

Supporting Information

Rossmann-Fold Methyltransferases:

Taking a “ β -turn” around Their Cofactor, S-Adenosylmethionine

Bhanu Pratap Singh Chouhan, Shayida Maimaiti, Madhuri Gade, and Paola Laurino*

Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa, Japan

Material and Methods:

Dataset creation for the Rossmann MTases: This dataset was created by downloading high-quality (< 2.5 Å) experimentally-solved crystal structures from PDB database (structures that are classified as Rossmann-fold) through the CATH structural classifications (v4.0.0)¹, SCOP structural classifications (version 1.75)² SAM/SAH cofactor containing protein structures from PDB³ along with the list of manually classified Rossmann structures as described in Laurino *et al.*, 2016⁴. This dataset was manually curated, and the final list of structures considered in this dataset are presented in Table S2.

Dataset creation for the non-MTase Rossmann structures and Nucleoside-cofactor containing proteins: This dataset was created by downloading all the protein structures (< 2.5 Å) that contain nucleoside-based cofactors mentioned in the Table S1 and S3. The Rossmann-fold containing proteins in this dataset were defined through the CATH structural classifications (v4.0.0). Our initial dataset consisted of ~8000 protein structures. Subsequently, CD-HIT⁵ clustering was performed at 80% sequence identity in order to remove redundancies and the final dataset was curated to ~4200 structures. These structures were then analyzed for β -turn formation around (< 4 Å) the cofactors (see *β -turn identification* section below).

Normalized B-factor analyses: We randomly picked representative structures belonging to different superfamilies of the Rossmann MTases (i.e. DNA, RNA, Protein etc.) in order to check for the main chain flexibility of the type-II' β -turn containing Gly-rich loop region. The B-factors (Temperature factors) of the main chain C_{α} atoms were considered as a measure to study the main chain flexibility. For each selected structure, the normalized B-factor value for C_{α} atoms was calculated by the following equation: $B' = B - \langle B \rangle / \sigma(B)$ ⁶. Here, $\langle B \rangle$ is the average value of all C_{α} atoms and $\sigma(B)$ is the standard deviation of the B-value (for the selected protein).

In-Silico Alanine Scanning: The type-II' β -turn residues in the Rossmann MTase dataset (Table S4) were mutated into Alanines by utilizing the Alasca program which is available from FoldX suite⁷. Herein, the difference in the Gibbs free energy change ($\Delta\Delta G$) is reported upon mutation of the residue. The numbers ($\Delta\Delta G$ values) demonstrate the importance of a residue towards the structure stability, i.e. a positive $\Delta\Delta G$ value indicates that the residue is important towards the structure stability. The $\Delta\Delta G$ values for β -turn residues in the Rossmann MTase dataset are reported in the Table S2.

β -turn identification: We utilized PROMOTIF⁸ program to identify β -turn regions in the non-MTase Rossmann structures and nucleoside-cofactor containing protein structures datasets (Table S1). Additionally, we also utilized BetaTurnTool18 program (<http://dunbrack.fccc.edu/betaturn/>) to count the occurrence of Type-II' β -turns and no significant difference was observed.

Site-directed mutagenesis: library was created by Q5 site-directed mutagenesis (NEB, E0554S) using the template pASK-IBA3+ vector ⁹. The primers were designed to introduce random mutations on the four amino acids (Ser, Gly, Ala, Gly, i, i+1, i+2, i+3) of the glycine-rich loop on the beta-turn of DNA methyltransferase M. *HaeIII*. Q5 site directed mutagenesis kit, *HaeIII* and *NcoI* restriction enzyme, cutsmart buffer purchased from the New England Biolab. We obtained approximately $\sim 10^6$ colonies (the expected diversity of the library was supposed to be 20^4), which indicate that diversity of all the amino acids is covered (Table S4).

Primers with four degenerate codon mutations

Name	Sequence
Q5m4HAE-F	nns nns ggattagatcttggcttcc
Q5m4HAE-R	nns nns aaaaagactaattaaatttgccatggagc

Wild type M. *HaeIII* protein sequence used in this study ⁹.

MGSSHHHHHHSSGLVPRGSSMANLISLFSGAGGLDLGFQKAGFRIIAANEYDKSIWKTY
 ESNHSAKLIKGDISKISSDEFKCDGIIGGPPCQSWSEGGSLRGIDDPRGKLFYEYIRILKQ
 KKPFFLAENVKGMLAQRHNKAVQEFIQEFDNAGYDVHIILLNANDYGVAQDRKRVFY
 IGFRKELNINYLPPIPHLIKPTLKDVIWDLKDNPIPALDKNKTNGNKCIPNHEYFIGSYST
 IFMSRNRVRQWNEPAFTVQASGRQCQLHPQAPVMLKVSKNLNKFVEGKEHLYRRLTV
 RECARVQGFPDDFIFHYESLNDGYKMIGNAVPVNLAYEIAKTIKSALEIRKGN

Cofactors	I	I'	II	II'	VIa1	VIa2	VIb	VIII	IV
NAD	643 (325)	9 (3)	178 (91)	26 (12)	5 (4)	4 (2)	12 (5)	243 (128)	770 (422)
NAI	130 (42)	5	43 (15)	4 (3)			4 (1)	42 (12)	180 (74)
B4P	13 (5)		10 (3)	1			1 (1)	4	15 (6)
AP5	17 (15)		18 (13)	1			11 (9)	2 (2)	46 (38)
APR	57 (8)		25 (6)	2	2 (1)	3 (2)		12 (3)	75 (11)
5AD	25 (8)		3 (2)		1 (1)			2	19 (5)
5GP	54 (10)	14 (12)	8 (5)	6 (3)	1		2	10 (1)	83 (42)
ADN	54 (14)	3 (1)	18 (7)	8 (3)			1	16 (9)	65 (20)
ADP	432 (82)	46 (14)	354 (119)	35 (14)	2 (1)	1	14 (6)	125 (20)	520 (129)
ADX	4 (1)		8 (4)				1	4 (1)	13 (5)
AGS	36 (6)	8 (4)	24 (10)	4 (1)				5 (2)	52 (23)
AHR	20	6	11	5			1	1	61
AMP	194 (35)	12 (2)	65 (13)	17 (3)	2	4	10 (4)	74 (16)	241 (55)
AMZ	2	5 (2)	2 (2)	1				2 (1)	15 (8)
APC	42 (4)		12 (3)	1			5 (1)	6	38 (11)
ATP	364 (61)	46 (9)	179 (55)	21 (6)	4	1	11 (5)	86 (21)	456 (115)
BMP	5							2	9
CDP	24 (3)	3	3 (3)	2				9 (3)	25 (3)
CTP	27	1	14 (1)	3			1	9	46 (5)
FAD	632 (57)	49 (5)	236 (16)	60 (3)	4		32	115 (7)	1000 (93)
FDA	44 (4)	8	27	2				6	92 (1)
GCP	22 (12)		28 (11)					1	22 (12)
GDD	9 (1)		1 (1)					4 (1)	13
GSP	43 (26)	3 (2)	24 (14)					5 (3)	34 (18)
GTP	179 (89)	6 (2)	72 (38)	9 (1)			1	37 (12)	181 (86)
IMP	69 (1)	1	1	4	2			3	33 (6)
M7G	14 (1)		1	3				4	9
MTA	44 (19)	5 (4)	4 (2)	8 (6)			2	5 (3)	53 (28)
NMN	9 (5)	1	4 (3)				1 (1)	2 (1)	14 (7)
NOS	4 (1)		1 (1)	1				8 (8)	13 (2)
R5P	9 (5)		1					1 (1)	7 (4)
RIP	4	5 (3)	6 (2)					1	10 (7)
SAH	286 (79)	12 (8)	78 (24)	194 (84)	2 (2)	1 (1)		49 (9)	558 (181)
SAM	202 (57)	11 (4)	51 (16)	111 (53)	11 (8)	1 (1)	4 (2)	43 (14)	349 (125)
SSA	4 (1)		1	1 (1)				7 (2)	7 (3)
U5P	37 (9)	1	13 (2)	3 (1)	3 (2)	3 (3)		9 (2)	47 (12)
UD1	32 (4)	1	26 (3)	1		3 (1)		22 (4)	42 (11)
UDP	94 (6)	2	22 (7)	10 (3)	3		3 (2)	24 (5)	146 (16)
UTP	21	1	8	3			1	10	25 (3)
XMP	20 (1)	2 (2)	1 (1)				1 (1)	2 (2)	14 (9)

Table S1. Number of observed distinct β -turn types within a 4 Å vicinity of various ribose-based cofactors (column 1 – for a detailed list see Table S3). Values within parentheses indicate the number of β -turns observed in Rossmann folds (CATH annotated) for the cofactor. Values outside the parentheses indicate the overall/total number of β -turn types observed for the cofactor. The Type-II' beta turn column is highlighted (in blue) and the observed values for NAD, FAD and SAM/SAH cofactors are also highlighted (in yellow). This supports our initial observation that the Type-II' β -turn within the vicinity of the cofactor could be a diagnostic feature of the Rossmann MTases.

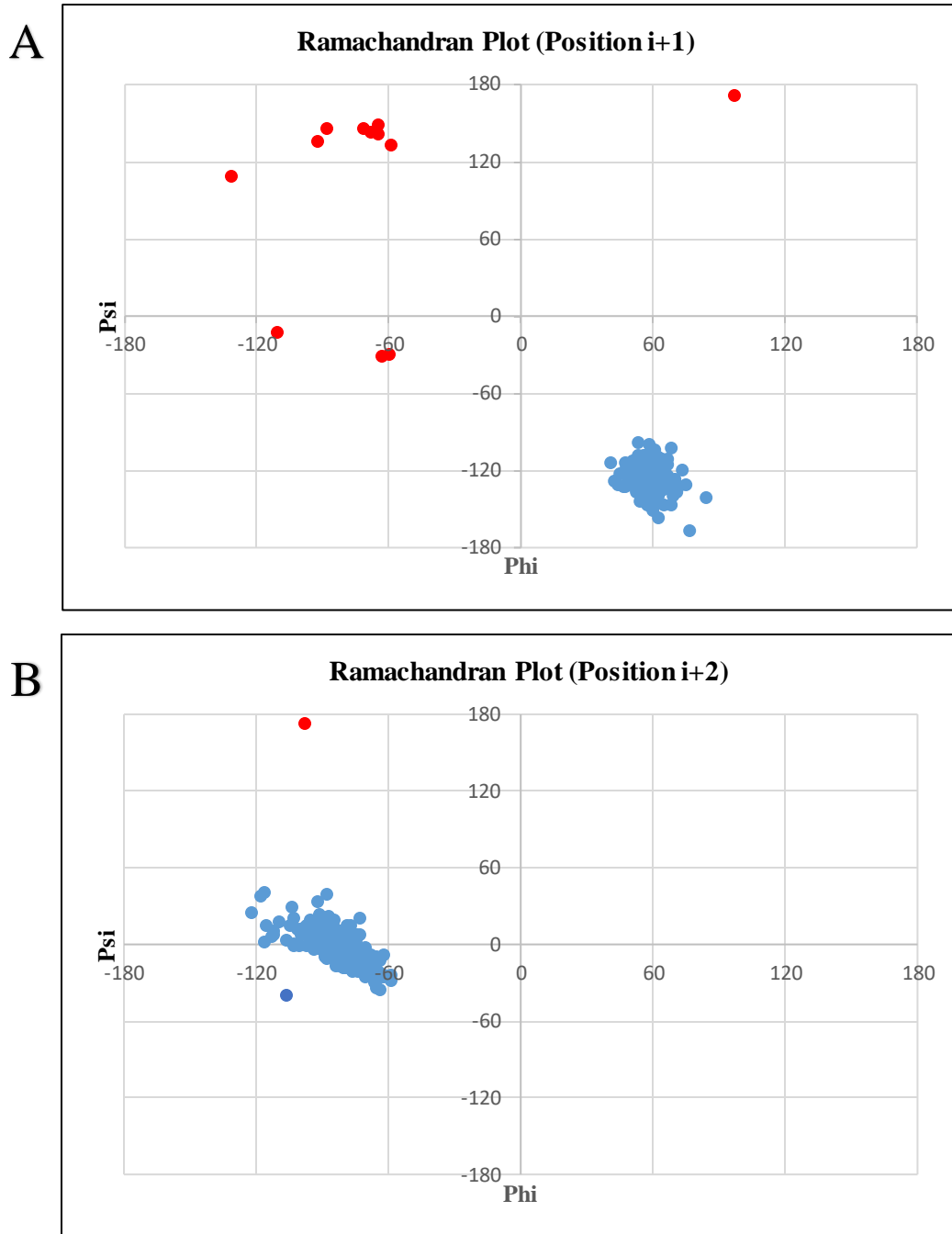


Figure S1. Ramachandran plot of position $i+1$ and $i+2$ from the Rossmann MTases dataset (Table S2). The expected torsion angle values for the two positions are **(A)** $\phi_{i+1}=60^\circ$ $\psi_{i+1}=-120^\circ$ and **(B)** $\phi_{i+2}=80^\circ$ $\psi_{i+2}=0^\circ$ and the residues that show deviation from these values are indicated in red. Refer Table S2 for list of PDB structures (12/205) that deviate from the expected torsion values (highlighted in the light brown shading) particularly in case of position $i+1$ i.e. ϕ_{i+1} and ψ_{i+1} (Panel A).

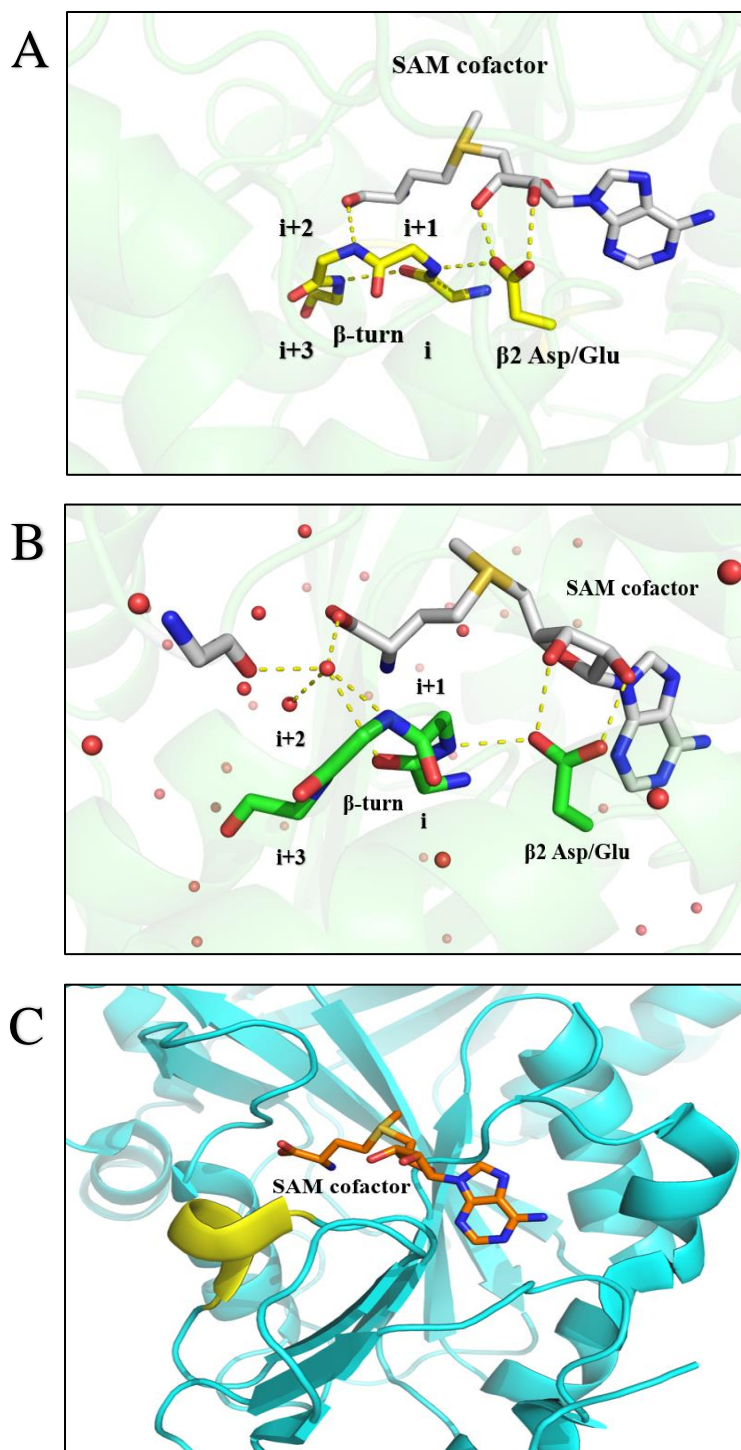


Figure S2. (A) Typical example of a Type-II' β -turn in contact with the cofactor SAM and the acidic residue from the tip of the $\beta 2$ strand (PDB: 2DPM) from *Streptococcus pneumoniae*. (B) An example of Type-II' β -turn in contact with the cofactor SAM through water molecules (PDB: 3UJ6) from *Plasmodium falciparum*. (C) An example of Type-II' β -turn absent in proximity ($\beta 1$ loop) to the cofactor - region highlighted in yellow (PDB: 5L6E) from *Homo sapiens*.

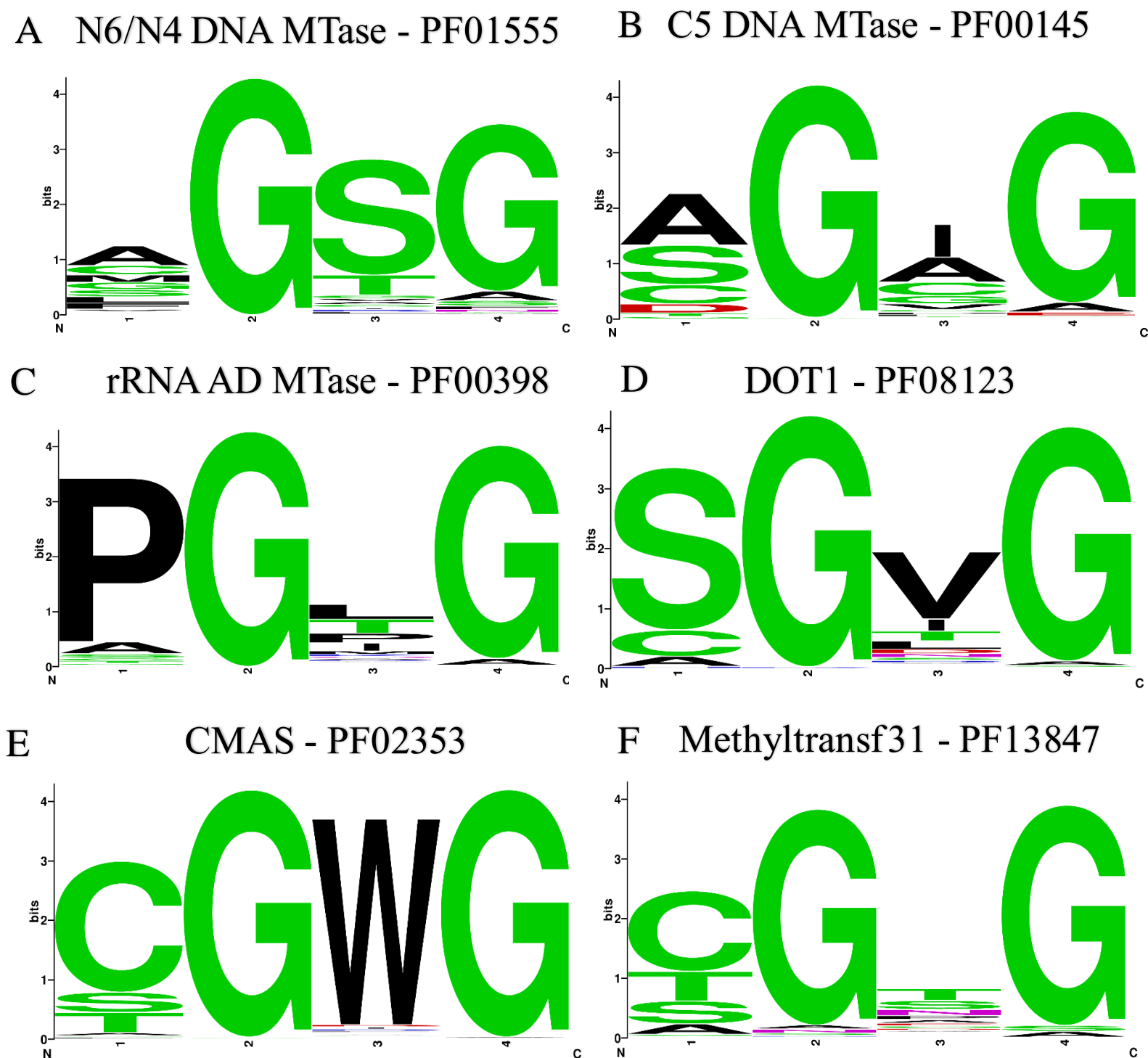


Figure S3. Weblogo analysis for Type-II' β -turn residues ('xGxG' motif) from some of the Rossmann MTase superfamilies based on the PFAM alignments. The PFAM¹⁰ families shown here belong to: **(A)** N6/N4 DNA MTase, **(B)** C5 DNA MTase, **(C)** rRNA Adenosine MTase, **(D)** DOT1 Histone MTase, **(E)** Mycolic acid synthetase and **(F)** Methyltransf_31 (contains the Arsenic SAM MTase). PFAM ID/Accession code for each family is mentioned along with the sequence logo (at the top).

PDB ID	Motif i-i+3	Begin-End	Phi (i+1)	Psi (i+1)	Phi (i+2)	Psi (i+2)	AlaScan (i)	AlaScan (i+1)	AlaScan (i+2)	AlaScan (i+3)
1AF7	ASTG	124<->127	54.2	-134.7	-78.7	14.7	0	1.46	-0.41	7.94
1DL5	GGTG	84<->87	58.9	-143.4	-82.2	0.4	2.55	2.84	1.54	3.86
1EIZ	AAPG	60<->63	-68.6	142.7	-81	-2.7	0	0	0.65	2.37
1F3L	CGTG	261<->264	57.8	-125.4	-76.5	-8.9	-0.44	3.54	-0.78	4.37
1FP2	GGTG	197<->200	64.6	-132.9	-95.8	11	3.41	3.82	-0.24	7.57
1G38	CAHG	48<->51	61.1	-115.1	-86.6	0.7	-0.69	0	-0.79	4.5
1G55	SGVG	11<->14	56.7	-128.9	-86.2	4	-2.26	2.59	1.58	8.92
1G60	MSGG	221<->224	64.5	-122.1	-76.7	-5.7	2.54	3.66	0.1	4.59
1HMY	AGLG	19<->22	60.5	-119.2	-89	-9.9	0	3.33	1.58	5.65
1I1N	SGSG	86<->89	47.6	-114	-67.1	-20.7	-1.99	2.63	-0.65	5.5
1IM8	CSRG	64<->67	49	-118.5	-83.1	-9.3	-0.03	-0.71	0.43	4.86
1JG1	TGSG	100<->103	62.8	-117.6	-82	-12.5	1.06	4.04	0.16	5.03
1JQD	GGAG	61<->64	62.7	-128.2	-84.3	1.5	4.23	3.66	0	10.15
1KYW	GGTG	209<->212	59.7	-121.7	-71.2	-11	3.43	3.33	-1.74	5.81
1L3I	CGTG	42<->45	49.3	-122.2	-74	-21.6	-0.36	2.22	-0.09	5.83
1NV8	TGSG	130<->133	52.1	-118.3	-69.4	-17.8	0.02	3.19	-0.72	5.39
1NW3	SGVG	164<->167	60.5	-127.3	-87.1	2.2	-1.68	4.74	1.98	6.29
1NW5	AGSG	251<->254	58.4	-99.8	-83.9	-14.5	0	2.44	-0.54	4.33
1O9G	CGSG	60<->63	55.6	-129.2	-71	-9.3	-0.08	2.57	-0.04	6.32
1OR8	SGTG	79<->82	48.8	-117.8	-72.4	-16.3	-2.38	4.52	-1.13	3.42
1P1C	FGMA	68<->71	55.8	-110.9	-115.2	14.8	3.41	2.47	-0.04	0
1PJS	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1Q0S	CGLL	33<->36	54.3	-125.3	-82.2	7.4	-0.52	3.85	2.06	3.88
1QAN	SGKG	39<->42	59.6	-120.5	-93.8	12.6	-0.7	3.65	1.19	4.75
1QYR	PGLA	46<->49	60.3	-121.5	-91.7	22.7	3.03	4.29	1.59	0
1QZZ	GGNG	191<->194	59.8	-135.1	-86.6	11.2	5.76	3.5	-1.6	7.39
1R18	SGSG	87<->90	41.3	-113.7	-65.8	-25.3	-1.9	1.28	-0.56	4.81
1RI4	CGKG	73<->76	69.3	-134.4	-78.5	9.4	-0.99	2.97	-0.33	6.31
1RJD	CGSD	106<->109	44.7	-130.5	-90.5	12.3	-0.78	4.51	0.67	0.32
1SQF	AAPG	255<->258	-65.1	141.1	-86.3	6.2	0	0	2.7	6.28
1TW2	GGKG	187<->190	69.8	-139.2	-79.6	6.4	4.19	3.01	-1.48	7.26
1U2Z	SGVG	400<->403	59.9	-116.8	-104.6	14.7	-1.54	4.22	2.26	6.27
1X19	GGIG	178<->181	50.8	-123	-70.8	-11.2	3.1	4.75	1.49	5.54
1XVA	CGTG	65<->68	62	-136.4	-75.3	-3.1	-1	6.08	-0.68	6.92
1ZG3	GGTG	208<->211	55.2	-129	-85	2.7	4.19	2.97	-0.17	3.78
2AN3	SGPT	580<->583	65.5	-146.2	-103.6	29.3	-2.51	4.06	2.1	0.6
2BH2	CGMG	295<->298	52.5	-130.9	-90.5	11.3	0.24	2.22	2.25	4.7
2BZG	CGKA	70<->73	51.1	-130.1	-117.8	38.4	0.07	3.41	0.57	0
2CL5	AYCG	67<->70	62.5	-109.5	-102.9	0.4	0	3.91	0.003	5.3
2DPM	VGGG	44<->47	54.5	-127.5	-78.1	-14.6	1.94	3.49	-0.73	6.7
2EFJ	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2EJT	SATG	56<->59	53.9	-98.1	-80.4	-18.8	-1.76	0	0.73	7.11
2FCA	TGKG	47<->50	66.6	-133.8	-72.7	-12	-0.77	3.79	0.3	6.5
2FK8	CGWG	82<->85	56.9	-135	-96.8	15	-0.21	3.01	2.99	3.67
2HMK	TFTG	69<->72	55.9	-131.9	-84.8	-3.2	1.85	2.52	0.76	6.84
2NXE	TGSG	129<->132	48.2	-119.6	-63.5	-12.2	-2.46	3.49	1.1	7.93
2P41	CGRG	82<->85	55.9	-130.5	-66.2	-25	0.22	2.62	-1.69	6.26
2PJD	CGAG	205<->208	71.5	-137.1	-92.8	4.1	0.33	4.62	0	3.97
2PWY	TGSG	102<->105	67.4	-111.3	-112.4	7.2	0.9	2.99	-0.28	4.64
2PXA	CGRG	82<->85	53.8	-133	-64.2	-22.3	-0.14	2.7	-2.19	7.79
2VDV	CGFG	104<->107	57.2	-129.3	-82.4	2.2	0.11	3.21	1.59	6.36
2VS1	SGVG	279<->282	62.5	-139.5	-74.8	7.1	-0.97	3.19	-0.02	8.22
2WA2	CGRG	83<->86	50.2	-128.9	-62.4	-25.2	0.01	2.2	-2.1	6.21
2WK1	VWRG	95<->98	50.8	-124.2	-75.2	-6.6	2.21	5.186	2.24	4.77
2XVM	CGNG	39<->42	56	-139.7	-78.4	7.4	0.003	3.16	0.42	6.71
2Y1X	CGSG	194<->197	55.8	-121.9	-64.6	-20.7	0.22	4.65	-0.16	4.31
2YX1	AGVG	204<->207	54.2	-134.5	-85.3	11.6	0	2.98	2.07	4.92
2ZW9	CGYD	116<->119	60.5	-150.6	-89.2	11.9	-0.7	5.56	2.02	-0.31

3AJD	AAPG	92<->95	-65.3	148.5	-78.3	-6.9	0	0	1.04	5.25
3AXT	SASG	61<->64	60.9	-103.6	-79.8	-18.5	-3.43	0	-0.6	5.58
3B3F	CGSG	194<->197	54.4	-126.9	-58.4	-23.6	-0.19	5.85	-0.72	3.92
3BGD	CGKA	65<->68	52.8	-129.7	-116.2	41.1	0.05	2.7	-0.08	0
3BWM	AYCG	67<->70	53.4	-108.5	-100.1	-1.5	0	3.38	0.25	6.17
3BXO	CGTG	47<->50	60	-125.6	-88.2	9.8	-0.17	4.11	1.46	2.94
3C0K	SYTG	229<->232	59.4	-148.3	-72.7	20.9	-1.78	-0.74	-1.16	7.43
3C3Y	VFTG	79<->82	52.3	-131.6	-87.3	2	2.8	0.7	-1.66	4.1
3CBG	VFRG	69<->72	64	-130.1	-85.4	13.9	1.02	0.97	-0.03	10.95
3DMF	AGYG	242<->245	54.7	-144.6	-82.9	5.3	0	2.41	1.25	6.13
3DUL	TLGG	67<->70	55.8	-127.3	-88.9	-0.5	0.57	0.84	3.99	4
3DXY	FGMG	72<->75	65.6	-122.1	-86.6	-0.4	3.13	3.94	0.11	8.08
3EMB	CGRG	82<->85	52	-129.2	-64.2	-25.7	-0.12	2.29	-1.79	5.24
3EVA	CGRG	82<->85	55.8	-132.5	-62	-23.6	-0.12	2.89	-0.83	6.16
3FRH	CGLN	112<->115	57.1	-138.3	-99.8	7.4	-0.69	4.28	3.1	1.28
3FTC	GGTG	39<->42	-131.5	109.3	-97.6	172.9	2.34	0.55	-0.65	-0.48
3FUU	PGLG	55<->58	64.4	-111.6	-92.7	-0.7	2.86	4.4	4.21	6.49
3FYC	LGKG	39<->42	57.4	-139.8	-81.4	-7.2	3.69	3.28	0.1	4.37
3FZG	CGFN	115<->118	61.6	-136.8	-92.2	6.5	-0.68	4.21	4.24	-2.1
3G2O	AGMG	72<->75	70.8	-131	-97.1	7.5	0	4.05	0.68	8.41
3G88	TGAG	89<->92	58.8	-124.8	-82.6	7.6	-0.36	3.95	0	4.75
3GCZ	CGRG	82<->85	59.4	-134.2	-67.4	-22.5	-0.28	3.51	-0.99	6.52
3GDH	CGVG	699<->702	49.4	-128.8	-91	11.7	-0.26	2.34	1.69	3.3
3GWZ	GGRG	191<->194	60.9	-133.1	-85	2.5	4.18	3.11	0.63	4.5
3I53	GGSG	178<->181	60.1	-142.1	-81.8	10.9	2.73	3.5	0.01	7.38
3JWG	CGEG	38<->41	58.9	-137.9	-74.4	3.6	-0.51	3.99	0.06	6.27
3TKA	FGRG	32<->35	61.5	-109.3	-112.5	11.1	3.67	2.73	-0.36	6.25
3LBF	TGSG	84<->87	58.4	-142.5	-71.1	-19.1	0.72	2.78	-0.39	5.36
3LCC	CGGG	75<->78	59.9	-128.3	-69.6	-8.4	-0.2	3.96	0.16	7.75
3LCV	CGLN	134<->137	61	-139.5	-101.7	12.6	-0.76	3.99	3.01	-6.33
3LKZ	CGRG	82<->85	57	-128.8	-66.5	-30	-0.15	3.1	-1.87	7.29
3MB5	VGSG	102<->105	61.3	-124.6	-87.9	-10.7	1.46	3.42	-0.07	4.1
3MAG	SAPG	69<->72	-88.6	145.3	-63.7	-35.7	-1.18	0	2.35	5.72
3NDI	CNDG	114<->117	55.4	-137.9	-74.7	8.3	-0.8	2.66	-0.9	7.89
3O7W	AGMD	99<->102	59.1	-140.7	-88	39.2	0	2.47	2.14	-1.3
3OFK	CAAG	53<->56	42.6	-128.9	-88.8	9.7	-0.05	0	0	3.39
3OU6	SGTG	55<->58	50	-127.5	-82.4	5.5	-2.05	1.69	1.2	4.37
3P2E	TGDG	33<->36	75.7	-131.1	-78.7	1.9	0.96	3.59	2.71	7.71
3P9C	GGVG	206<->209	66.7	-126.5	-88.3	6	3.45	3.36	0.14	8.82
3P9N	AGSG	52<->55	68.4	-103.2	-79.9	-18.4	0	3.05	0.08	4.26
3PFG	CGTC	59<->62	60.7	-127.9	-84.7	6.3	0.33	3.35	1	7.24
3PVC	FGTG	67<->70	67.2	-115	-102.7	-1.7	5.92	3.18	-0.4	7.61
3T7T	CGTG	55<->58	52.8	-137.2	-70.6	-7.4	-0.05	2.77	-0.7	7.59
3TM4	CGSG	226<->229	50.5	-126.4	-81.6	4.1	0.3	3.46	-1.84	5.6
3TKY	GGTG	212<->215	71.6	-132.3	-91.1	9.9	5.64	3.24	-0.7	5.61
3TQS	PGRG	41<->44	61.2	-125.2	-85.8	-4.4	1.54	3.95	-1.66	6.29
3UJ6	SGLG	64<->67	59	-133.6	-79.6	-3.4	-1.94	2.7	2.23	9.4
3V97	CGSG	198<->201	62.2	-143.3	-62.2	-8.3	0.27	2.99	-1.09	6.43
3VC1	CGRG	106<->109	61.5	-143.9	-81.4	-0.8	-0.3	2.93	-0.51	3.83
3WST	TGTG	73<->76	64.3	-115.5	-87	-9.2	0.29	3.4	1.04	3.94
4BLW	AGAG	43<->46	77.4	-166.4	-83.6	-16.5	0	4.01	0	3.37
4A6D	GGAG	188<->191	59.7	-127.5	-69.8	-16.1	2.54	2.95	0	3.31
4ATN	ACPG	220<->223	-59.1	132.4	-90.6	-3.3	0	-0.18	2.01	5.01
4AZW	CAQG	62<->65	58.9	-123.9	-95.7	19.5	-0.45	0	0.15	5.37
4C05	AGTG	94<->97	54.7	-123.1	-83.3	-2.9	0	2.65	-0.24	4.27
4CTJ	CGRG	82<->85	54.3	-130.5	-64.5	-25.9	-0.15	2.75	-2.02	5.51
4D7K	GGRG	187<->190	57.8	-122.9	-86.1	-9.7	5.77	3.48	-0.002	3.8
4DCM	CGNG	237<->240	57.7	-123.3	-77.3	-6	-0.43	3.03	1.02	8.29
4DMG	SYVG	223<->226	47.4	-132.3	-78.9	4.9	-0.84	0.63	1.69	7.55
4DF3	IASG	86<->89	57	-135.4	-90.9	12.4	2.13	0	-0.62	6.2

4DKJ	AGIG	18<->21	57.4	-130	-96.1	10.4	0	3.37	2.39	10.1
4EVI	GGTG	209<->212	62.8	-138.4	-91	23	6.4	2.68	-0.33	7.67
4F84	CGRG	114<->117	59.1	-123.3	-97	11.3	-1.28	5.88	-1.26	3.58
4FSD	CGTG	92<->95	61.2	-138.3	-74.2	-5.8	-0.18	7.94	0.27	7.87
4GC9	PGPG	64<->67	58	-128.8	-59.2	-28.8	2.37	3.13	2.83	3.4
4GEK	CSLG	65<->68	54.3	-119.5	-95.7	5.9	-0.32	-0.7	2.56	7.81
4HC4	AGTG	91<->94	61.9	-121.8	-79.1	-5.6	0	4.27	-0.53	3.46
4HSG	CGTG	281<->284	56.3	-128.8	-68.3	-8.2	0.32	2.99	-0.68	5.41
4KDC	CGGG	65<->68	57.7	-130.8	-84.1	5.9	0.36	3.69	5.83	7.92
4KIB	SGSG	167<->170	58.1	-119.9	-68.1	-12.5	-1.89	2.66	-0.32	7.06
4KRG	AGIG	60<->63	53.2	-135.1	-70.4	-2	0	3.94	1.49	8.35
4KVZ	CGQG	69<->72	55.8	-134.4	-84.6	18.5	-0.09	3.19	-0.09	8.67
4L7V	TGSG	84<->87	60.4	-122.9	-76.1	-21.7	0.77	3.57	-0.12	5.5
4LWP	SGTG	63<->66	55.8	-108.9	-75.4	-17.7	-2.08	4.08	-0.56	9.81
4M37	AGSG	104<->107	56.6	-125.4	-69.5	-18.5	0	2.63	-0.07	3.63
4N49	AGPG	278<->281	-71.4	146.3	-77	-12.9	0	3.79	2.81	6.82
4NEC	CGTG	48<->51	58.6	-132.4	-82.9	6.3	-0.66	3.54	1.27	8.36
4OA5	TCVG	65<->68	53.9	-116.4	-109.9	17.5	0.86	0.62	2.04	3.92
4ONJ	SGIG	489<->492	63.1	-126	-103.1	21.1	0.01	3.48	3.49	4.44
4OBW	GGSG	120<->123	-60.2	-30.1	-65.7	-28.7	-0.54	3.04	-1.83	7.68
4PGG	GGIG	207<->210	65.9	-126.1	-92.1	4.8	1.2	3.35	0.432	5.55
4PIO	SGTS	87<->90	56.2	-130.9	-79.3	-8.7	-2.27	2.7	-0.51	-0.2
4PYN	AYCG	110<->113	56.6	-107.8	-106.1	3.8	0	1.85	0.15	4.34
4QDJ	CGVG	71<->74	60.7	-131.2	-92.8	19	-0.18	3.32	1.68	3.18
4QTU	CGSG	56<->59	54.1	-129.1	-93.6	10.5	-0.27	3.94	0.03	5.93
4YMG	CYSG	74<->77	51.4	-112.3	-96.9	-1.6	-0.001	1.09	-0.2	6.56
4R05	CGRG	82<->85	54.8	-124.6	-65.1	-33.4	0.1	2.7	-1.07	3.79
4RTS	VGAG	36<->39	48.7	-129.3	-70.2	-25.5	1.5	2.36	0	2.5
4RVG	SNTG	115<->118	60.2	-134.2	-69.9	-11.7	-2.11	0.11	-0.76	6.38
4RX1	TGDG	33<->36	66.8	-124.3	-74.3	-12	0.93	4.72	0.4	6.87
4X7U	VWRG	86<->89	52	-118.8	-82.5	-6.9	2.23	4.03	0.46	5.19
5E72	MGAG	188<->191	54.7	-129.7	-78.2	-1.8	2.71	3.74	0	4.63
5BSZ	CGTG	54<->57	59.8	-130.8	-72.9	-6.9	0.37	3.12	-0.26	5.91
5BW4	TGDA	39<->42	73.6	-119.2	-80.7	-8.8	1.49	3.92	1.57	0
5CCB	TGSG	112<->115	63.3	-113.5	-113	6.6	1.03	3.33	-1.6	2.87
5CVD	AGIG	70<->73	56.1	-135.7	-78.6	-2.2	0	4.13	2.74	6.73
5D4U	VYLW	55<->58	-76.2	127.3	-111.0	-13.2	2.26	2.41	4.61	5.24
5DLY	CGQG	69<->72	56.5	-137.7	-84.7	14	0.33	3.28	-0.61	6.83
5DNK	CAAG	80<->83	52.6	-128.5	-87.7	6.1	-0.27	0	0	8.58
5E9W	CGKG	206<->209	58.1	-139.7	-89.5	9.3	-0.49	2.89	-0.86	5.91
5EGP	CGIG	55<->58	64.8	-126.6	-98.3	13.3	-0.64	3.36	2.01	5.24
5EOV	ASTG	91<->94	-67.0	-49.5	-107.3	169.5	0	-0.55	0.76	1.55
5ERG	TGSG	119<->122	53.1	-133.8	-84.8	4.3	0.61	3.32	-1.39	4.06
5F8C	CGSG	180<->183	55	-134.2	-68.7	-8.6	0.31	2.97	-0.06	5.85
5F2K	VATG	65<->68	50.4	-125.9	-82.6	-8.6	1.23	0	0.74	4.62
5FA8	CGDG	37<->40	56.9	-129.4	-79.1	6	0.41	2.61	2.24	5.49
5FTW	CSTG	102<->105	47.3	-132.9	-79	11.7	0.12	0.68	-0.44	8.28
5FUB	CGTG	124<->127	59.7	-123.2	-69.3	-10.1	0.49	4.2	-0.47	4.92
5GM1	CGTG	86<->89	64.2	-135.6	-99.8	11.4	0.22	2.93	-0.18	6.98
5GOZ	CGRG	82<->85	59.7	-119.9	-82.5	-17.1	0.45	3.05	-1.48	7.13
5GWX	TGTG	68<->71	55	-133.7	-84.8	8.1	-2.36	3.02	1.56	7.38
5HJK	AGVI	192<->195	60.3	-135.7	-92.3	33.5	0	2.74	2.45	2.22
5HOQ	VRFG	78<->81	55.9	-128.7	-87.3	0.4	2.77	0.33	2.21	6.29
5ICC	GGTG	196<->199	68.9	-147.5	-86.4	16.9	5.37	2.87	-0.82	7.65
5JDY	CGFG	48<->51	59	-135.2	-78.4	4.2	-0.36	2.72	2.23	5.13
5JGL	CGIG	55<->58	62.8	-122.8	-89.5	13.4	-0.58	3.3	0.06	6.17
5K8V	CGSG	194<->197	57.7	-120.1	-76.3	-10.4	0.25	6.56	-1.53	2.64
5KPC	CGHG	136<->139	84.7	-140.5	-81.3	9.9	-0.52	2.72	-0.05	7.9
5KVA	VYTG	102<->105	50	-127.9	-89.2	-6.1	2.62	2.18	0.6	3.75
5L6D	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

5LOG	VFTG	68<->71	56.7	-126.8	-84.9	4.7	2.17	3.71	0.84	9.74
5M58	CGSG	48<->51	58.3	-140.4	-85.2	8.7	-0.15	2.66	-0.21	6.34
5MGZ	CGSG	48<->51	61.7	-136.6	-90.8	16.2	-0.11	2.72	1.02	6.9
5MPT	AGTG	1993<->1996	51.1	-119.8	-77.5	-3.4	0	2.42	-0.58	5.55
5N5D	TFGG	66<->69	44.9	-122	-91.8	4.8	-0.5	3.07	0.29	6.87
5O4H	AYLS	50<->53	-92.2	135.9	-106	-40.2	0	2.24	3.15	-0.57
5T6B	CNDG	114<->117	56.6	-134.3	-77.3	14.1	-0.73	0.5	-3.2	6.05
5THY	AGTG	206<->209	62.1	-133.4	-82.6	-6.5	0	3.45	-0.09	6.42
5TYQ	CGSG	132<->135	56.1	-122.3	-68.6	-23.1	-0.52	2.9	-1.1	8.23
5WP4	AGIG	62<->65	63	-156.8	-73.7	-6.4	0	2.82	1.1	6.6
5X62	CGTG	147<->150	70	-126.7	-93.8	-3.2	0.26	3.83	-0.56	3.61
5XLX	CSSG	110<->113	57.3	-134	-81.4	6.6	0.07	-1.15	-0.16	7.49
5XVK	SGPT	64<->67	57.5	-145.9	-93.7	16.5	-2.55	3.58	2.16	-0.08
5YJF	SGPT	64<->67	58.1	-146.3	-92.1	10	-0.29	3.65	2.34	-1.71
6B3A	CGDG	281<->284	61.8	-120.5	-93	7.1	0.07	3.46	2.03	5.04
6B92	GASC	112<->115	-63.5	-31.5	-79.2	-17.4	3.43	0	-0.9	-0.05
6C5B	AGYG	183<->186	61.3	-134.5	-84.9	4.2	0	2.82	1.67	5.86
6C8R	CSTG	78<->81	58.1	-141.1	-116.1	2.5	-0.02	-0.47	0.23	6.27
6FDF	SGIG	14<->17	58.2	-133.6	-93.8	13.9	-2.4	3.5	3.21	9.68

Table S2. Manually curated Rossmann MTases dataset. This table highlights a list of PDB structures constituting the dataset along with the observed torsion angle values (ϕ, ψ) for the positions $i+1$ and $i+2$ for the type-II' turn. Structures where a β -turn was not observed are highlighted in yellow and the structures with noted deviation in torsion angle values (from the expected values) are highlighted in light brown. Change in Gibbs free energy ($\Delta\Delta G$) calculated by Alascan (FoldX suite) for all the turn residues (position $i, i+1, i+2, i+3$) are also shown (highlighted in blue).

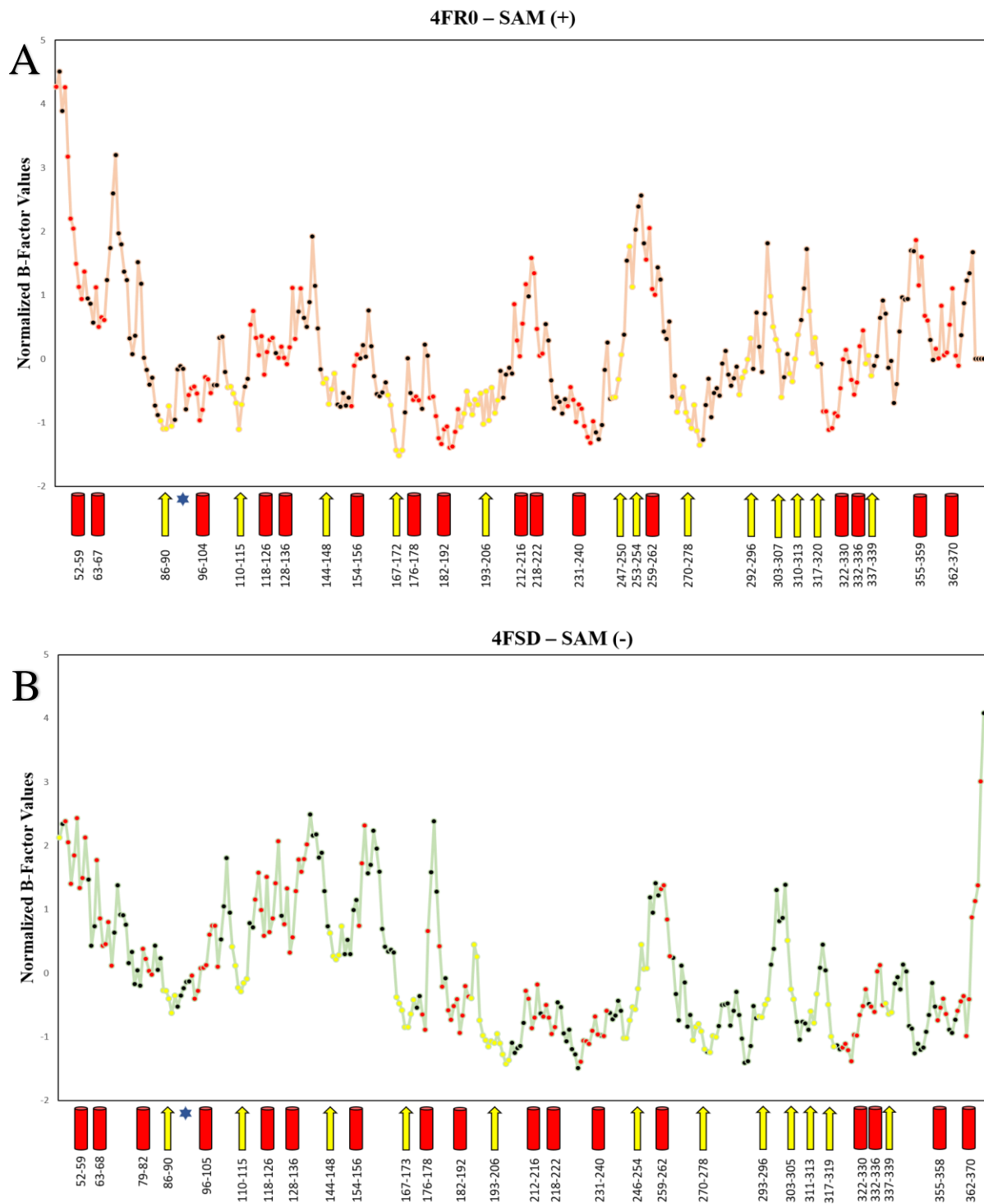


Figure S4. (A) Normalized Crystallographic $C\alpha$ B-factor plot for the randomly picked ArsMarsenic(III) S-adenosylmethionine methyltransferase with As(III) from *Cyanidioschyzon sp.* bound to S-Adenosyl methionine (PDB: 4FR0). (B). SAM absent (PDB: 4FSD). Secondary structure assignment is based on DSSP classification, β -strands on the curve are highlighted in yellow, α -helices in red and loops in black (residue positions are indicated along the X-axis) and the β -turn containing “Glycine-rich” loop region has been highlighted (between β -strand 86-90 and α -helix 96-104).

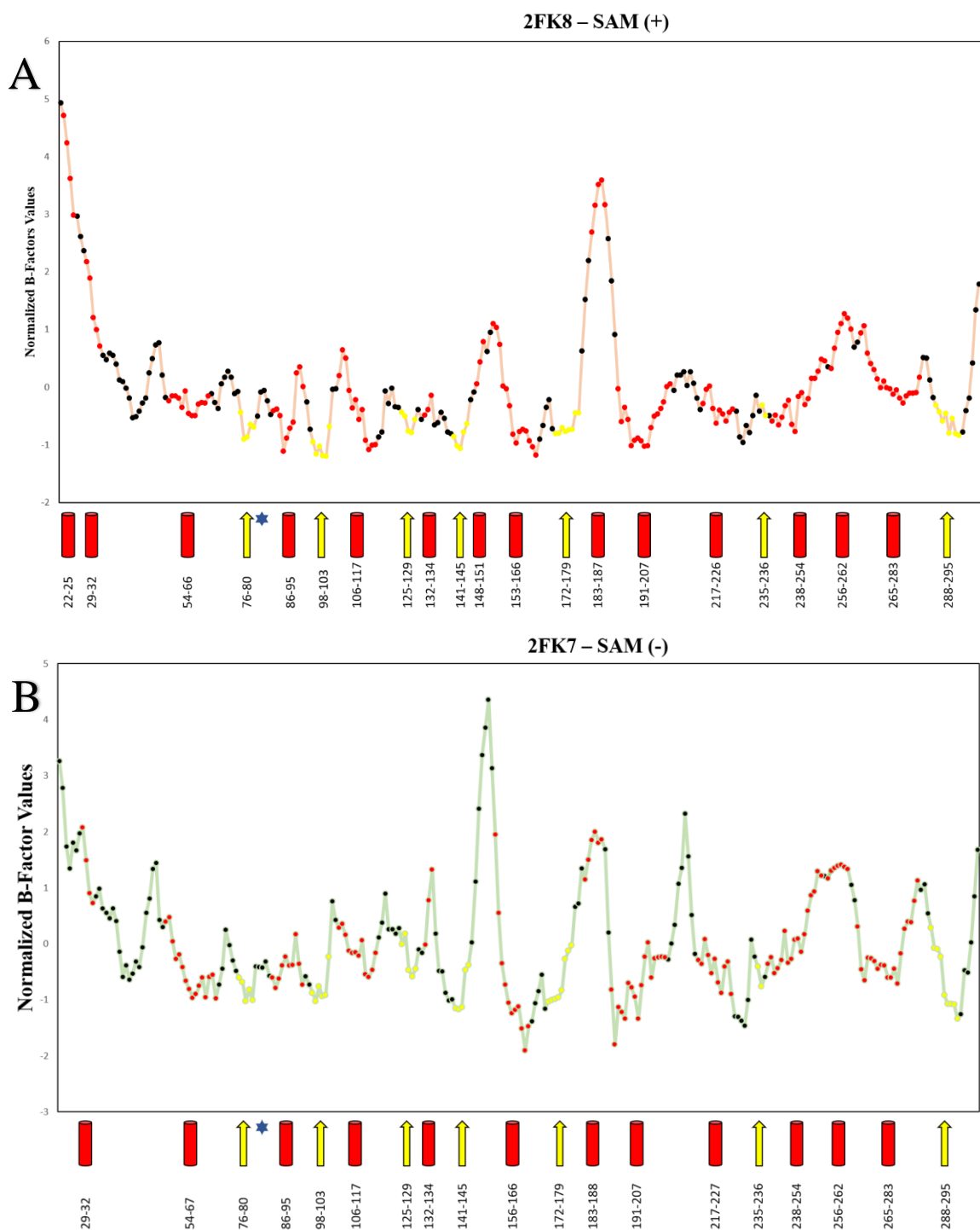


Figure S6. (A) Normalized Crystallographic C α B-factor of the Hma (MmaA4) from *Mycobacterium tuberculosis*, bound to SAM (PDB: 2FK8). (B) SAM absent (PDB: 2FK7). Secondary structure assignment is based on DSSP classification, β -strands on the curve are highlighted in yellow, α -helices in red and loops in black (residue positions are indicated along the X-axis) and the β -turn containing “Glycine-rich” loop region has been highlighted (between β -strand 76-80 and α -helix 86-95).

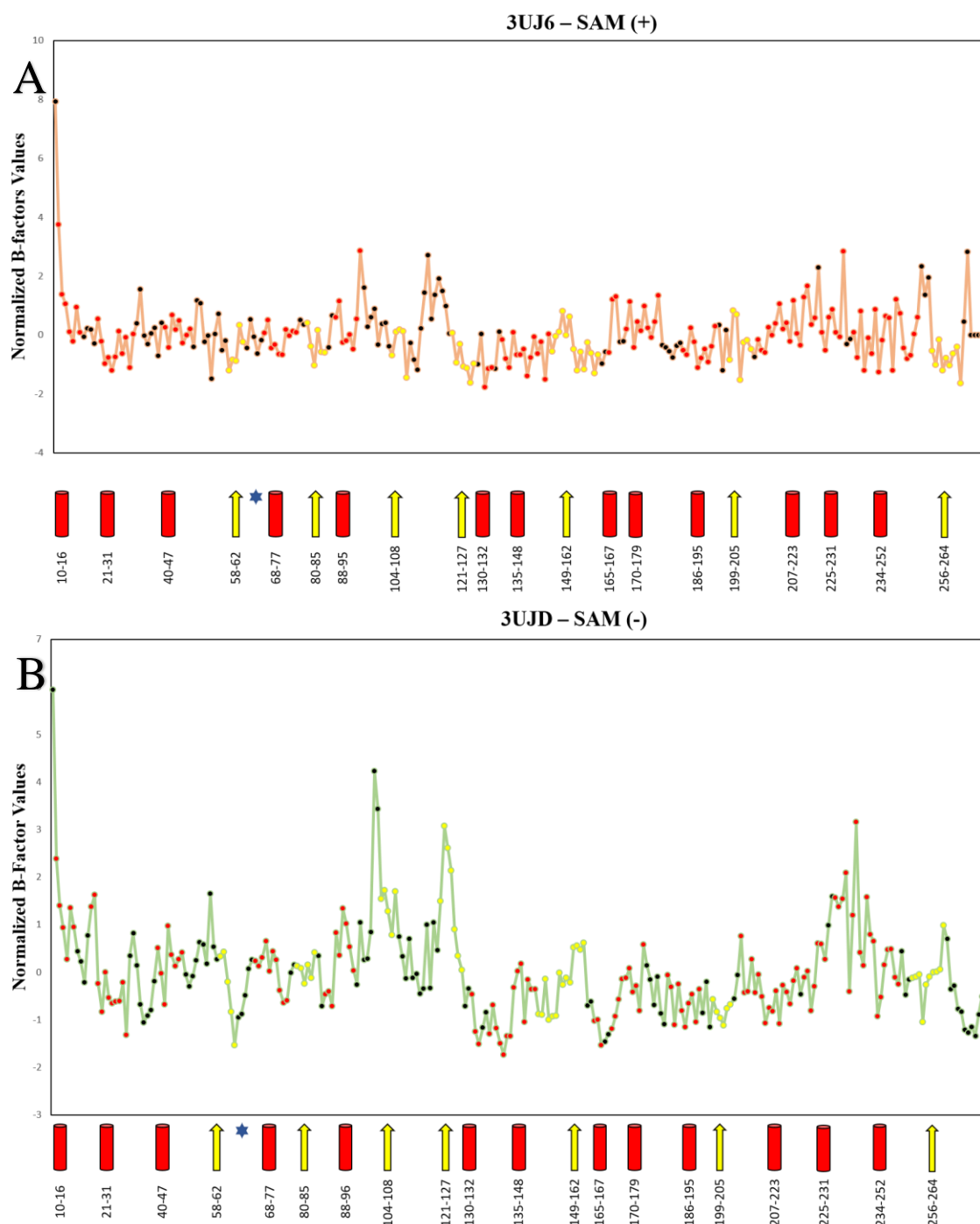


Figure S7. (A) Normalized Crystallographic α B-factor of Phosphoethanolamine methyltransferase bound to SAM from *Plasmodium falciparum* (PDB ID = 3UJ6). **(B)** SAM absent. Secondary structure assignment is based on DSSP classification, β -strands on the curve are highlighted in yellow, α -helices in red and loops in black (residue positions are indicated along the X-axis) and the β -turn containing “Glycine-rich” loop region has been highlighted (between β -strand 58-62 and α -helix 68-77).

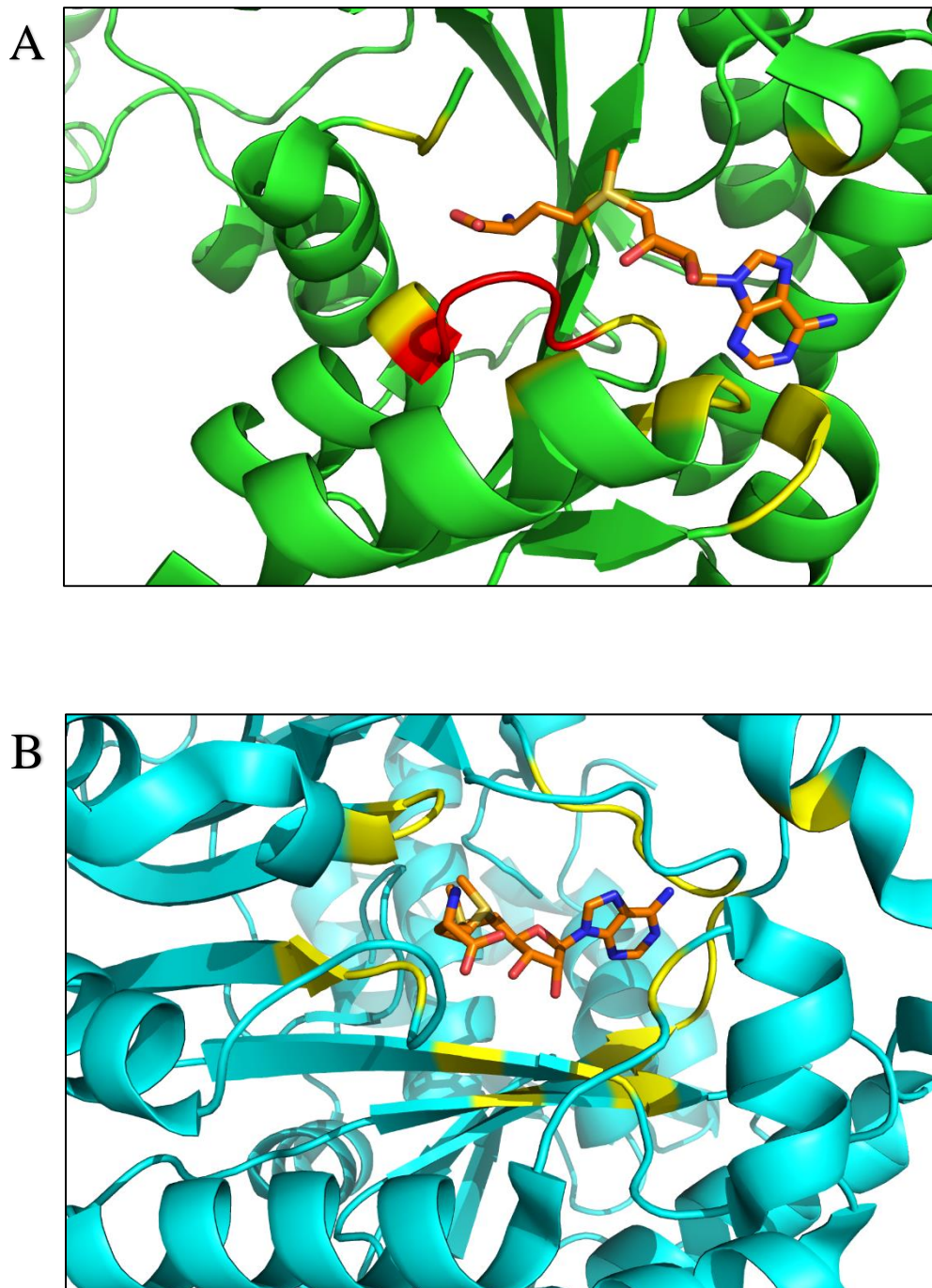


Figure S8. An example of a Type-II' β -turn (from the Gly-rich loop region - shown in red) in vicinity of the cofactor SAM from crystal structure of a DNA MTase MboIIA (M.MboIIA) from the bacterium *Moraxella bovis* (PDB: 1G60) with yellow regions indicating vicinity within 4 Å of the cofactor SAM. **(B)** An example of cofactor SAM in a pocket wherein no Type-II' β -turn formation is observed in a 4 Å vicinity from the crystal structure of HemN, a radical SAM enzyme from *Escherichia coli* (strain K12) (PDB: 1OLT).

Abbreviation	Cofactor complete name	Number of structures (overall) < 2.5 Å	Number of structures (Rossmann - CATH) < 2.5 Å
NAD	Nicotinamide Adenine Dinucleotide	1035	509
NAI	1,4-Dihydronicotinamide Adenine Dinucleotide	185	58
B4P	Bis (Adenosine)-5'-Tetraphosphate	14	4
AP5	Bis (Adenosine)-5'-Pentaphosphate	53	28
APR	Adenosine-5-Diphosphoribose	53	11
5AD	5'-Deoxyadenosine	23	5
5GP	Guanosine-5'-Monophosphate	79	25
ADN	Adenosine	5	3
ADP	Adenosine-5-Diphosphate	1195	284
ADX	Adenosine-5-Phosphosulfate	14	9
AGS	Phosphothiophosphoric Acid-Adenylate Ester	60	12
AHR	Alpha-L-Arabinofuranose	39	0
AMP	Adenosine Monophosphate	462	87
AMZ	Aminoimidazole 4-Carboxamide Ribonucleotide	14	5
APC	Diphosphomethylphosphonic Acid Adenosyl Ester	89	15
ATP	Adenosine-5'-Triphosphate	704	117
BMP	6-Hydroxiuridine-5'-Phosphate	49	0
CDP	Cytidine-5'-Diphosphate	34	7
CTP	Cytidine-5'-Triphosphate	77	19
FAD	Flavin-Adenine Dinucleotide	1512	214
FDA	Dihydroflavine-Adenine Dinucleotide	70	2
GCP	Phosphomethylphosphonic Acid Guanylate Ester	34	15
GDD	Guanosine-5'-Diphosphate-Alpha-D-Mannose	11	1
GSP	5'-Guanosine-Diphosphate-Monothiophosphate	48	26

GTP	Guanosine-5'-Triphosphate	292	67
IMP	Inosinic Acid	95	16
M7G	7N-Methyl-8-Hydroguanosine-5'-Diphosphate	14	2
MTA	5'-Deoxy-5'-Methylthioadenosine	46	22
NMN	Beta-Nicotinamide-Ribose Monophosphate	28	8
NOS	Inosine	13	7
R5P	Ribose-5-Phosphate	11	3
RIP	Ribose (Pyranose Form)	18	8
SAH	S-Adenosyl-L-Homocysteine	659	189
SAM	S-Adenosylmethionine	331	85
SSA	5'-O- (N- (L-Seryl) -Sulfamoyl) Adenosine	5	1
U5P	Uridine-5'-Monophosphate	71	9
UD1	Uridine-Diphosphate-N-Acetylglucosamine	59	12
UDP	Uridine-5'-Diphosphate	318	43
UTP	Uridine 5'-Triphosphate	42	3
XMP	Xanthosine-5'-Monophosphate	16	4

Table S3. List of cofactors considered in this study, their abbreviation (according to PDB), complete cofactor name (according to PDB), list of structures per cofactor in our initial dataset and number of Rossmann structures per cofactor in our dataset.

	Ser9	Gly10	Ala11	Gly12
Colony 1	Gly	Arg	Leu	Asn
Colony 2	Val	Val	Leu	Ser
Colony 3	Pro	Ala	Ile	Tyr
Colony 4	Gly	Gly	Thr	Ser
Colony 5	Asn	Glu	Asn	Tyr
Colony 6	Gly	His	Ser	Cys
Colony 7	His	Leu	Gly	Phe
Colony 8	Ala	Pro	His	Thr
Colony 9	Ala	Gln	Asn	Tyr
Colony 10	Pro	His	Lys	Leu

Table S4. List of representative colonies of naïve library of *M. HaeIII* which shows coverage of amino acid diversity. Random 10 colonies were picked from the naïve library of *M. HaeIII* and colony PCR was performed and further sequenced by the sanger sequencing, which shows coverage for all amino acids.

Supporting References:

- (1) Dawson, N. L., Lewis, T. E., Das, S., Lees, J. G., Lee, D., Ashford, P., Orengo, C. A., and Sillitoe, I. (2017) CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 45, D289–D295.
- (2) Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol.* 247, 536–540.
- (3) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- (4) Laurino, P., Toth-Petroczy, A., Meana-Paneda, R., Lin, W., Truhlar, D. G., and Tawfik, D. S. (2016) An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol.* 14, e1002396.
- (5) Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 22, 1658–1659.
- (6) Yuan, Z., Zhao, J., and Wang, Z. X. (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.* 16, 109–114.
- (7) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–388.
- (8) Hutchinson, E. G., and Thornton, J. M. (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* 3, 2207–2216.
- (9) Rockah-Shmuel, L., and Tawfik, D. S. (2012) Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res.* 40, 11627–11637.
- (10) Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–285.