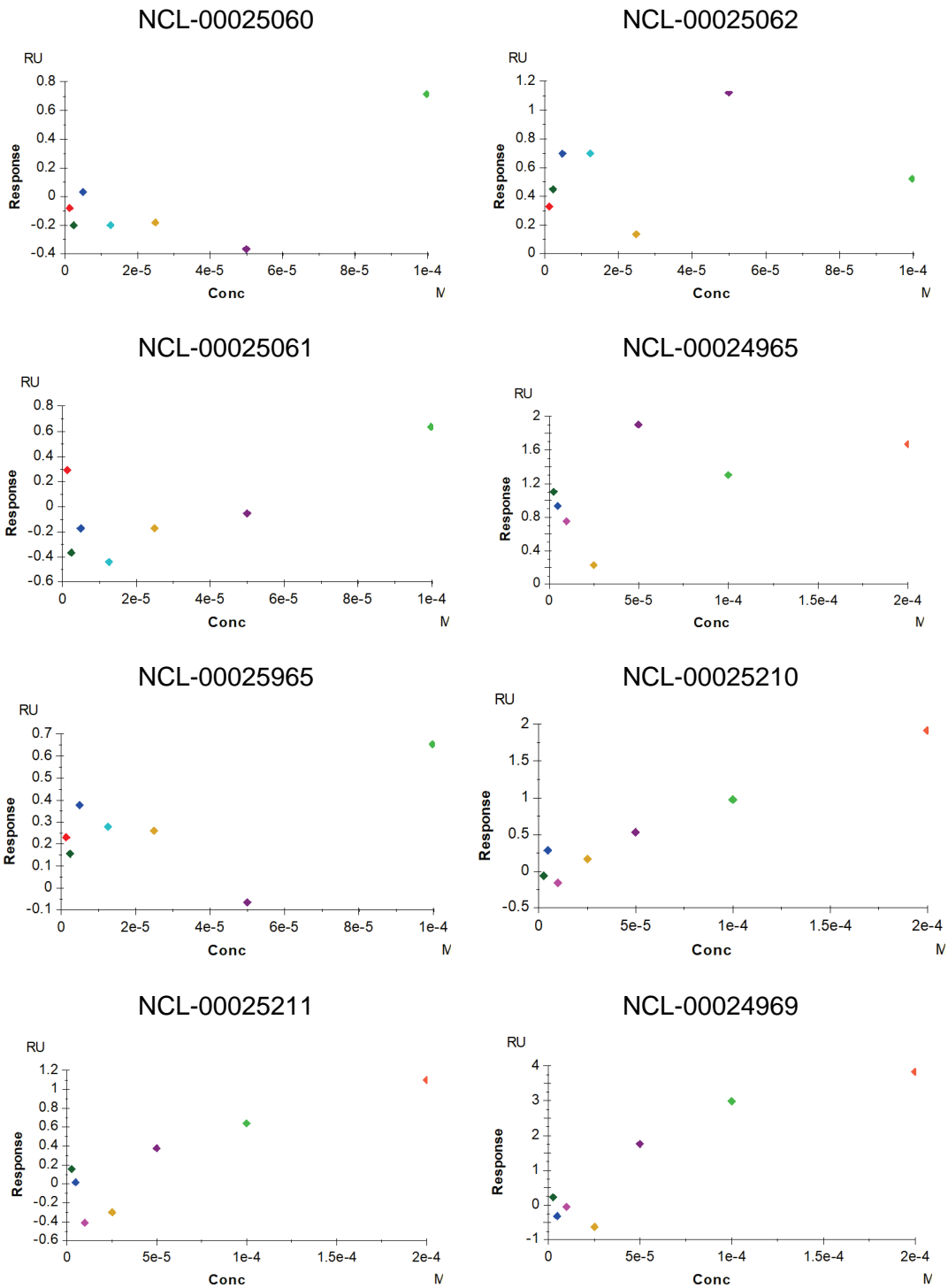
Appendix K.

MALDI-TOF Mass Spectrometry Analysis of H3/H4 peptide (circa 1mg/ml) by Severn Biotech.



Appendix L.

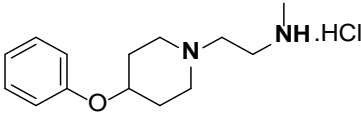
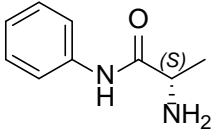
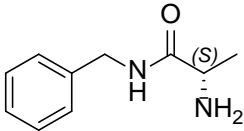
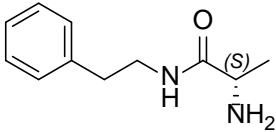
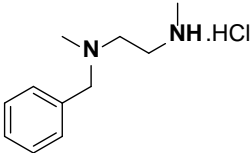
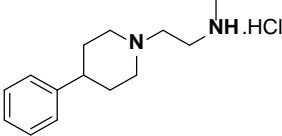
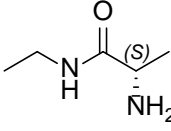
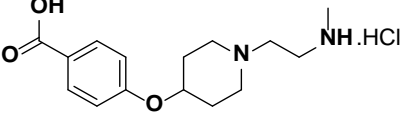
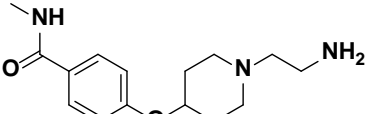
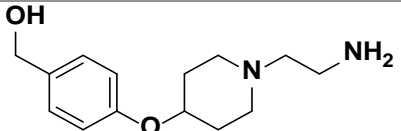
SPR Analysis of Bisubstrate Fragment Binding to HsCARM1^{CAT}.



Appendix M.

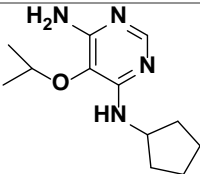
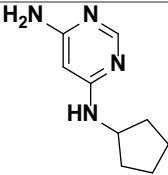
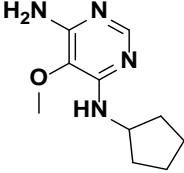
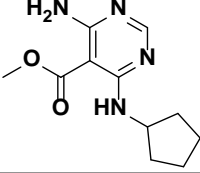
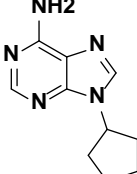
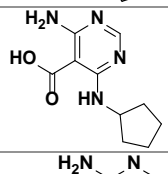
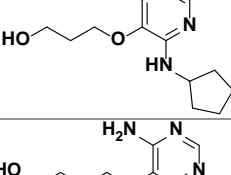
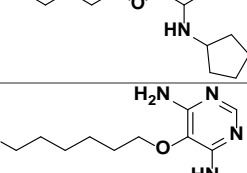
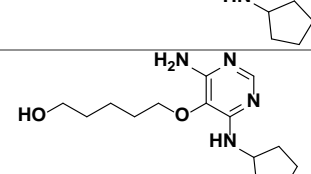
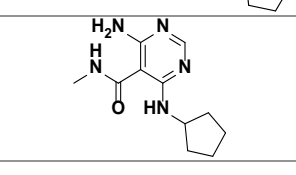

Summary of Ligands Tested Against HsCARM1^{CAT}.

For each ligand, the ID and chemical structure is shown, together with the K_d values derived from SPR experiments with at least two biological replicates with ligand concentration ranging from 0- 200 μM . If a CARM1-ligand structure could be obtained by X-ray crystallography it is indicated with a yes. K_d values are shown in μM , \pm standard deviation (nb= no binding observed; na = not analysed; / = no saturation/ K_d value).

NCL ID	Structure	K_d (SPR) [μM]	X-Ray Structure
NCL-00024056		5.18 ± 1.50	Yes
NCL-00024052		8.26 ± 0.92	Yes
NCL-00024053		36.21 ± 1.89	Yes
NCL-00024054		13.37 ± 0.76	Yes
NCL-00024055		5.18 ± 1.50	Yes
NCL-00024057		45.06 ± 15.57	No
NCL-00024058		nb	Yes
NCL-00024359		nb	No
NCL-00024360		nb	No
NCL-00024361		nb	No

NCL-00024852		3.21 ± 1.43	No
NCL-00024854		/	Yes
NCL-00024855		/	Yes
NCL-00024856		10.19 ± 3.76	Yes
NCL-00024857		2.39 ± 0.61	No
NCL-00024858		/	Yes
NCL-00024859		/	No
NCL-00024860		9.11 ± 4.76	No
NCL-00024861		5.30 ± 1.83	Yes
NCL-00024862		2.93 ± 1.06	Yes
NCL-00024969		nb	No
NCL-00025005		/	No
NCL-00025060		nb	No
NCL-00025061		nb	Yes



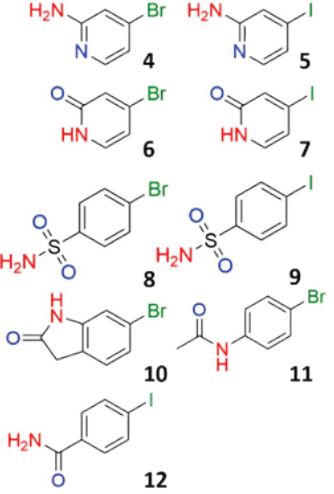
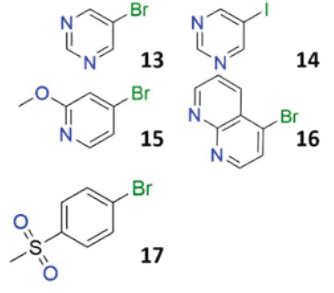
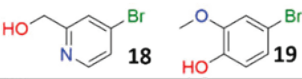
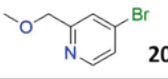
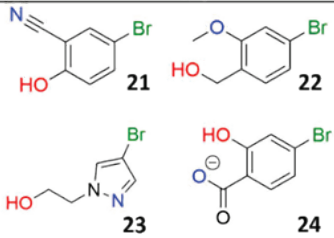
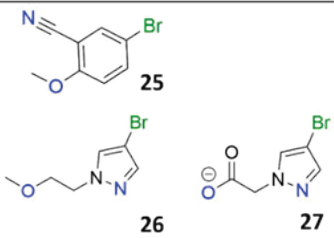
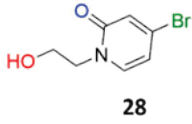
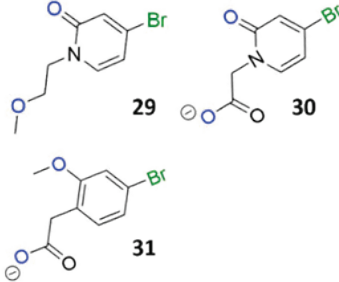
NCL-00025062		nb	No
NCL-00024965		nb	No
NCL-00024971		/	No
NCL-00024972		/	No
NCL-00025069		31.36 ± 1.64	Yes
NCL-00025070		54.75 ± 31.67	Yes
NCL-00025093		/	No
NCL-00025094		~80	No
NCL-00025210		nb	No
NCL-00025211		nb	No
NCL-00025138		/	No
NCL-00025544		na	Yes
NCL-00025547		na	Yes

NCL-00024109		~ 72	No
NCL-00024110		/	No
NCL-00024111		/	No
NCL-00024112		/	No
NCL-00024113		/	No
NCL-00024114		/	No
NCL-00024115		/	No
NCL-00024116		/	No
NCL-00024117		/	No
NCL-00024118		/	No
NCL-00024119		/	No

Appendix N.

FragLite Set.

Stock: 500 mM in 100 % DMSO. The final CARM1-crystal soak concentration was 5 mM. The figure was taken, with permission from Wood et al. (2019).

Connectivity	Donor / Acceptor	Acceptor / Acceptor
1	 1 2	 3
2	 4 5 6 7 8 9 10 11 12	 13 14 15 16 17
3	 18 19	 20
4	 21 22 23 24	 25 26 27
5	 28	 29 30 31

Appendix O.

Crystallographic Data and Refinement Statistic for HsCARM1^{CAT}-SAH-Ligand Crystals.

Collection data for the high-resolution shell are given in parentheses.

CARM1 Ligand	NCL-00024058	NCL-00025070	NCL-00025069	NCL-00024056
Space Group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Wavelength (Å)	0.97950	0.82656	0.97621	0.97953
Unit cell dimension (a, b, c) (Å)	74.996, 98.946, 206.982	74.788, 98.343, 206.800	74.513, 98.418, 206.199	74.783, 98.478, 206.879
(α, β, γ) (°)	90,90,90	90,90,90	90, 90, 90	90, 90, 90
Resolution (Å)	206.98-2.8	71.36-2.21	40.3-1.68	50.78-2.04
R_{meas}	0.23(1.17)	0.29 (2.22)	0.091 (2.90)	0.030 (0.122)
I/σ (I)	7.4 (1.8)	4.9 (0.9)	11 (0.8)	8.1 (1.1)
Completeness (%)	100 (100)	100 (100)	100 (99.9)	100 (99.8)
CC ½	0.992 (0.642)	0.989 (0.363)	0.997 (0.337)	0.996 (0.432)
Redundancy	7.3 (7.4)	7.6 (7.5)	7.4 (7.6)	5.4 (5.3)
No. Reflections (all/free)	38770/1923	152958/76225	172958/8675	98599/4437
R-factor/R-free	0.193/0.262	0.206/	0.192/0.226	0.203/0.239
Overall B-factor (Å²)	34.56	32.31	29.82	29.12
Protein (No)	10997	10997	10997	10997
Ligand (No)	84	144	168	154
Water (No)	678	218	621	650
RMSD Deviations Bonds (Å)	0.011	0.0075	0.0093	0.0043
Angles (°)	1.528	1.645	1.710	1.657
Ramachandran Favoured (%)	96.99	96.40	96.55	97.12
Ramachandran outliers (%)	0.21	0.81	0.81	0.55

CARM1 Ligand	NCL-00024858	NCL-00024855	NCL-00024856	NCL-00024861
Space Group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Wavelength (Å)	0.976250	0.979500	0.979500	0.979500
Unit Cell Dimension (a, b, c) (Å)	74.613 98.117 206.168	75.176 98.744 207.452	75.2228 98.854 208.137	74.871 98.106 207.581
(α, β, γ) (°)	90,90,90	90,90,90	90, 90, 90	90,90,90
Resolution (Å)	88.94-1.68	207.3-1.83	208.14-1.82	207.58-1.80
R_{meas}	0.158(2.767)	0.158 (2.257)	0.159 (2.132)	0.128 (1.092)
I/σ (I)	7.8 (0.9)	6.9 (0.8)	7.5 (0.9)	8.5 (1.5)
Completeness (%)	100 (98.8)	100 (99.8)	100 (100)	99.5 (98.8)
CC ½	0.995 (0.281)	0.996 (0.301)	0.997 (0.334)	0.994 (0.685)
Redundancy	7.1 (6.9)	6.4 (6.6)	6.5 (6.7)	6.6 (6.7)
No. Reflections (all/free)	173849/8780	136212/6875	139570/6786	141207/7023
R-factor/R-free	0.192/0.239	0.214/0.253	0.207/0.250	0.195/0.235
Overall B-Factor (Å²)	24.32	30.12	27.28	25.24
Protein (No.)	10997	10997	10997	10997
Ligand (No.)	156	152	152	152
Water (No.)	978	619	696	1029
RMSD Deviations Bonds (Å)	0.0085	0.0082	0.00074	0.0083
Angles (°)	1.578	1.550	1.510	1.553
Ramachandran fa- voured (%)	96.85	97.42	96.33	96.48
Ramachandran outli- ers (%)	0.0	0.13	0.22	0.22

CARM1 Ligand	NCL-00024862	NCL-00024852	NCL-00024052	NCL-00024054
Space Group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Wavelength (Å)	0.976250	0.979500	0.97693	0.97950
Unit Cell Dimension (a, b, c) (Å)	74.762	74.710	74.921	74.967
(α, β, γ) (°)	98.168	97.896	98.43	98.773
	206.228	205.458	206.75	206.77
	90, 90, 90	90,90,90	90,90,90	90,90,90
Resolution (Å)	57.21-2.10	205.46-1.89	206.75-1.74	50.78-1.97
R_{meas}	0.163 (1.517)	0.228 (2.656)	0.137 (0.627)	0.153 (0.102)
I/σ (I)	9.3 (1.8)	5.1 (0.8)	15.1 (6.6)	12.2 (6.7)
Completeness (%)	100 (100)	100 (100)	98.8 (100)	100 (100)
CC ½	0.997 (0.620)	0.994 (0.433)	0.993 (0.43)	0.995 (0.54)
Redundancy	7.4 (7.1)	6.3 (6.2)	7.3 (2.4)	7.3 (2.4)
No. Reflections (all/free)	89412/4478	121236/6001	157403/6894	110686/5897
R-factor/R-free	0.193/0.247	0.229/0.273	0.189/0.214	0.177/0.210
Overall B-factor (Å²)	25.24	28.31	24.84	27.42
Protein (No.)	10997	10997	10997	10997
Ligand (No.)	152	104	120	148
Water (No.)	704	400	678	642
RMSD Deviations Bonds (Å)	0.0143	0.0089	0.0180	0.0138
Angles (°)	1.986	1.547	1.304	1.534
Ramachandran favoured (%)	96.3	96.6	97.3	96.8
Ramachandran outliers (%)	0.21	0.73	0.11	0.22

CARM1 Ligand	NCL-00024053	NCL-00025544	NCL-00024854
Space Group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Wavelength (Å)	0.96730	0.91165	0.97625
Unit Cell Dimension (a, b, c) (Å) (α, β, γ) (°)	75.0, 98.8, 206.8 90, 90, 90	75.2, 99.3, 208.0 90, 90, 90	74.8, 98.2, 206.3 90, 90, 90
Resolution (Å)	76.8-1.95	89.77-2.88	103.36-1.84
R_{meas}	0.040 (0.098)	0.443 (4.322)	0.194 (3.394)
I/σ(I)	12.6 (6.6)	5.4 (0.8)	6.6 (0.7)
Completeness (%)	99.2 (99.2)	100 (100)	100 (100)
CC^{1/2}	0.995 (0.631)	0.990 (0.332)	0.995 (0.324)
Redundancy	7.2 (2.4)	13.6 (13.5)	7.2 (7.0)
No. Reflections (all/free)	112856/5320	36110/1892	132399/6610
R-factor/R-free	0.186/0.223	0.199/0.282	0.208/0.253
Protein (No.)	10997	10997	10997
Ligand (No.)	156	184	164
Water (No.)	678	18	522
Overall B-Factor (Å²)	24.52	53.66	28.31
RMSD Deviations			
Bonds (Å)	0.0021	0.0056	0.0073
Angles (°)	1.326	1.430	1.505
Ramachandran favoured (%)	96.9	94.1	96.7
Ramachandran outlier (%)	0.25	1.10	0.73

CARM1 Ligand	NCL-00025547	NCL-00024055
Space Group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Wavelength (Å)	0.911654	0.979402
Unit cell dimension (a, b, c) (Å) (α, β, γ) (°)	75.340, 99.077, 208.290 90, 90, 90	75.0, 98.6, 206.7 90, 90, 90
Resolution (Å)	71.89-2.29	53.76-2.21
R_{meas}	0.251 (4.285)	0.06 (0.741)
I/σ(I)	7.6 (0.8)	8.2 (1.0)
Completeness (%)	100 (100)	97.9 (100)
CC^{1/2}	0.992 (0.391)	0.997 (0.437)
Redundancy	13.8 (14.0)	8.9 (3.9)
No. Reflections (all/free)	71013/3584	105356/4820
R-factor/R-free	0.210/0.277	0.186/0.216
Protein (No.)	10997	10997
Ligand (No.)	180	154
Water (No.)	95	692
Overall B-factor (Å²)	48.03	26.73
RMSD Deviations		
Bonds (Å)	0.0077	0.0043
Angles (°)	1.702	1.648
Ramachandran outliers (%)	0.51	0.29
Ramachandran favoured (%)	96.9	95.9