Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase

Michael Dukatz, Sabrina Adam, Mahamaya Biswal, Jikui Song, Pavel Bashtrykov, & Albert Jeltsch

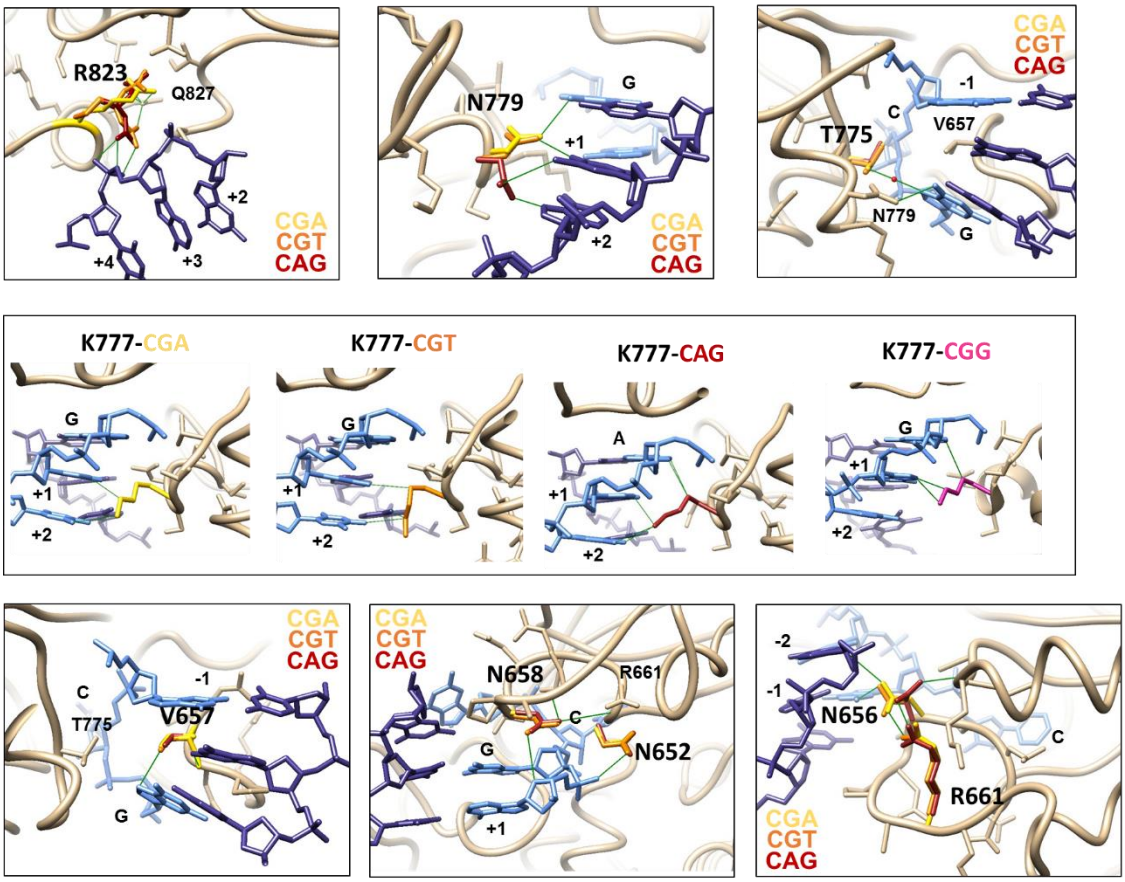
SUPPLEMENTARY INFORMATION

Supplementary Figures

Supplementary Tables

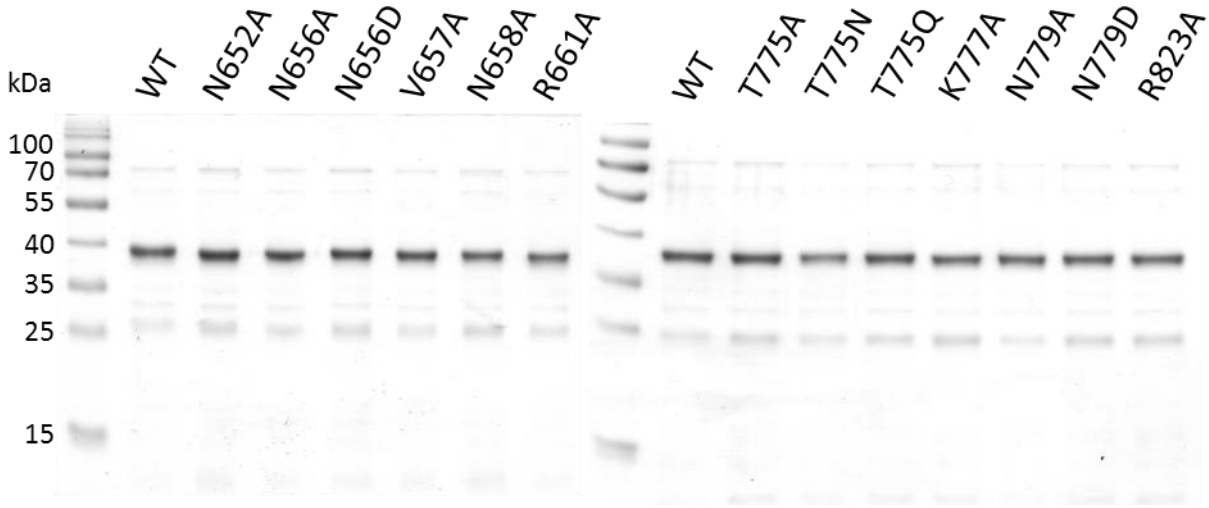
Supplementary References

Supplementary Figure S1.



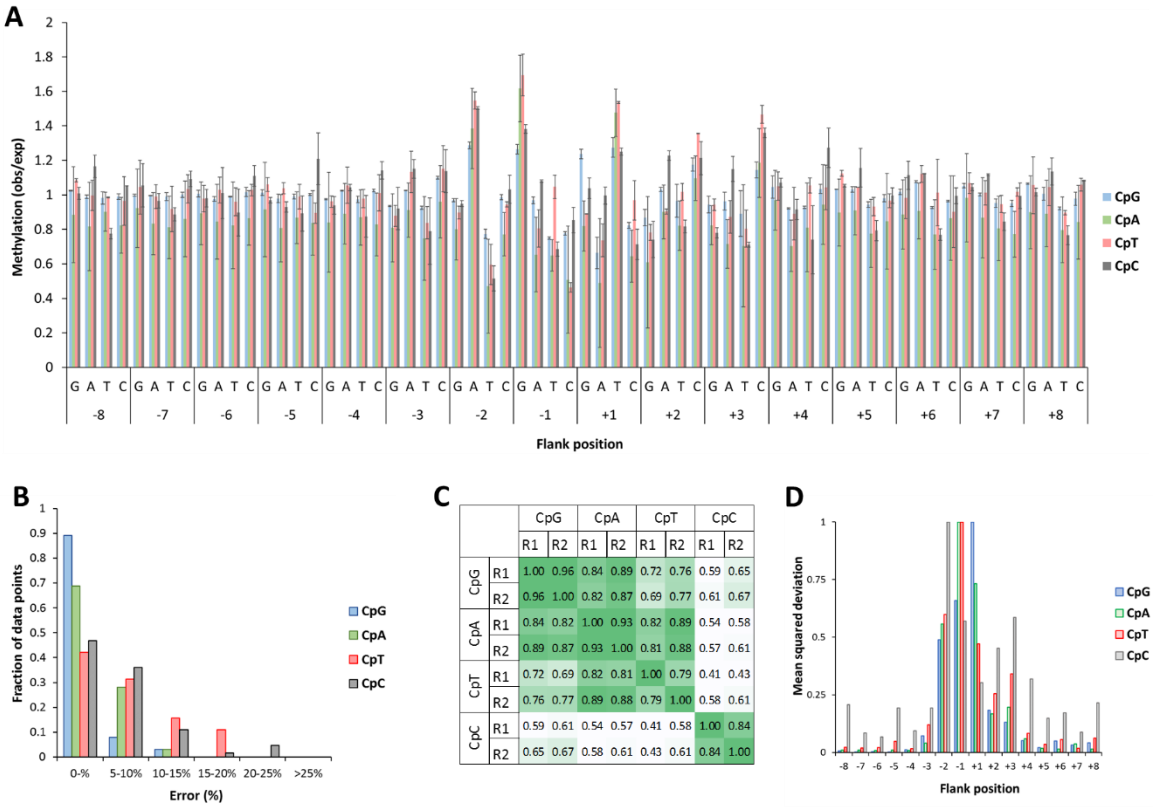
Structural snapshots showing the contacts of the residues investigated in this study in three different DNMT3B structures (CGA, CGT, CAG, see Supplementary Table S1). Green lines refer to contact distances smaller than 4 Å.

Supplementary Figure S2.



Example of a Coomassie BB stained SDS-PAGE gel image of the purified proteins used in this study.

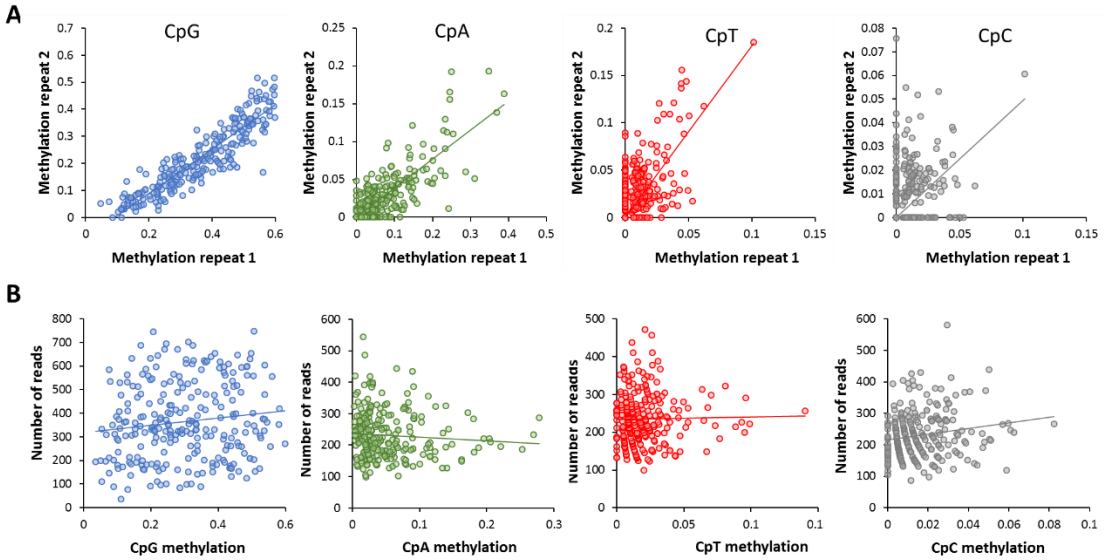
Supplementary Figure S3.



		CpG		CpA		CpT		CpC	
		R1	R2	R1	R2	R1	R2	R1	R2
CpG	R1	1.00	0.96	0.84	0.89	0.72	0.76	0.59	0.65
	R2	0.96	1.00	0.82	0.87	0.69	0.77	0.61	0.67
CpA	R1	0.84	0.82	1.00	0.93	0.82	0.89	0.54	0.58
	R2	0.89	0.87	0.93	1.00	0.81	0.88	0.57	0.61
CpT	R1	0.72	0.69	0.82	0.81	1.00	0.79	0.41	0.43
	R2	0.76	0.77	0.89	0.88	0.79	1.00	0.58	0.61
CpC	R1	0.59	0.61	0.54	0.57	0.41	0.58	1.00	0.84
	R2	0.65	0.67	0.58	0.61	0.43	0.61	0.84	1.00

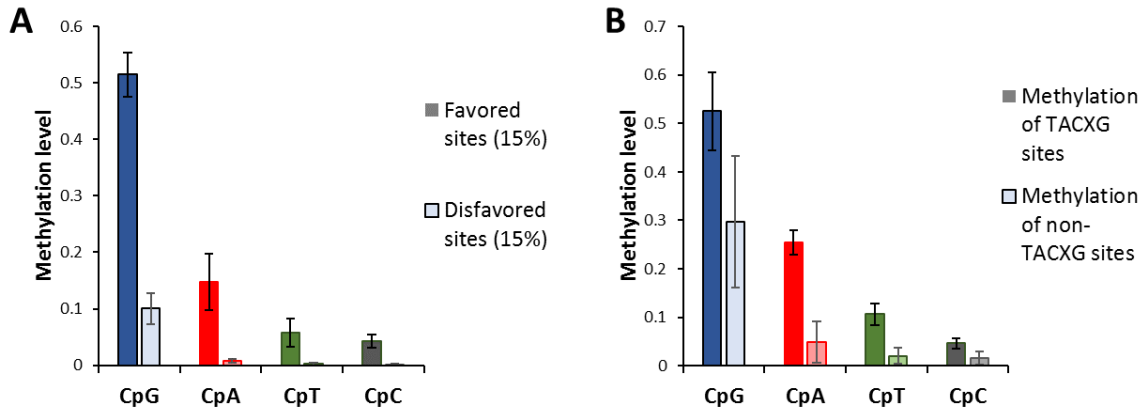
Controls related to the global base enrichment and depletion analysis. (**A**) Compilation of the averaged frequencies of each base at the -8 to +8 flank site. The error bars display the SD of the two measurements. (**B**) Histogram of the distribution of error levels (in %) among the data points. Note that in the CpG and CpA data sets >98% of the points have error margins below 10%. For the CpT and CpC data sets with lower overall methylation relative errors are larger, but even here >90% of the data points have errors below 10% and only one data point had an error above 20% (25%). (**C**) Pairwise correlation of the flank profiles determined in the individual repeats (R1 and R2) of the methylation experiments illustrating the high similarity of the experimental repeats. Shown are Pearson correlation factors. (**D**) Measurement of the position specific variance. For the averaged CpG, CpA, CpT and CpC data sets, the squared values of the deviation of the base frequencies from 0.25 were averaged for all -8 to +8 flank positions and plotted scaled to the largest value. The data show that the -2 to +3 flank positons have the largest influence on the methylation rate.

Supplementary Figure S4.



Controls related to the NNCXNN base enrichment and depletion analysis for WT DNMT3B. (**A**) Correlation of NNCXNN methylation levels observed in the two experimental repeats. (**B**) Lack of correlation of methylation levels with the number of reads observed for the corresponding bin.

Supplementary Figure S5.



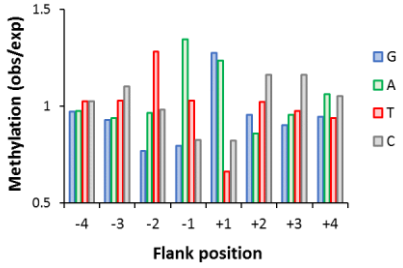
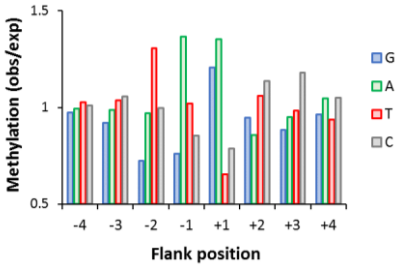
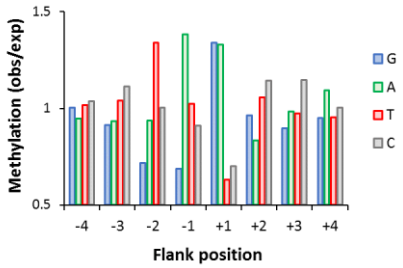
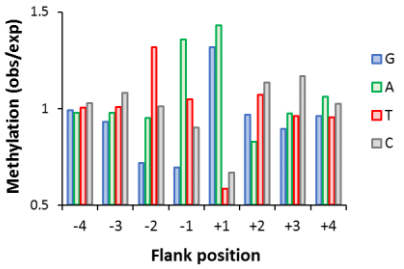
Flanking sequence preferences of WT DNMT3B in CpG and non-CpG methylation. (**A**) Average methylation rate of the 15% of the most preferred and most disfavored flanking sequences. Note the stronger flanking sequence effects in non-CpG methylation. Error bars show the SD. (**B**) Preference for TACXG sequences in non-CpG methylation. Note the elevated preference in CpA and CpT sequence context.

Supplementary Figure S6.

		WT		N652A		N656A		N656D		V657A		N658A		R661A		K777A		N779A		N779D		R823A	
		R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
WT	R1	1.00	0.96	0.95	0.95	0.86	0.86	0.77	0.77	0.96	0.96	0.90	0.92	0.85	0.86	0.00	0.04	0.84	0.85	0.58	0.57	0.95	0.95
	R2	0.96	1.00	0.95	0.94	0.81	0.83	0.76	0.76	0.94	0.95	0.85	0.88	0.87	0.87	0.09	0.11	0.80	0.84	0.50	0.50	0.94	0.93
N652A	R1	0.95	0.95	1.00	0.97	0.83	0.84	0.77	0.78	0.96	0.96	0.90	0.92	0.85	0.86	0.11	0.14	0.83	0.86	0.55	0.55	0.94	0.93
	R2	0.95	0.94	0.97	1.00	0.84	0.84	0.79	0.79	0.95	0.95	0.90	0.92	0.86	0.88	0.12	0.17	0.84	0.87	0.58	0.57	0.94	0.93
N656A	R1	0.86	0.81	0.83	0.84	1.00	0.98	0.65	0.65	0.84	0.84	0.84	0.87	0.80	0.79	-0.04	0.00	0.98	0.96	0.83	0.83	0.80	0.79
	R2	0.86	0.83	0.84	0.84	0.98	1.00	0.67	0.66	0.85	0.85	0.85	0.87	0.82	0.80	-0.04	-0.01	0.97	0.96	0.82	0.81	0.81	0.80
N656D	R1	0.77	0.76	0.77	0.79	0.65	0.67	1.00	0.98	0.73	0.77	0.73	0.74	0.87	0.88	-0.15	-0.11	0.63	0.64	0.38	0.39	0.82	0.82
	R2	0.77	0.76	0.78	0.79	0.65	0.66	0.98	1.00	0.74	0.78	0.74	0.73	0.87	0.88	-0.15	-0.12	0.63	0.64	0.39	0.39	0.82	0.83
V657A	R1	0.96	0.94	0.96	0.95	0.84	0.85	0.73	0.74	1.00	0.98	0.86	0.87	0.82	0.83	0.08	0.13	0.84	0.85	0.56	0.55	0.96	0.94
	R2	0.96	0.95	0.96	0.95	0.84	0.85	0.77	0.78	0.98	1.00	0.87	0.89	0.83	0.85	0.08	0.12	0.84	0.85	0.56	0.55	0.96	0.95
N658A	R1	0.90	0.85	0.90	0.90	0.84	0.85	0.73	0.74	0.86	0.87	1.00	0.96	0.78	0.77	-0.13	-0.09	0.84	0.82	0.66	0.68	0.88	0.88
	R2	0.92	0.88	0.92	0.92	0.87	0.87	0.74	0.73	0.87	0.89	0.96	1.00	0.80	0.80	-0.04	-0.01	0.86	0.86	0.67	0.70	0.89	0.89
R661A	R1	0.85	0.87	0.85	0.86	0.80	0.82	0.87	0.87	0.82	0.83	0.78	0.80	1.00	0.97	-0.02	0.02	0.79	0.81	0.54	0.53	0.84	0.83
	R2	0.86	0.87	0.86	0.88	0.79	0.80	0.88	0.88	0.83	0.85	0.77	0.80	0.97	1.00	0.03	0.07	0.77	0.80	0.53	0.51	0.85	0.85
K777A	R1	0.00	0.09	0.11	0.12	-0.04	-0.04	-0.15	-0.15	0.08	0.08	-0.13	-0.04	-0.02	0.03	1.00	0.97	-0.01	0.06	-0.10	-0.11	0.01	-0.02
	R2	0.04	0.11	0.14	0.17	0.00	-0.01	-0.11	-0.12	0.13	0.12	-0.09	-0.01	0.02	0.07	0.97	1.00	0.02	0.09	-0.08	-0.09	0.05	0.02
N779A	R1	0.84	0.80	0.83	0.84	0.98	0.97	0.63	0.63	0.84	0.84	0.84	0.86	0.79	0.77	-0.01	0.02	1.00	0.97	0.84	0.83	0.79	0.78
	R2	0.85	0.84	0.86	0.87	0.96	0.96	0.64	0.64	0.85	0.85	0.82	0.86	0.81	0.80	0.06	0.09	0.97	1.00	0.78	0.78	0.80	0.79
N779D	R1	0.58	0.50	0.55	0.58	0.83	0.82	0.38	0.39	0.56	0.56	0.66	0.67	0.54	0.53	-0.10	-0.08	0.84	0.78	1.00	0.95	0.52	0.52
	R2	0.57	0.50	0.55	0.57	0.83	0.81	0.39	0.39	0.55	0.55	0.68	0.70	0.53	0.51	-0.11	-0.09	0.83	0.78	0.95	1.00	0.52	0.52
R823A	R1	0.95	0.94	0.94	0.94	0.80	0.81	0.82	0.82	0.96	0.96	0.88	0.89	0.84	0.85	0.01	0.05	0.79	0.80	0.52	0.52	1.00	0.98
	R2	0.95	0.93	0.93	0.93	0.79	0.80	0.82	0.83	0.94	0.95	0.88	0.89	0.83	0.85	-0.02	0.02	0.78	0.79	0.52	0.52	0.98	1.00

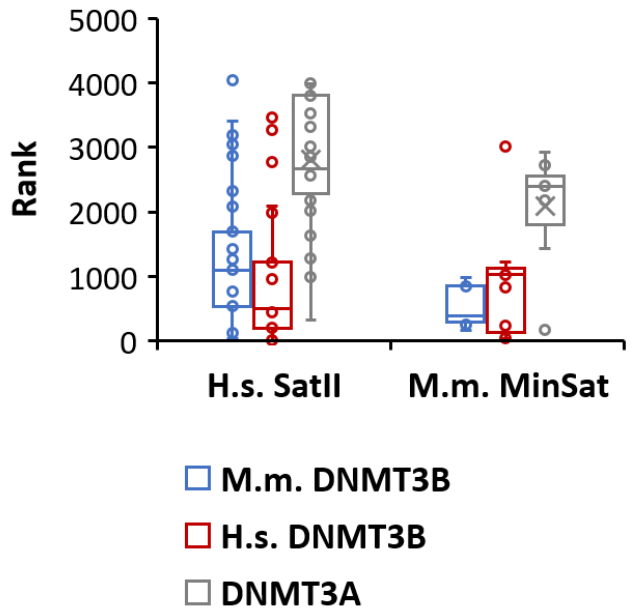
Pairwise correlation of the flank profiles determined in the individual repeats of the CpG methylation experiments of WT and mutant DNMT3B. The data illustrate the high similarity of the experimental repeats. Shown are Pearson correlation factors.

Supplementary Figure S7.

	CpG profiles determined from methylation analysis of CN substrate pool	CpG profiles determined from methylation analysis of hemimethylated CG substrate pool
WT DNMT3B		
R823A		

Comparison of the enrichment and depletion of bases in the -4 to +4 flank region of CpG substrates methylated by WT DNMT3B and R823A. The figure shows the N4 flanking preference profiles determined in the analysis of the methylation of CN substrate pools (taken from Figure 5) and of hemimethylated CpG substrate pools.

Supplementary Figure S8.



Boxplot of the ranks of the CpG sites in the human SatII repeat (n=67) and murine minor satellite repeat (n=11) in the NNNCGNNN flanking preferences of human and mouse DNMT3B and DNMT3A. A low rank indicates high preference. The boxes display the 1st and 3rd quartiles with medians indicated by vertical lines. Whiskers display the data range. Murine DNMT3B and DNMT3A data were taken from (1). Note that DNMT3A disfavors both sequences.

Supplementary Tables

Supplementary Table S1: Compilation of DNMT3B-DNA structures used in this work.

Name of complex	DNA sequence context	PDB number	Reference
CGA	CATGCGATCT	6U8P	Gao et al., 2020
CGT	CATGCGTTCT	6U8V	Gao et al., 2020
CAG	CATGCAGTCT	6U8X	Gao et al., 2020
CGG	AATTCGGAAA	6KDA	Lin et al., 2020
CGT2	CATGCGTTCT	6KDB	Lin et al., 2020

Supplementary Table S2: DNA contacts of the DNMT3B amino acid residues studied in this work. Numbers refer to atom distances in Å.

Residue	Position	CGA complex	CGT complex	CAG complex
N652	+1	Ade-OP1 N652-ND2 3.2	Ade-OP1 N652-ND2 3.3	Ade-OP1 N652-ND2 3.2
N656	-2	Ade'-O4' N656-ND2 4.3	Ade'-N3 N656-ND2 3.9	Ade'-O4' N656-OD1 4.1
	-1	Gua-N2 N656-ND2 3.7 Cyt'-O2 N656-ND2 4.4	Gua-N2 N656-ND2 3.7 Cyt'-O2 N656-ND2 4.3	
V657	-1		Gua-C1' V657-CG 3.8 Gua C2' V657-CG 3.8	
	Target C	Gua'-N2 V657-O 3.3 Gua'-N3 V657-O 3.7	Gua'-N2 V657-O 3.3 Gua'-N3 V657-O 4.3	Gua'-N2 V657-O 3.3 Gua'-N3 V657-O 4.3
	CpN site	Gua-N2 V657-CG 3.7 Gua-N7 V657-CG 3.4 Gua-C5' V657-CG 3.5	Gua-N3 V657-O 3.8	Ade-N7 V657-CG 3.5
N658	CpN site	Gua-O3' N658-ND2 4.3 Gua-O4' N658-ND2 4.4 Gua-N3 N658-OD1 4.4 Gua-N7 N658-OD1 4.4	Gua-O3' N658-ND2 4.1 Gua-O4' N658-ND2 4.2 Gua-N3 N658-OD1 4.6 Gua-N7 N658-OD1 4.5	Ade-O3' N658-ND2 4.4 Ade-O4' N658-ND2 4.1 Ade-N3 N658-ND2 4.6 Ade-N7 N658-OD1 4.5
	+1	Ade-O4' N658-ND2 3.9	Thy-O4' N658-OD2 3.8	Gua-O4' N658-ND2 3.8
T775	Target C	Gua-C2' T775-CG2 4.1 Gua-C3' T775-CG2 4.4	Gua-C2' T775-CG2 4.1 Gua-C3' T775-CG2 4.4	Gua-C2' T775-CG2 4.2 Gua-C3' T775-CG2 4.5
	CpN site	Gua-N7 T775-OG1 4.4	Gua-N7 T775-OG1 4.5	Ade-N7 T775-OG1 4.5
K777	CpN site	Gua-N7 K777-CG 4.1 Gua-C8 K777-CG 4.1	Gua-C8 K777-CG 4.4	Ade-N7 K777-CG 3.7 Ade-C8 K777-CG 3.7
	+1	Ade-N7 K777-CE 3.3 Ade-N6 K777-CG 3.9	Thy-C7 K777-CG 3.5	Gua-N7 K777-NZ 3.1 Gua-O6 K777-NZ 4.1
	+2	Thy-O4 K777-NZ 4.4 Thy-C7 K777-NZ 3.8	Ade'-N6K777-CE 3.6 Thy-O4 K777-CE 3.6	Thy-O4 K777-NZ 3.5 Thy-C7 K777-NZ 3.7
	+3		Gua'-O6 K777-NZ 4.5 Gua'-N7 K777-NZ 4.8	
N779	CpN site	Gua-O6 N779-ND2 3.2 Cyt'-N4 N779-ND2 4.1	Gua-O6 N779-ND2 2.9 Cyt'-N4 N779-ND2 4.3	
	+1	Ade-N6 N779-ND2 3.8 Thy'-O4 N779-ND2 2.8	Thy-O4 N779 ND2 5.0 Ade'-N6 N779ND2 3.6	
	+2	Ade'-N6 N779-ND2 4.3	Ade'-N6 N779-ND2 4.5	Ade'-N7 N779-ND2 3.9 Ade'-OP2 N779-OD1 3.5
R823	+3	Gua'-OP1 R823-NE 4.5	Gua'-OP1 R823-NZ 4.0	Gua'-OP1 R823-NE 4.5
	+4	Ade' OP2 R823-N 4.3	Ade' OP2 R823-N 4.3 R823-NZ 2.6 Ade'-O5' R823-NZ 3.3	Ade' OP2 R823-N 4.5 R823-NZ 3.4 Ade'-O5' R823-NZ 2.7

Supplementary Table S3: Protein contacts of the residues studied in this work. Numbers refer to atom distances in Å.

Residue 1	Residue 2	Atom 1	Atom 2	CGA	CGT	CAG
N652	**N658**	N652-ND2	N658-N	2.7	2.7	2.7
N656	**N658**	N652-ND2	N658-N	2.7	2.7	2.7
	R661	N656- OD1	R661- NH2	2.6	2.6	2.7
	M703	N656- ND2	M703-N			3.5
V657	**T775**	V657-CG2	T775-CG2	4.9	4.9	5.3
N658	**R661**	N658-O	R661-NH2	3.8	4.1	3.5
T775	**K777**	T775-OG1	K777-O	OG1 3.8	3.7	3.9
		T775-OG1	K777-N	OG1 3.6	3.6	3.4
	N779	T775-OG1	N779-OD1	4.1	3.9	

Supplementary Table S4: Summary of the NGS sequencing data used in this study.

Enzyme	Sequence context	Concentration [µM]	Number of reads	Average methylation (%)
DNMT3B WT	CG	0.5	47424	23.7
		1	45734	38.3
	CA	0.5	30213	2.9
		1	28040	7.4
	CT	0.5	30729	1.1
		1	28670	3.2
	CC	0.5	31270	1.1
		1	27576	2.4
	HM CG	0.5	909803	11.6
		1	631992	37.9
N652A	CG	1	12608	36.1
		2	31551	44.8
	CA	1	12162	6.1
		2	30009	10.3
	CT	1	12069	1.9
		2	27703	3.2
	CC	1	12721	1.7
		2	31881	2.4
N656A	CG	0.5	40523	17.9
		1	27707	28.1
	CA	0.5	25768	3.6
		1	17644	7.3
	CT	0.5	26365	1.3
		1	17963	2.7
	CC	0.5	27368	1.0
		1	18290	2.0
N656D	CG	5	42586	11.4
		10	44724	12.2
	CA	5	26889	1.0
		10	27554	1.3
	CT	5	26735	0.6
		10	27541	0.6
	CC	5	29128	0.7
		10	29519	0.9
V657A	CG	1	32673	37.8
		2	30748	43.6
	CA	1	31013	4.5
		2	28644	7.6
	CT	1	28541	5.8
		2	26911	9.2
	CC	1	32625	2.0
		2	30436	3.7
N658A	CG	3	20296	16.0
		6	22440	20.3
	CA	3	20562	4.6
		6	22497	6.1
	CT	3	19487	1.2
		6	20793	1.4
	CC	3	21376	0.9
		6	23573	0.9
R661A	CG	1.5	37235	29.2
		3	44184	39.2
	CA	1.5	23343	5.3
		3	26909	7.7
	CT	1.5	23548	2.5
		3	27043	3.6
	CC	1.5	24091	1.7
		3	27464	2.3
K777A	CG	1	11839	48.3
		2	13262	56.8
	CA	1	9236	16.3
		2	9882	24.2
	CT	1	9308	4.1
		2	9595	8.1

	CC	1	9817	6.3
		2	10730	10.6
N779A	CG	0.5	42293	26.1
		1	40945	34.0
	CA	0.5	26360	5.3
		1	25258	9.8
	CT	0.5	26428	1.6
		1	25223	3.7
	CC	0.5	27770	1.3
		1	26154	2.8
N779D	CG	0.5	18340	11.5
		1	20120	25.7
	CA	0.5	18300	2.3
		1	19489	5.8
	CT	0.5	17274	0.6
		1	18573	1.7
	CC	0.5	19883	1.0
		1	20698	1.9
R823A	CG	1.5	44328	30.7
		3	38794	36.6
	CA	1.5	25999	3.7
		3	23440	5.8
	CT	1.5	25824	1.3
		3	23529	1.9
	CC	1.5	27690	1.3
		3	24452	2.0
	HM CG	3	974408	23.6
no enzyme	CG		13629	0.5
	CA		12608	0.6
	CT		12809	0.3
	CC		13981	0.4

Supplementary References

1. Gao, L., Emperle, M., Guo, Y., Grimm, S.A., Ren, W., Adam, S., Uryu, H., Zhang, Z.M., Chen, D., Yin, J. *et al.* (2020) Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nature communications*, **11**, 3355.