DNA methyltransferase DNMT3A forms interaction networks with the CpG site and flanking sequence elements for efficient methylation

Michael Dukatz, Marianna Dittrich, Elias Stahl, Sabrina Adam, Alex de Mendoza, Pavel Bashtrykov, & Albert Jeltsch*

Supplemental information

Supplemental Figures

Supplemental Figure 1. Exemplary images of the purified DNMT3A proteins.

Supplemental Figure 2. +2/+3 flank preferences for different bases at the +1 flank position.

Supplemental Figure 3. Detailed outer flank sequence preference profiles of DNMT3A mutants.

Supplemental Figure 4. Exemplary data of the fits of CpG and non-CpG methylation of WT and selected mutants.

Supplemental Tables

Supplemental Table 1. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the CpG substrate.

Supplemental Table 2. Correlation of flanking preference profiles of CG methylation.

Supplemental Table 3. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the outer flank substrates.

Supplemental Table 4. Correlation of outer flank flanking preference profiles in CpG methylation.

Supplemental Table 5. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the CpN substrate.

Supplemental Table 6. Correlation of flanking preference profiles of CpN methylation.

Supplemental Figure 1. Exemplary images of the purified DNMT3A proteins. Shown are Coomassie BB stained SDS-polyacrylamide gels.
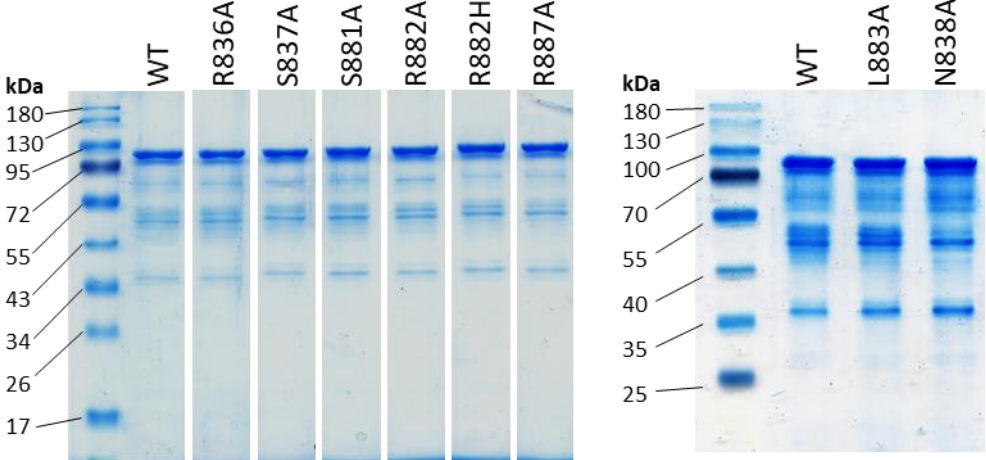
Supplemental Figure 2. +2/+3 flank preferences for different bases at the +1 flank position.
Preferences for all 16 dinucleotides at the +2/+3 flank positions were determined for each possible base at the +1 flank site.
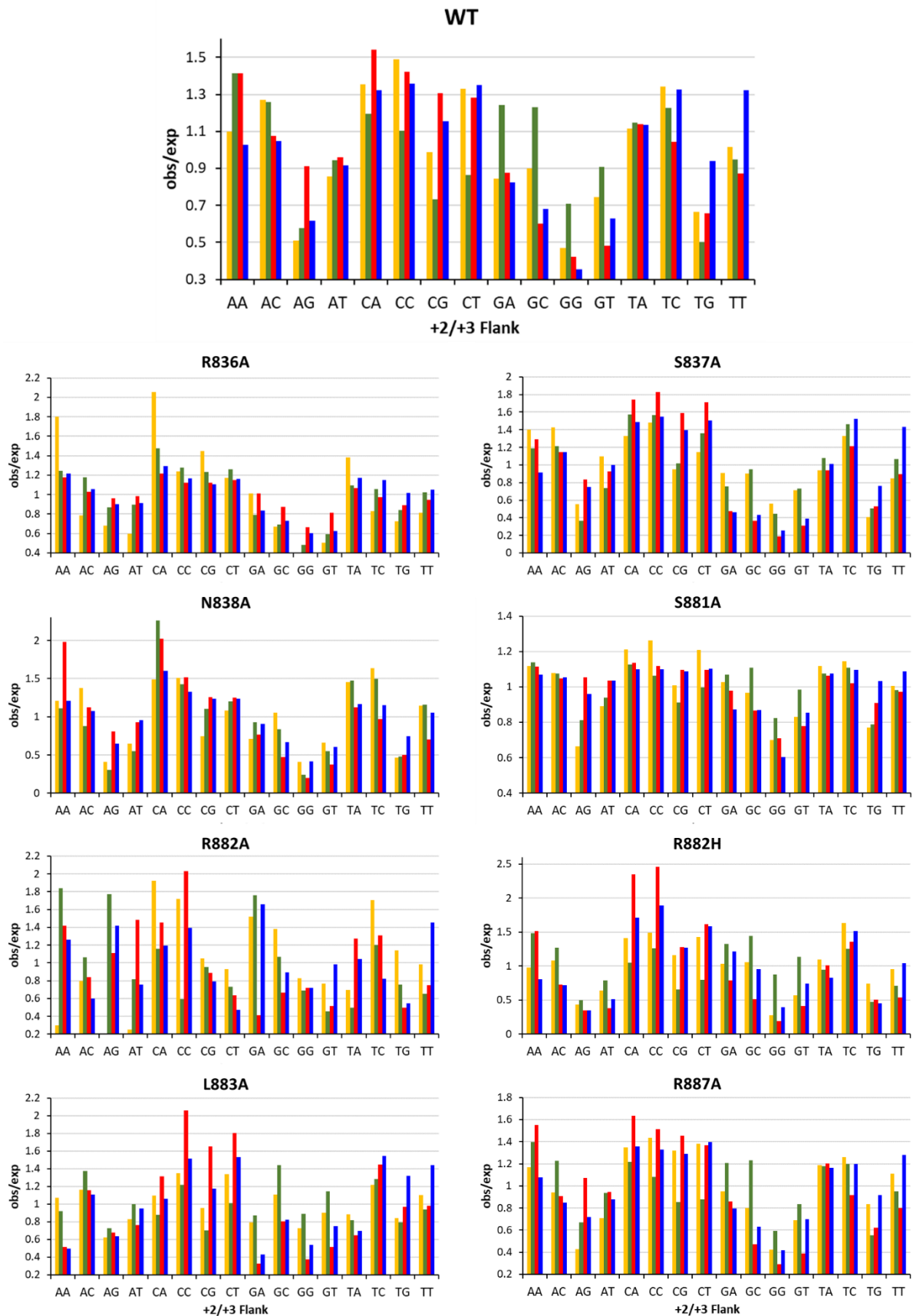
Supplemental Figure 3. Detailed outer flank sequence preference profiles of DNMT3A WT and mutants. Shown are -8 to -4 and +4 to +8 profiles of the obs/exp values of bases in the methylated sequences.

WT
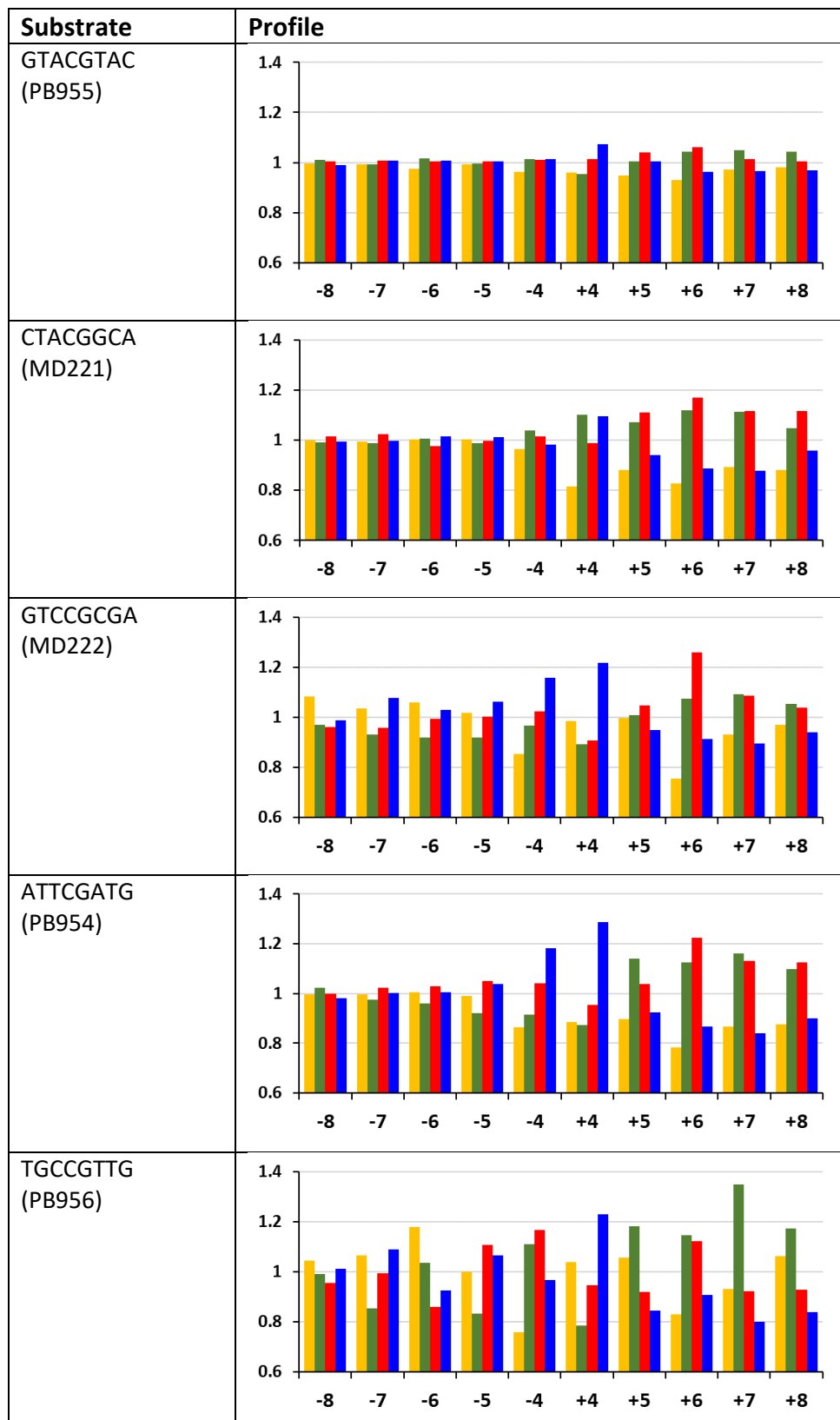
Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	
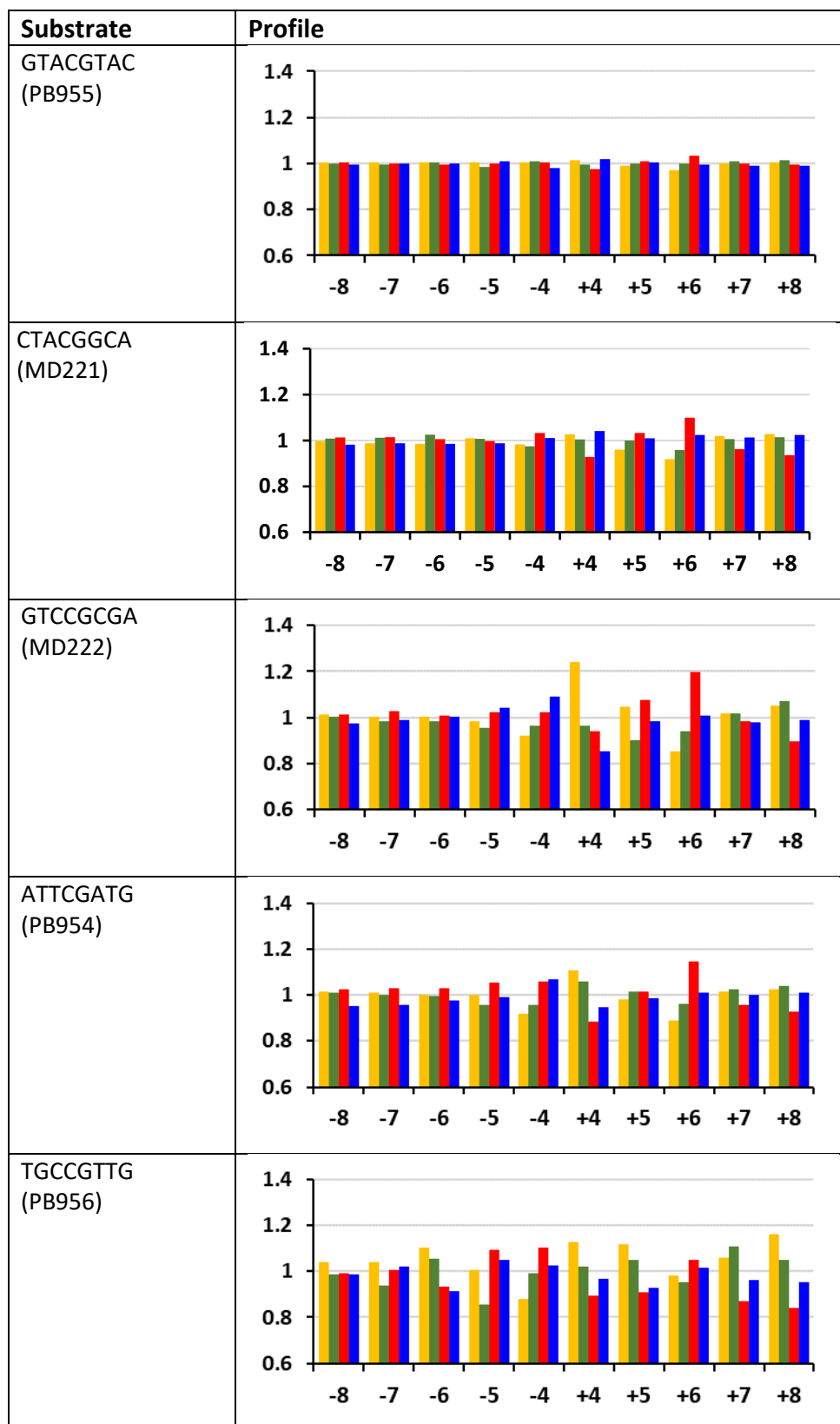
R836A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	

S837A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	

S881A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	
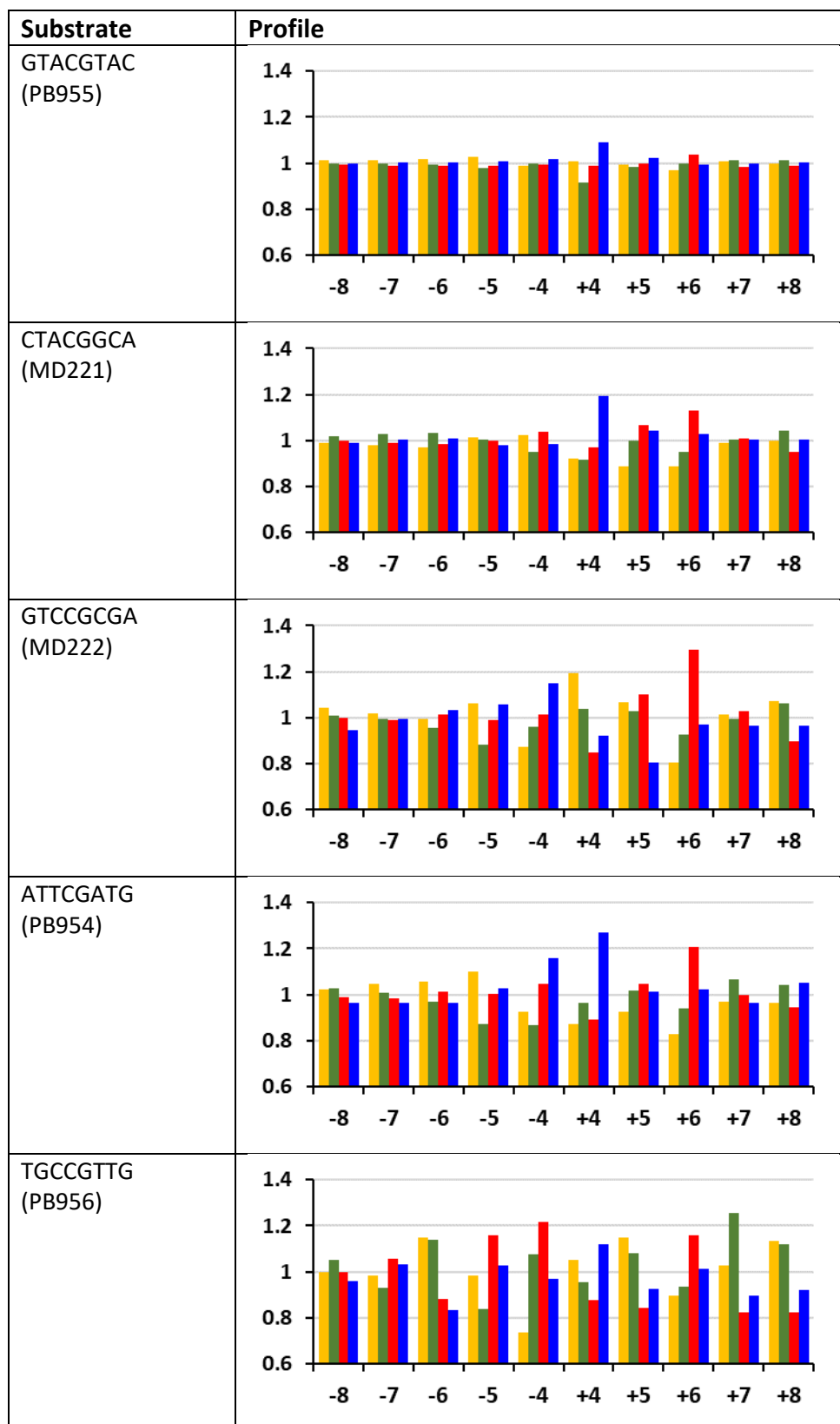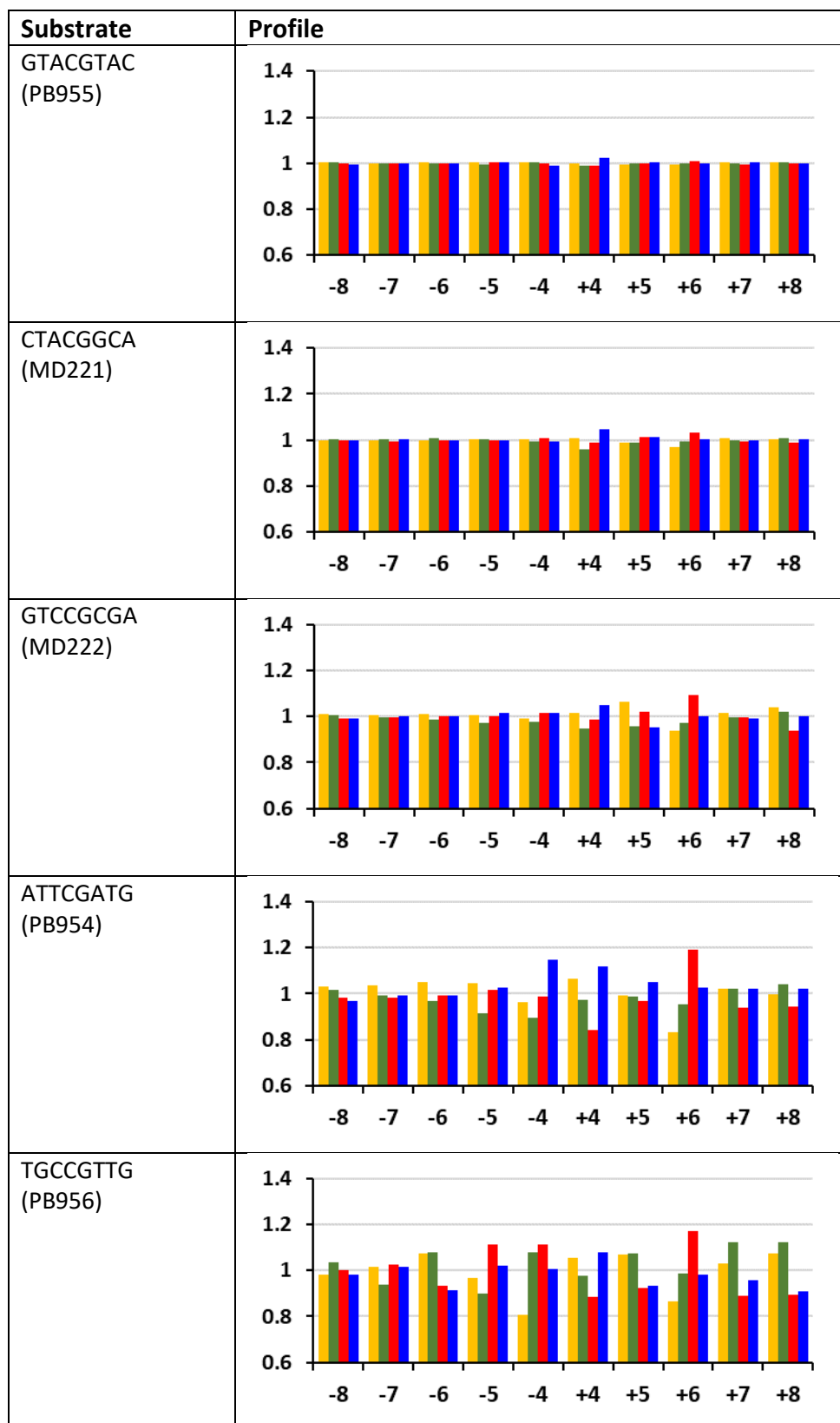
Supplemental Figure 3 (cont.)

R882A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	

R882H

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	
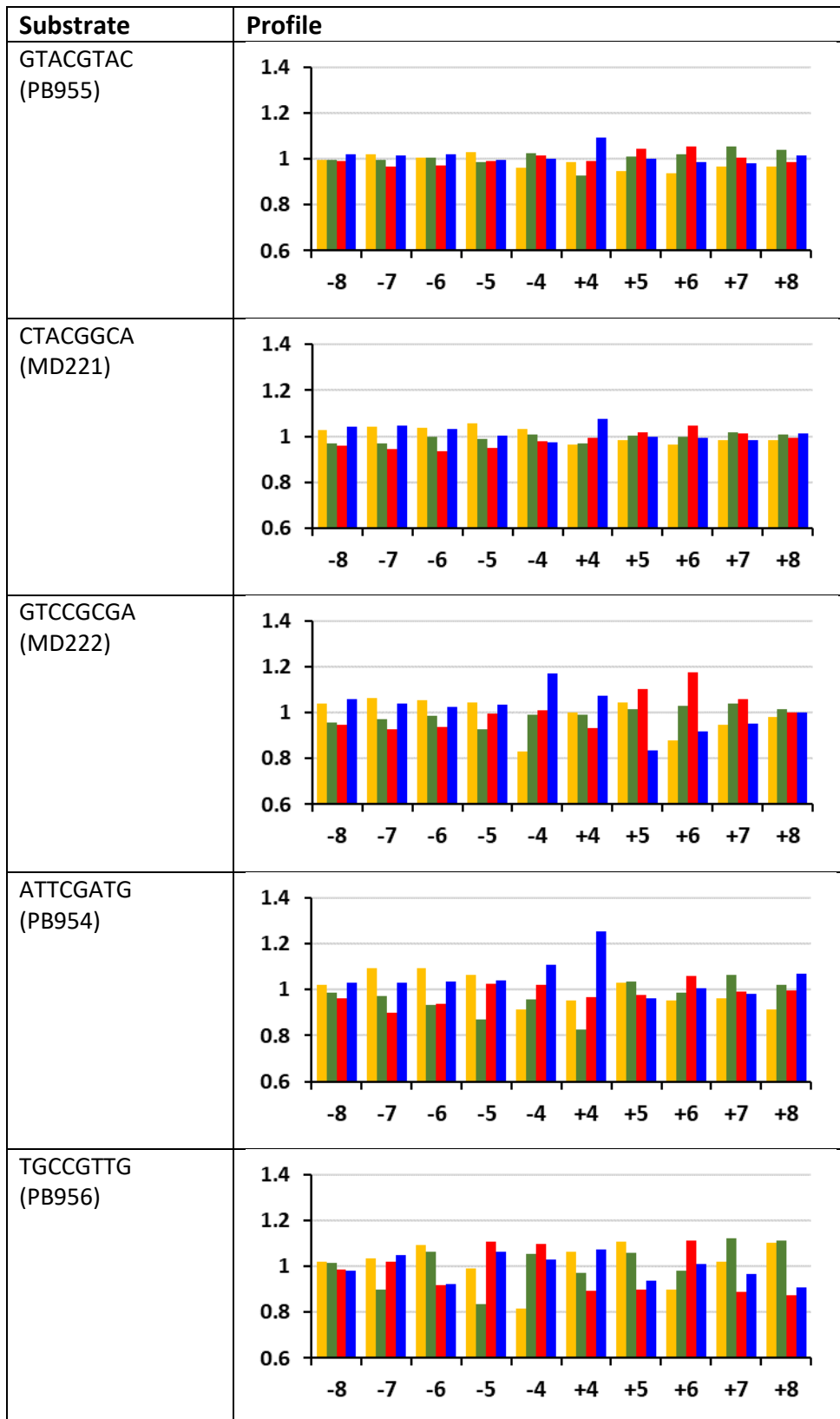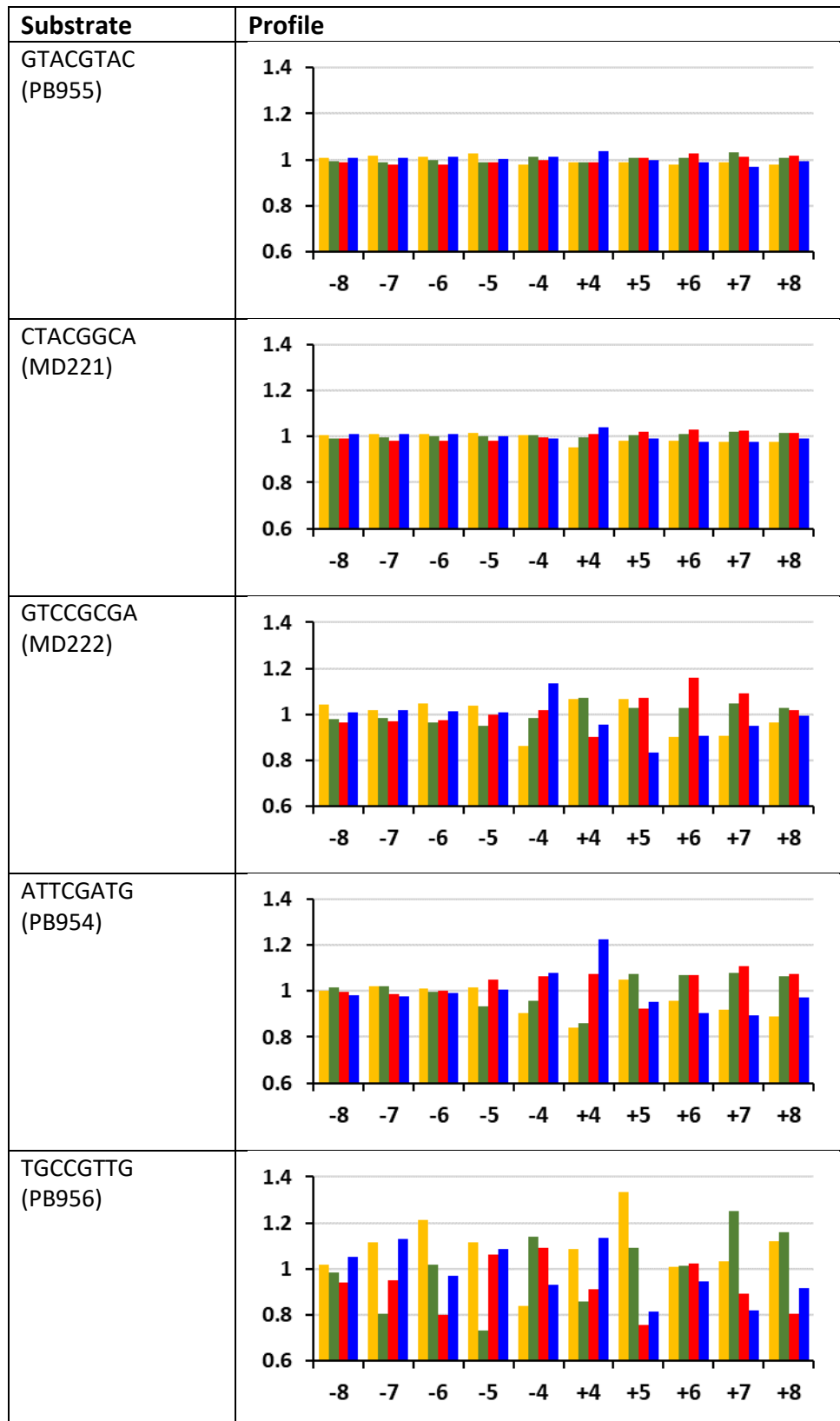
Supplemental Figure 3 (cont.)

L883A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	
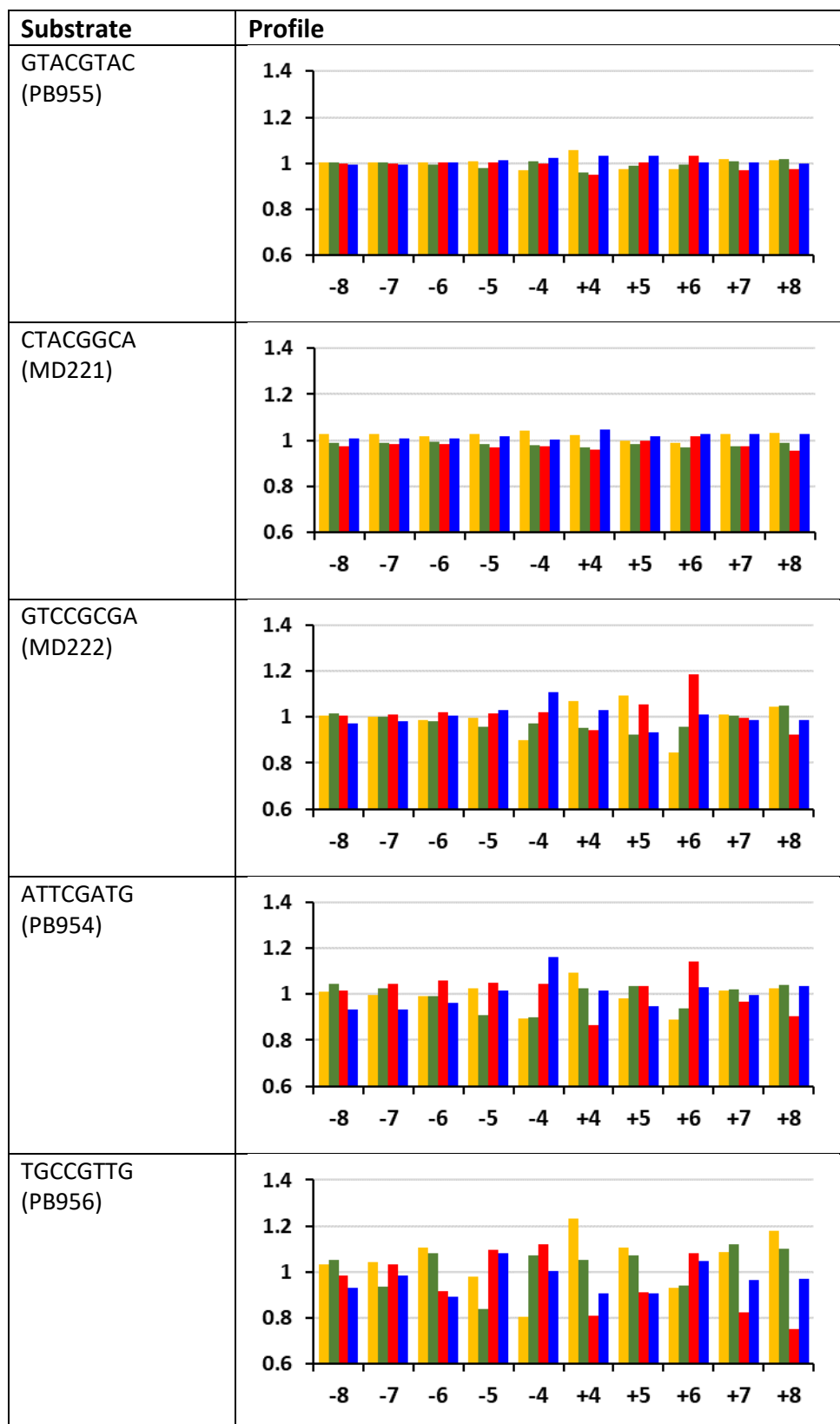
R887A

Substrate	Profile
GTACGTAC (PB955)	
CTACGGCA (MD221)	
GTCCGCGA (MD222)	
ATTCGATG (PB954)	
TGCCGTTG (PB956)	

Supplemental Figure 4. Exemplary data of the fits of CpG and non-CpG methylation of WT and selected mutants.
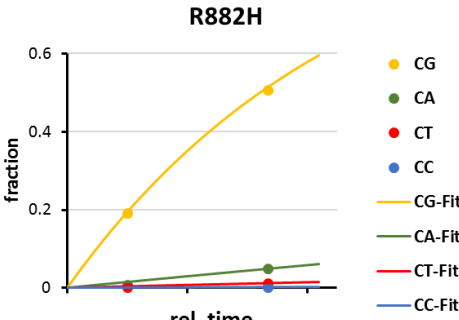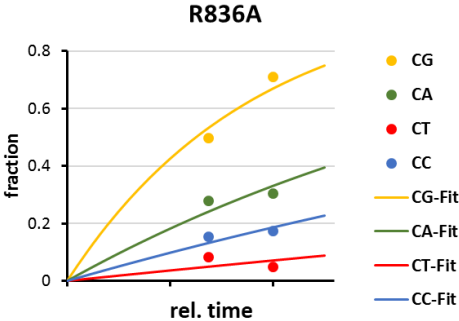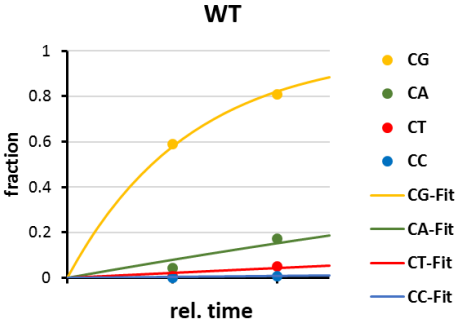
Supplemental Table 1. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the CpG substrate. All reactions were incubated for 60 min.

Enzyme	Repeat	c(E) [µM]	#Met	#Unmet	#Total	Fraction methylated
DNMT3A WT	R1	2	19273	19836	39109	0.493
	R2	4	20569	18826	39395	0.522
R836A	R1	3	19050	19043	38093	0.500
	R2	6	21699	15664	37363	0.581
N838A	R1	1	47153	200732	247885	0.190
	R2	2	54092	181774	235866	0.229
S837A	R1	2	25913	40713	66626	0.389
	R2	4	28367	32248	60615	0.468
S881A	R1	3	39517	18157	57674	0.685
	R2	6	36620	10981	47601	0.769
R882A	R1	10	1414	63856	65270	0.022
	R2[1]	4	16219	50199	66418	0.244
R882H	R1	3	4815	51921	56736	0.085
	R2	6	6833	66271	73104	0.093
L883A	R1	0.5	39999	22472	62471	0.640
	R2	1	33733	20818	54551	0.618
R887A	R1	2	20876	28581	49457	0.422
	R2	4	15433	27659	43092	0.358
no enzyme	R1	0	68	34969	35037	0.002

[1] MBP-tag cleaved off.

Supplemental Table 2. Pearson correlation factors of the -5 to +5 regions of the flanking sequence profiles determined in the independent experimental repeats.

Enzyme	R1 vs. R2
WT	0.985
R836A	0.996
S837A	0.987
N838A	0.998
S881A	0.991
R882A	0.920
R882H	0.995
L883A	0.993
R887A	0.993

Enzyme	Average vs. Gao et al., 2020
WT	0.936

Supplemental Table 3. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the outer flank substrates.

Enzyme	Substrate	Sequence context	Repeat	C(E) [µM]	t (min)	#Met	#Unmet	#Total	Methylated (%)
DNMT3A	PB954	ATT CG ATG	R1	0.5	60	23719	183904	207623	11.42%
			R2	1	60	55777	131628	187405	29.76%
	PB955	GTA CG TCA	R1	0.5	60	20484	119257	139741	14.66%
			R2	1	60	103670	21133	124803	83.07%
	PB956	TGC CG TTG	R1	0.5	60	7487	177460	184947	4.05%
			R2	1	60	29584	124338	153922	19.22%
	MD221	CTA CG GCA	R1	0.5	60	4806	18826	23632	20.34%
			R2	1	60	14572	6073	20645	70.58%
	MD222	GTC CG CGA	R1	0.5	60	9697	31722	41419	23.41%
			R2	1	60	9876	3972	13848	71.32%
R836A	PB954	ATT CG ATG	R1	3	30	15846	62891	78737	20.13%
			R2	3	120	53289	50064	103353	51.56%
	PB955	GTA CG TCA	R1	3	30	61744	35780	97524	63.31%
			R2	3	120	79819	9481	89300	89.38%
	PB956	TGC CG TTG	R1	3	30	16334	55607	71941	22.70%
			R2	3	120	38942	23399	62341	62.47%
	MD221	CTA CG GCA	R1	3	30	34135	55850	89985	37.93%
			R2	3	120	69185	32274	101459	68.19%
	MD222	GTC CG CGA	R1	3	30	26805	41183	67988	39.43%
			R2	3	120	16632	13652	30284	54.92%
S837A	PB954	ATT CG ATG	R1	2	30	1826	56611	58437	3.12%
			R2	2	120	6775	44427	51202	13.23%
	PB955	GTA CG TCA	R1	2	30	42010	33259	75269	55.81%
			R2	2	120	52306	6368	58674	89.15%
	PB956	TGC CG TTG	R1	2	30	1703	67182	68885	2.47%
			R2	2	120	6549	68642	75191	8.71%
	MD221	CTA CG GCA	R1	2	30	12059	45976	58035	20.78%
			R2	2	120	28833	26645	55478	51.97%
	MD222	GTC CG CGA	R1	2	30	7184	33187	40371	17.79%
			R2	2	120	12838	19664	32502	39.50%
S881A	PB954	ATT CG ATG	R1	2	30	15280	42639	57919	26.38%
			R2	2	120	34470	19390	53860	64.00%
	PB955	GTA CG TCA	R1	2	30	53539	8663	62202	86.07%
			R2	2	120	61882	4453	66335	93.29%
	PB956	TGC CG TTG	R1	2	30	10876	57581	68457	15.89%
			R2	2	120	39313	23546	62859	62.54%
	MD221	CTA CG GCA	R1	2	30	46692	14923	61615	75.78%
			R2	2	120	63843	4822	68665	92.98%
	MD222	GTC CG CGA	R1	2	30	39353	13719	53072	74.15%
			R2	2	120	16040	3265	19305	83.09%
R882A	PB954	ATT CG ATG	R1	4	30	2651	85038	87689	3.02%
			R2	4	120	14709	71720	86429	17.02%
	PB955	GTA CG TCA	R1	4	30	28853	90545	119398	24.17%
			R2	4	120	89433	46035	135468	66.02%
	PB956	TGC CG TTG	R1	4	30	3487	70047	73534	4.74%
			R2	4	120	39313	23546	62859	62.54%
	MD221	CTA CG GCA	R1	4	30	30987	65914	96901	31.98%
			R2	4	120	60486	14943	75429	80.19%
	MD222	GTC CG CGA	R1	4	30	21734	67861	89595	24.26%
			R2	4	120	25538	27120	52658	48.50%
R882H	PB954	ATT CG ATG	R1	4	30	7670	54168	61838	12.40%
			R2	4	120	35713	33025	68738	51.96%
	PB955	GTA CG TCA	R1	4	30	29852	39988	69840	42.74%
			R2	4	120	63207	7767	70974	89.06%
	PB956	TGC CG TTG	R1	4	30	1649	73366	75015	2.20%
			R2	4	120	9517	38007	47524	20.03%
	MD221	CTA CG GCA	R1	4	30	46061	25002	71063	64.82%
			R2	4	120	62975	4915	67890	92.76%
	MD222	GTC CG CGA	R1	4	30	21677	29064	50741	42.72%
			R2	4	120	14592	9116	23708	61.55%
L883A	PB954	ATT CG ATG	R1	0.5	30	3208	55521	58729	5.46%
			R2	0.5	120	18026	57343	75369	23.92%
	PB955	GTA CG TCA	R1	0.5	30	11631	65602	77233	15.06%
			R2	0.5	120	38790	33421	72211	53.72%
	PB956	TGC CG TTG	R1	0.5	30	990	77762	78752	1.26%
			R2	0.5	120	4931	61158	66089	7.46%
	MD221	CTA CG GCA	R1	0.5	30	22604	45138	67742	33.37%
			R2	0.5	120	44375	15354	59729	74.29%
	MD222	GTC CG CGA	R1	0.5	30	5381	41014	46395	11.60%
			R2	0.5	120	12429	27336	39765	31.26%
R887A	PB954	ATT CG ATG	R1	3	30	12502	130921	143423	8.72%
			R2	3	120	48652	62434	111086	43.80%
	PB955	GTA CG TCA	R1	3	30	79387	59262	138649	57.26%
			R2	3	120	66732	6638	73370	90.95%
	PB956	TGC CG TTG	R1	3	30	9674	132510	142184	6.80%
			R2	3	120	46734	84973	131707	35.48%
	MD221	CTA CG GCA	R1	3	30	100652	74459	175111	57.48%
			R2	3	120	125159	15066	140225	89.26%
	MD222	GTC CG CGA	R1	3	30	36221	46225	82446	43.93%
			R2	3	120	35156	13442	48598	72.34%
no enzyme	PB954	ATT CG ATG	R1			356	95662	96018	0.37%
	PB955	GTA CG TCA	R1			222	73098	73320	0.30%
	PB956	TGC CG TTG	R1			184	60457	60641	0.30%
	MD221	CTA CG GCA	R1			412	92381	92793	0.44%
	MD222	GTC CG CGA	R1			532	91076	91608	0.58%

Supplemental Table 4. Correlation of outer flank flanking preference profiles in CpG methylation. Pearson correlation factors of the -8 to +8 regions of the flanking sequence profiles determined in the independent experimental repeats.

Enzyme	R1 vs. R2				
	CTA<u>CG</u>GCA	GTC<u>CG</u>CGA	ATT<u>CG</u>ATG	GTA<u>CG</u>TCA	TGC<u>CG</u>TTG
WT	0.951	0.923	0.976	0.821	0.955
R836A	0.879	0.966	0.915	0.913	0.950
S837A	0.909	0.971	0.931	0.943	0.945
S881A	0.781	0.749	0.978	0.799	0.974
R882A	0.966	0.875	0.914	0.945	0.718
R882H	0.919	0.891	0.976	0.853	0.910
L883A	0.947	0.945	0.981	0.950	0.944
R887A	0.666	0.950	0.954	0.932	0.965

Supplemental Table 5. Compilation of conditions, read counts and methylation levels of the Deep Enzymology experiments conducted with the CpN substrate.

Enzyme	Context	Repeat	c(E) [µM]	t [min]	#Met	#Unmet	#Total	Methylated (%)
DNMT3A WT	CG	R1	2	30	29456	20256	49712	59.25%
		R2	2	120	28134	6444	34578	81.36%
	CA	R1	2	30	2325	44250	46575	4.99%
		R2	2	120	6480	29820	36300	17.85%
	CT	R1	2	30	607	45034	45641	1.33%
		R2	2	120	1937	33927	35864	5.40%
	CC	R1	2	30	154	34511	34665	0.44%
		R2	2	120	311	22823	23134	1.34%
R836A	CG	R1	4	30	23458	9386	32844	71.42%
		R2	4	120	17174	17131	34305	50.06%
	CA	R1	4	30	9110	20142	29252	31.14%
		R2	4	120	9601	23802	33403	28.74%
	CT	R1	4	30	1540	27376	28916	5.33%
		R2	4	120	3046	31018	34064	8.94%
	CC	R1	4	30	3359	15425	18784	17.88%
		R2	4	120	3619	19110	22729	15.92%
N838A	CG	R1	2	60	25281	59869	85150	29.69%
		R2	2	60	8362	38274	46636	17.93%
	CA	R1	2	60	908	45433	46341	1.96%
		R2	2	60	266	25800	26066	1.02%
	CT	R1	2	60	233	44293	44526	0.52%
		R2	2	60	95	24879	24974	0.38
	CC	R1	2	60	132	39407	39539	0.38%
		R2	2	60	57	23202	23259	0.25
S837A	CG	R1	4	30	25511	16297	41808	61.02%
		R2	4	120	31464	7313	38777	81.14%
	CA	R1	4	30	3946	36659	40605	9.72%
		R2	4	120	9867	30514	40381	24.43%
	CT	R1	4	30	851	39615	40466	2.10%
		R2	4	120	2731	37948	40679	6.71%
	CC	R1	4	30	200	29193	29393	0.68%
		R2	4	120	436	25373	25809	1.69%
S881A	CG	R1	4	30	48216	8969	57185	84.32%
		R2	4	120	37850	3832	41682	90.81%
	CA	R1	4	30	7642	47722	55364	13.80%
		R2	4	120	11936	31468	43404	27.50%
	CT	R1	4	30	1969	53194	55163	3.57%
		R2	4	120	3963	39001	42964	9.22%
	CC	R1	4	30	437	38685	39122	1.12%
		R2	4	120	735	26955	27690	2.65%
R882A	CG	R1	4	30	1904	18993	20897	9.11%
		R2	4	120	15092	33345	48437	31.16%
	CA	R1	4	30	162	20736	20898	0.78%
		R2	4	120	1146	43802	44948	2.55%
	CT	R1	4	30	86	20786	20872	0.41%
		R2	4	120	346	44070	44416	0.78%
	CC	R1	4	30	57	18021	18078	0.32%
		R2	4	120	178	37001	37179	0.48%
R882H	CG	R1	4	30	8698	36110	44808	19.41%
		R2	4	120	16393	15782	32175	50.95%
	CA	R1	4	30	572	43217	43789	1.31%
		R2	4	120	1766	29637	31403	5.62%
	CT	R1	4	30	191	44210	44401	0.43%
		R2	4	120	498	30516	31014	1.61%
	CC	R1	4	30	151	38214	38365	0.39%
		R2	4	120	168	23925	24093	0.70%
L883A	CG	R1	0.5	30	9333	41212	50545	18.46%
		R2	0.5	120	20432	29560	49992	40.87%
	CA	R1	0.5	30	721	46633	47354	1.52%
		R2	0.5	120	1762	46265	48027	3.67%
	CT	R1	0.5	30	385	46530	46915	0.82%
		R2	0.5	120	920	45935	46855	1.96%
	CC	R1	0.5	30	201	41083	41284	0.49%
		R2	0.5	120	248	37167	37415	0.66%
R887A	CG	R1	2	30	9123	20140	29263	31.18%
		R2	2	120	25893	18214	44107	58.70%
	CA	R1	2	30	491	28151	28642	1.71%
		R2	2	120	2270	37415	39685	5.72%
	CT	R1	2	30	169	28223	28392	0.60%
		R2	2	120	527	37977	38504	1.37%
	CC	R1	2	30	76	23263	23339	0.33%
		R2	2	120	173	30653	30826	0.56%
no enzyme	CG	R1	0	-	36	11879	11915	0.30%
	CA	R1	0	-	59	9871	9930	0.59%
	CT	R1	0	-	39	10666	10705	0.36%
	CC	R1	0	-	43	10000	10043	0.43%

Supplemental Table 6. Correlation of flanking preference profiles of CpN methylation. Pearson correlation factors of the -4 to +4 regions of the flanking sequence profiles determined in the independent experimental repeats.

Enzyme	R1 *vs.* R2				CpG (CN) *vs.* CpG (CG)
	CpG	CpA	CpT	CpC	
WT	0.952	0.985	0.962	0.700	0.951
R836A	0.882	0.985	0.954	0.957	0.948
S837A	0.969	0.987	0.987	0.793	0.956
N838A	0.993	0.819	0.670	low	0.998
S881A	0.971	0.992	0.992	0.899	0.958
R882A	0.973	0.802	low	low	0.966
R882H	0.984	0.941	0.648	low	0.949
L883A	0.986	0.942	0.903	0.450	0.942
R887A	0.990	0.967	0.834	low	0.874