Supplementary Data

Glenna Meister[1], Srinivasan Chandrasegaran[2], Marc Ostermeier[1]

[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218

[2]Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205

EXPANDED MATERIALS AND METHODS

Plasmids

Restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs (Ispwich, MA) and used according to manufacturer's instructions. Agarose gel electrophoresis and PCR were performed essentially as described (19). *Escherichia coli* K-12 strain ER2267 [F′ *proA*$^{+}$*B*$^{+}$ *lacI*^{q} *Δ(lacZ)M15 zzf::mini-Tn*10 (Kan^{R})/ *Δ(argF-lacZ)U169 glnV44* e14^{-}(McrA^{-}) *rfbD1? recA1 relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10*] was obtained from New England Biolabs (Ipswich, MA) and was used as a host for DNA cloning experiments and methylation protection assays. Cells were grown in LB media supplemented with chloramphenicol (50 μg/ml) and/or ampicillin (100 μg/ml). All DNA primers were obtained from Invitrogen (Carlsbad, CA).

Construction of pDIM-N7 M.EcoHK31IA and pAR M.EcoHK31IB vectors

All primers are listed in Table S1 (Supplementary Data). Plasmids were derived from pDIM-N7-MHhaI[1-302] and pAR-MHhaI[29-327] (15). The pAR plasmid had an NcoI site replaced with an NdeI site for cloning purposes.

The pET3a-M38 and pET3a-C23 plasmids containing the *M.EcoHK31IA* and *M.EcoHK31IB* genes respectively were a gift from P.C Shaw (17). The natural *M.EcoHK31IA* gene also encodes the *M.EcoHK31IB* sequence in a different reading frame. The pETa-M38 plasmid contains the *M.EcoHK31IA* sequence with the *M.EcoHK31IB*'s start codon silently removed to prevent expression of both fragments from a single vector. This modified *M. EcoHK31IA* gene was amplified by PCR using HK31IA-for and HK31IA(NcoI)-rev primers. The M. EcoHK31IA fragment was digested with NdeI and NcoI and ligated into a similarly digested pDIM-N7 MHhaI[1-302] to make pDIM-N7 M.EcoHK31IA. Two additional silent mutations were made to the M.EcoHK31IA fragment to remove AUG codons which could act as an internal translation site for the M.EcoHK31IB fragment.

The *M.EcoHK31IB* gene was amplified with the HK31IB-for and HK31IB–rev primers, digested with NdeI and SpeI, and ligated into a similarly digested pAR plasmid. The truncated *M.EcoHK31IB* fragments were amplified using the appropriate forward primer and HK31IB-rev to create truncations ranging from 35 to 50 amino acid residues. Forward primers truncated the N-terminus of the M.EcoHK31IB up to the amino acid residue indicated (ex. HK31IB Δ35-for removes the first 34 amino acid residues) and appended an ATG start codon immediately preceding the residue (see Figure S1 for

deletion fragments). The amplified truncated fragment was then inserted into pAR plasmid using NdeI and SpeI restriction sites. See Figure S2 for detailed plasmid maps.

Construction of pDIM-N7 Tyr123-α and pAR β-Tyr456 plasmids for combinatorial library

All primer sequences are found in Table S2 in supplementary data. We initially made several modifications to the plasmids and gene sequences to facilitate cloning and the detection of methylation using EagI for restriction endonuclease assays. Both the *M.EcoHK31IA* and *M.EcoHK31IB* genes had EcoRI sites silently removed, and the zinc finger sequences for Tyr123 and Tyr456 (20) had NdeI and SpeI sites, which were silently removed. The pDIM-N7 M.EcoHK31IA plasmid also had an EagI restriction site removed (that was located near the intended EagI target site) in order to create a larger digestion fragment that could be detected in restriction endonuclease assays.

The target sites were designed with a methylation site (5'-CGGCCG-3') flanked by the Tyr123 and Tyr456 zinc finger binding sites (Figure 1A). This methylation site is also recognized by the EagI (5'-CGGCCG-3') and EaeI (5'-YGGCCR-3') restriction enzymes. The zinc finger DNA binding sites are separated from the methylation site by 0, 1, 2 or 3 base pairs (labeled 0 bp, 1 bp, 2 bp, and 3 bp target sites) with the methylation site centered between the zinc finger binding sites. Oligonucleotides (TyrZF 0 bp Target-for and rev, TyrZF 1 bp Target-for and rev, TyrZF 2 bp Target-for and rev, and TyrZF 3 bp Target-for and rev) containing the zinc finger binding sites and methylation site were annealed, phosphorylated and ligated into the pDIM-N7 M.EcoHK31IA plasmid between the XmaI and EcoRI sites.

The zinc finger protein Tyr123 was appended to the N-terminus of the M.EcoHK31IA fragment either directly (linker 0L) or using amino acid linkers of various sizes. Linkers were designed to be flexible and are repeating GGGGS sequences to create linkers of the length 5, 10 or 15 residues long (linkers 5L, 10L or 15L). The Tyr123 zinc finger sequence was amplified by PCR with the Tyrzif123-for primer and either the Tyrzif123-0L-rev, Tyrzif123-5L-rev, Tyrzif123-10L-rev, or Tyrzif123-15L-rev primers to attach the amino acid linkers. The length of the 10L and 15L linkers for the Tyr123-10link and Tyr123-15link PCR reactions required the use of the previous linker size fragment (Tyr123-5L or Tyr123-10L) as templates.

The M.EcoHK31IA fragments were amplified and overlapping sequences to the Tyr123 or linkers created using either HK31IA-0L-for, HK31IA-5L-for, HK31I-10L-for or HK31I-15L-for primers and the HK31IA (SalI)-rev primer. The zinc finger Tyr123 and HK31IA fragments with corresponding linkers were fused by overlap extension PCR using Tyr123-for and HK31IA (SalI)-rev primers. Each construct was digested with NdeI and SalI and ligated into the four similarly digested pDIM-N7 target plasmids, containing each of the four previously described target sites. Ligation using the SalI site removed portions of an unused TetR operon and three potential methylation sites to simplify restriction endonuclease assays. This created sixteen pDIM-N7 Tyr123-M.EcoHK31IA (pDIM-N7 Tyr123-α) plasmids containing every possible combination of target site (0 bp, 1 bp, 2 bp, or 3 bp) and Tyr123-M.EcoHK31IA linker length (0L, 5L, 10L or 15L).

The Tyr456 zinc finger protein was fused to the C-terminus of the M.EcoHK31IB Δ42 fragment using the identical linkers as the Tyr123-M.EcoHK31IA fusions (0L, 5L,

10L or 15L). M.EcoHK31IB Δ42 fragments were amplified with HK31IB Δ42-for and either HK31IB-0L-rev, HK31IB-5L-rev, HK31IB-10L-rev or HK31I-15L-rev primers. As before, due to the long length of the 10L and 15L linkers, the HK31IB-10L and HK31IB-15L fragments were amplified using the previous linker size fragments (βΔ42-5L or βΔ42-10L) as a template. The Tyr456 fragments were amplified with the Tyrzif456-rev primer and either Tyrzif456-0L-for, Tyrzif456-5L-for, Tyrzif456-10L-for, or Tyrzif456-15L-for primers. Fusions of the HK31IBΔ42 and Tyr456 fragments with corresponding linker sizes were assembled and amplified by overlap extension PCR using HK31IB Δ42-for and Tyrzif456-rev primers. Fusions were digested with NdeI and SpeI and ligated into similarly digested pAR plasmids creating four unique pAR M.EcoHK31IB Δ42-Tyr456 (pAR βΔ42-Tyr456) plasmids with the four different linker sizes. These plasmids were cotransformed with the sixteen pDIM-N7 Tyr123-α plasmids into ER2267 cells to form a 64 member combinatorial M.EcoHK31I Tyr123-α/βΔ42-Tyr456 library.

Construction of plasmids for M.EcoHK31IB deletion optimization and control plasmids

During the course of our investigation we noted that the methyltransferase appeared to contain a bias against the NotI non-target site. To eliminate this bias we made additional modifications to the pDIM-N7 plasmids. The EagI non-target site was modified so that the overlapping NotI site was removed by changing base pairs flanking the EagI site to adenines. This created identical flanking nucleotides as the target EagI site (Figure 2B). A non-target methylation site overlapped by an EaeI restriction site on

the pAR-NdeI plasmid was also removed due to an overlapping Dcm (CCAGG) methylation site which blocked digestion by EaeI.

We also created several control plasmids to test if the zinc finger binding sites and zinc finger proteins were necessary for targeted methylation. Control target sites lacking the zinc finger binding sites were designed to have an EagI site with similar flanking base pairs as the 3 bp target site but with a random sequence in place of the zinc finger binding sites (Figure 2B). The control target site oligonucleotides Control site (3bp)-for and Control site (3bp)-rev were annealed, phosphorylated and ligated into the pDIM-N7 Tyr123-15L-α plasmid between the XmaI and EcoRI sites in place of the original target site. The M.EcoHK31IA gene without Tyr123 was amplified using the HK31IA-for and HK31IA(SalI)-rev primers and inserted into the pDIM-N7 plasmids containing the 3 bp target site and the plasmid containing a control target site to create the pDIM-N7 α plasmids.

New M.EcoHK31IB-5L-Tyr456 truncations were created by amplifying the M.EcoHK31IBΔ42-5L-Tyr456 fragment with the appropriate HK31IB deletion forward primer (Table S3, supplementary data) and Tyrzif456-rev primers. Truncations were created up to 70 amino acid residues. The truncated fusions were then ligated into the pAR plasmid using the NdeI and SpeI sites. M.EcoHK31IB fragments without zinc finger proteins were also amplified using the deletion forward primer and HK31IB-rev primer and were ligated into the pAR plasmid using the same restriction sites.

A full length M.EcoHK31IB-5L-Tyr456 fusion was also created by amplifying the full length M.EcoHK31IB gene with HK31IB-for and HK31IB-5L-rev primers and fusing to the Tyrzif456-5L fragment by overlap extension PCR as performed before. The

full length fusion was then inserted in the new pAR plasmid using the NdeI and SpeI restriction sites.

Oligonucleotides

Table S1. Oligonucleotide sequences for construction of pDIM-N7 M.EcoHK31IA and pAR M.EcoHK31IB and truncation plasmids

Primer	Sequence 5'-3'
HK31IA -for	CGG GGG CAT ATG AAA AAG AAA CCA CTG AAA CAG
HK31IA(NcoI)-rev	CCT ATA GCC ATG GCT GTT ACG ATA ATG CCT CGT TGA GC
Hk31IB-for	GGA GAT ATA CAT ATG CAA AAC TCA TCG
HK31IB Δ35-for	CCG CAC AAC ATA TGA TAC GTT GGT CGA ATT CG
HK31IB Δ37-for	GCG GCA TAT GTG GTC GAA TTC GGG TAT GGC G
HK31IB Δ38-for	GCG GCA TAT GTC GAA TTC GGG TAT GGC GTT TCG
HK31IB Δ39-for	GCG GCA TAT GAA TTC GGG TAT GGC GTT TCG
HK31IB Δ40-for	CCG CAC AAC ATA TGT CGG GTA TGG CGT TTC CGT GG
HK31IB Δ41-for	GCG GCA TAT GGG TAT GGC GTT TCG TGG AGA G
HK31IB Δ42-for	GCG GCA TAT GAT GGC GTT TCG TGG AGA G
HK31IB Δ45-for	CCG CAC AAC ATA TGC GTG GAG AGT ATT GG
HK31IB Δ50-for	CCG CAC AAC ATA TGA TGC AAA ATA CGG TGG
HK31IB-rev	CTT CCC CAC TAG TCT ACT GCT CTG AAT CG

Table S2: Oligonucleotide sequences for construction of pDIM-N7 Tyr123-α and pAR βΔ42-Tyr456 combinatorial library

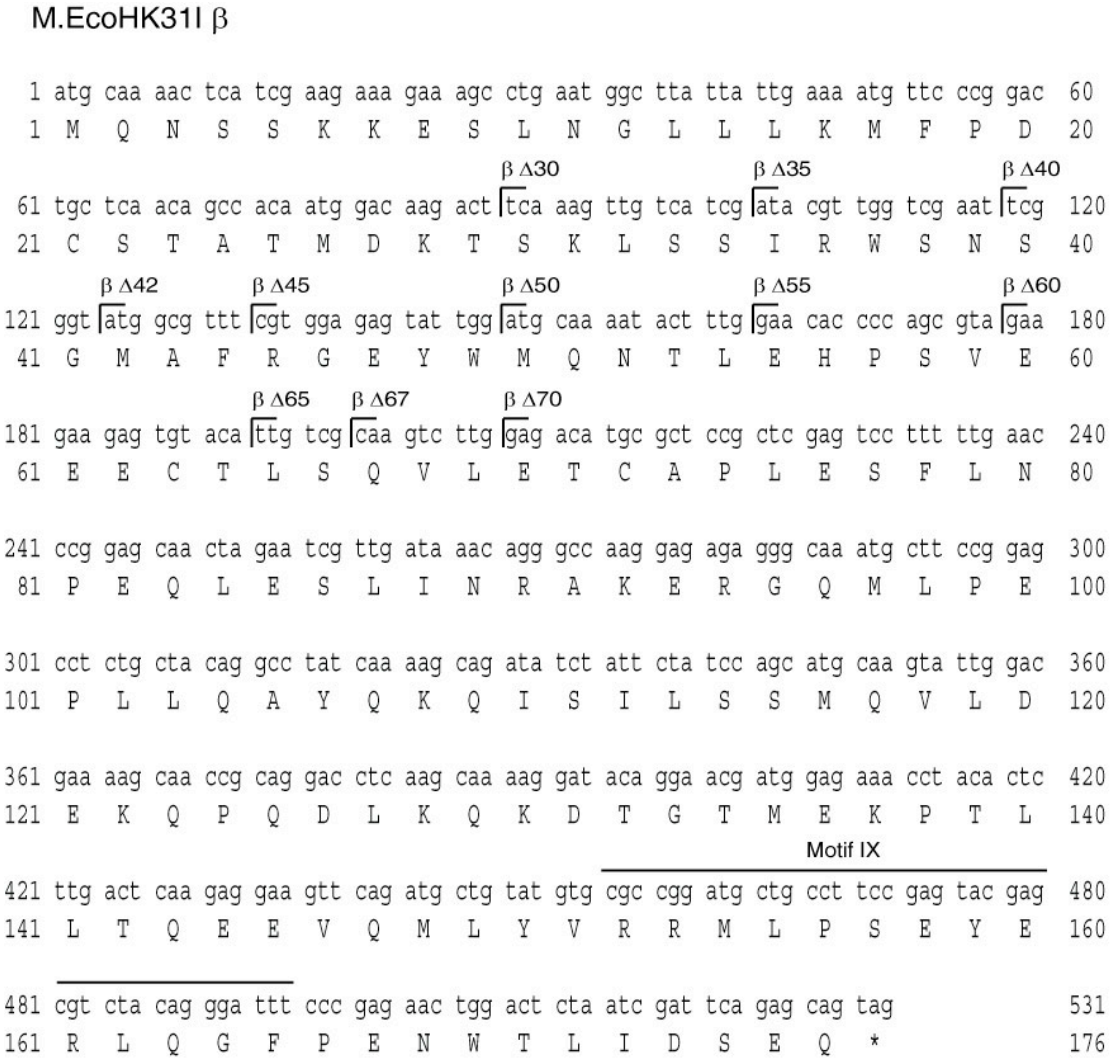
Primer	Sequence 5'-3'
TyrZF 0 bp Target-for	CCG GGA AGC TTG AAG GGG AAC GGC CGG TGG ATG ACG AG
TyrZF 0 bp Target-rev	AAT TCT CGT CAT CCA CCG GCC GTT CCC CTT CAA GCT TC
TyrZF 1 bp Target-for	CCG GGA AGC TTG AAG GGG AAG CGG CCG TGT GGA TGA CGA G
TyrZF 1 bp Target-rev	AAT TCT CGT CAT CCA CAC GGC CGC TTC CCC TTC AAG CTT C
TyrZF 2 bp Target-for	CCG GGA AGC TTG AAG GGG AAG TCG GCC GCT GTG GAT GAC GAG
TyrZF 2 bp Target-rev	AAT TCT CGT CAT CCA CAG CGG CCG ACT TCC CCT TCA AGC TTC
TyrZF 3 bp Target-for	CCG GGA AGC TTG AAG GGG AAG TAC GGC CGA CTG TGG ATG ACG AG
TyrZF 3 bp Target-rev	AAT TCT CGT CAT CCA CAG TCG GCC GTA CTT CCC CTT CAA GCT TC
Tyrzif123-for	GGG CGG CAT ATG GGC AGC AGC CAT CAT CAT C
Tyrzif123-0L-rev	GTG GTT TCT TTT TGA CCA GTC CCT TCT TAT TCT GAT GAG TAC G
Tyrzif123-5L-rev	TGA TCC TCC ACC GCC GAC CAG TCC CTT CTT ATT CTG ATG AGT ACG
Tyrzif123-10L-rev	GCT ACC TCC TCC CCC TGA TCC TCC ACC GCC GAC CAG TCC
Tyrzif123-15L-rev	CTT TTT TGA TCC GCC TCC GCC GCT ACC TCC TCC CCC TG
HK31IA-0L-for	GGG ACT AGA CAA AAA GAA ACC ACT GAA ACA GTA CAA GGT ATC C
HK31IA-5L-for	GGC GGT GGA GGA TCA AAA AAG AAA CCA CTG AAA CAG TAC AAG G
HK31IA-10L-for	CAG GGG GAG GAG GTA GCA AAA AGA AAC CAC TGA AAC AGT ACA AG
HK31IA-15L-for	GGC GGA GGC GGA TCA AAA AAG AAA CCA CTG AAA CAG TAC AAG G
HK31IA (SalI)-rev	CCG CCC GTC GAC TTA CGA TAA TGC CTC GTT GAG CAC C
HK31IB Δ42-for	GCG GCA TAT GAT GGC GTT TCG TGG AGA G
HK31IB-0L-rev	GCT GCT GCC CTG CTC TGA ATC GAT TAG AGT CCA GTT C
HK31IB-5L-rev	TGA TCC TCC ACC GCC CTG CTC TGA ATC GAT TAG AGT CCA G
HK31IB-10L-rev	CCT CCT CCC CCT GAT CCT CCA CCG CCC TGC
HK31IB-15L-rev	CTG ATC CGC CTC CGC CGC TAC CTC CTC CCC CTG ATC C
Tyrzif456-0L-for	GAT TCA GAG CAG GGC AGC AGC CAT CAT CAT CAT C
Tyrzif456-5L-for	CAG GGC GGT GGA GGA TCA GGC AGC AGC CAT CAT CAT CAT C
Tyrzif456-10L-for	GGC GGT GGA GGA TCA GGG GGA GGA GGT AGC GGC AGC AGC CAT CAT C
Tyrzif456-15L-for	GCG GCG GAG GCG GAT CAG GCA GCA GCC ATC ATC ATC
Tyrzif456-rev	CCG CAC TAG TTT AGA CCA GTC CCT TCT TAT TCT GAT GAG TAC G

Table S3: Oligonucleotide sequences for construction of optimized M.EcoHK31I-zinc finger fusions

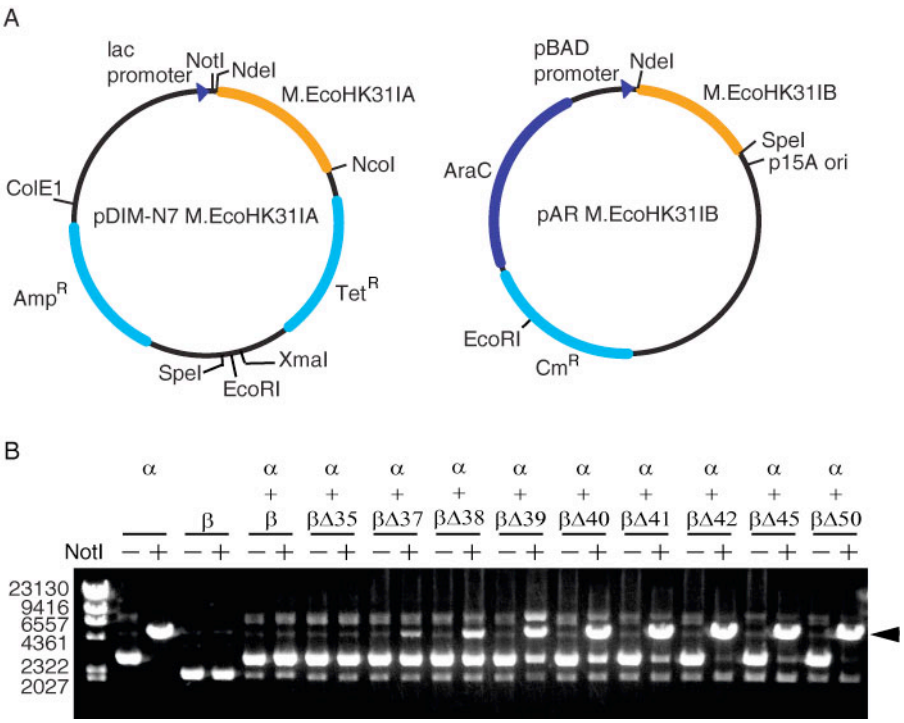
Primer	**Sequence 5'-3'**
Control site (3 bp)-for	CCG GGA AGC TTT CGT ACG GCC GAC TAC TTG ATG AG
Control site (3 bp)-rev	AAT TCT CAT CAA GTA GTC GGC CGT ACG AAA GCT TC
HK31IB-for	GGA GAT ATA CAT ATG CAA AAC TCA TCG
HK31IB Δ42-for	GCG GCA TAT GAT GGC GTT TCG TGG AGA G
HK31IB Δ50-for	CCG CAC AAC ATA TGA TGC AAA ATA CGG TGG
HK31IB Δ55-for	GGG CGT CAT ATG GAA CAC CCC AGC GTA GAA GAA GAG TGT AC
HK31IB Δ60-for	GGG CGT CAT ATG GAA GAA GAG TGT ACA TTG TCG CAA GTC TTG G
HK31IB Δ65-for	GGG GTA CAT ATG TTG TCG CAA GTC TTG GAG ACA TGC
HK31IB Δ66-for	GGA CGC ATA TGT CGC AAG TCT TGG AGA CAT GCG CTC C
HK31IB Δ67-for	GGA CGC ATA TGC AAG TCT TGG AGA CAT GCG CTC CGC
HK31IB Δ68-for	GGA CGC ATA TGG TCT TGG AGA CAT GCG CTC CGC TCG
HK31IB Δ69-for	GGA CGC ATA TGT TGG AGA CAT GCG CTC CGC TCG
HK31IB Δ70-for	CGG GGA CAT ATG GAG ACA TGC GCT CCG CTC G
HK31IB-rev	CTT CCC CAC TAG TCT ACT GCT CTG AAT CG

Table S4: Bisulfite sequencing primers for optimized M.EcoHK31I-ZF fusion plasmids.

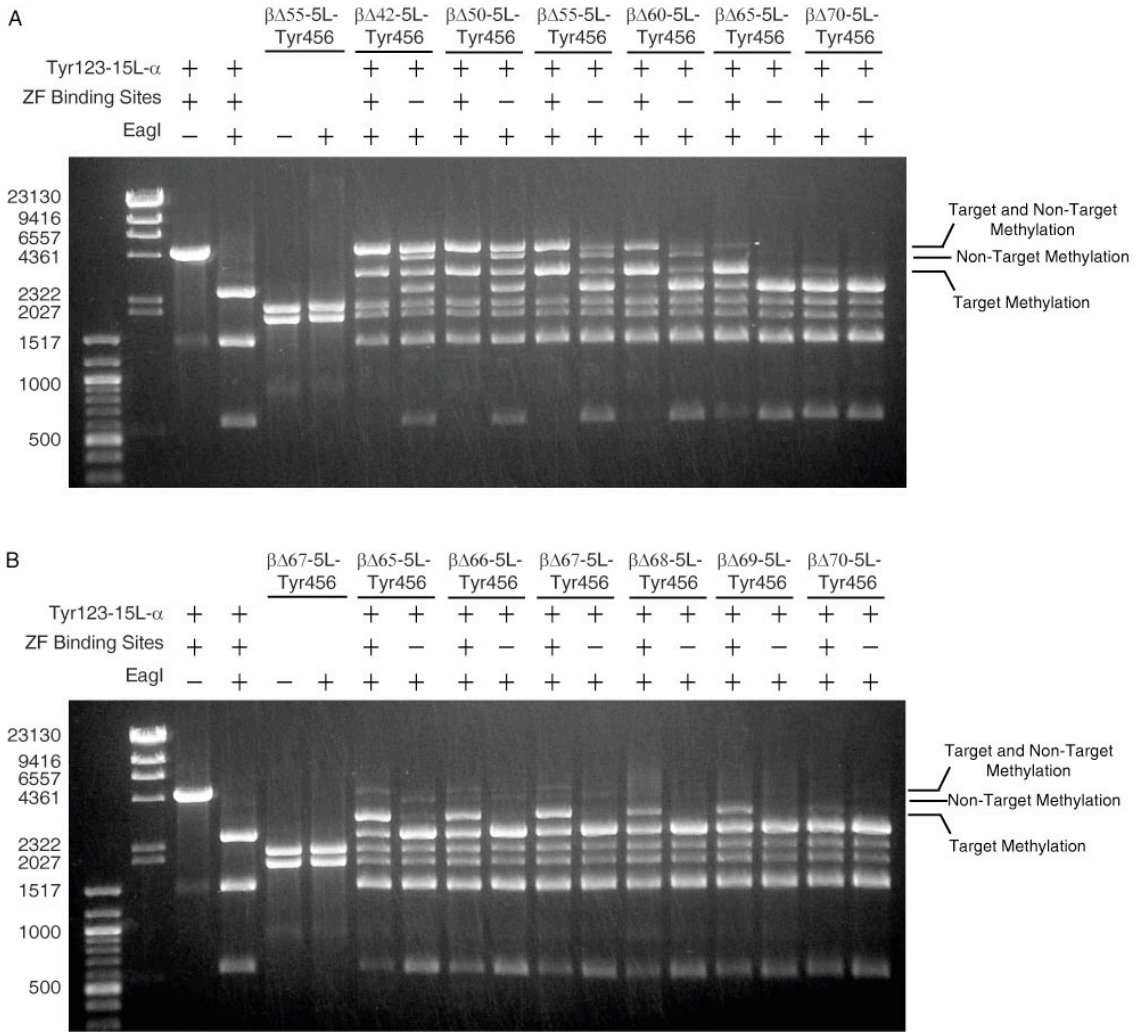
Primer	**Sequence 5'-3'**
Bisulfite Top Target-for	GGC TTC ATA TGG TGT TGT GTT GTG GTG TAT GGA GTT GG
Bisulfite Top Target-rev	CCA CCA CTA GTC GTC ATA ATA TTC ACT CCA AAA CAA TAA AAA CAT TTC AAT TTA CTC
Bisulfite Bottom Target-for	CGT GCA TAT GTG GTA TTT ATT TTA GAG TGA TGA AAA TGT TTT AGT TTG TTT ATG GAA AAT GG
Bisulfite Bottom Target-rev	GAC GAC TAG TCC ACC TCA ACC TAA ATA AAA ACC AAC AAC ACC
Bisulfite Non-Target-for	GGC GGC ATA TGG TTT GAA ATT AAT TTT TAT TAA AGG GAA TAA AAG TTG GAG TTT TAT TGT GG
Bisulfite Non-Target-rev	CCG GCC ACT AGT CCC ACA TTC CAA ACA CTT ATA AAA TTT TTC CAT ATA ACT ACC



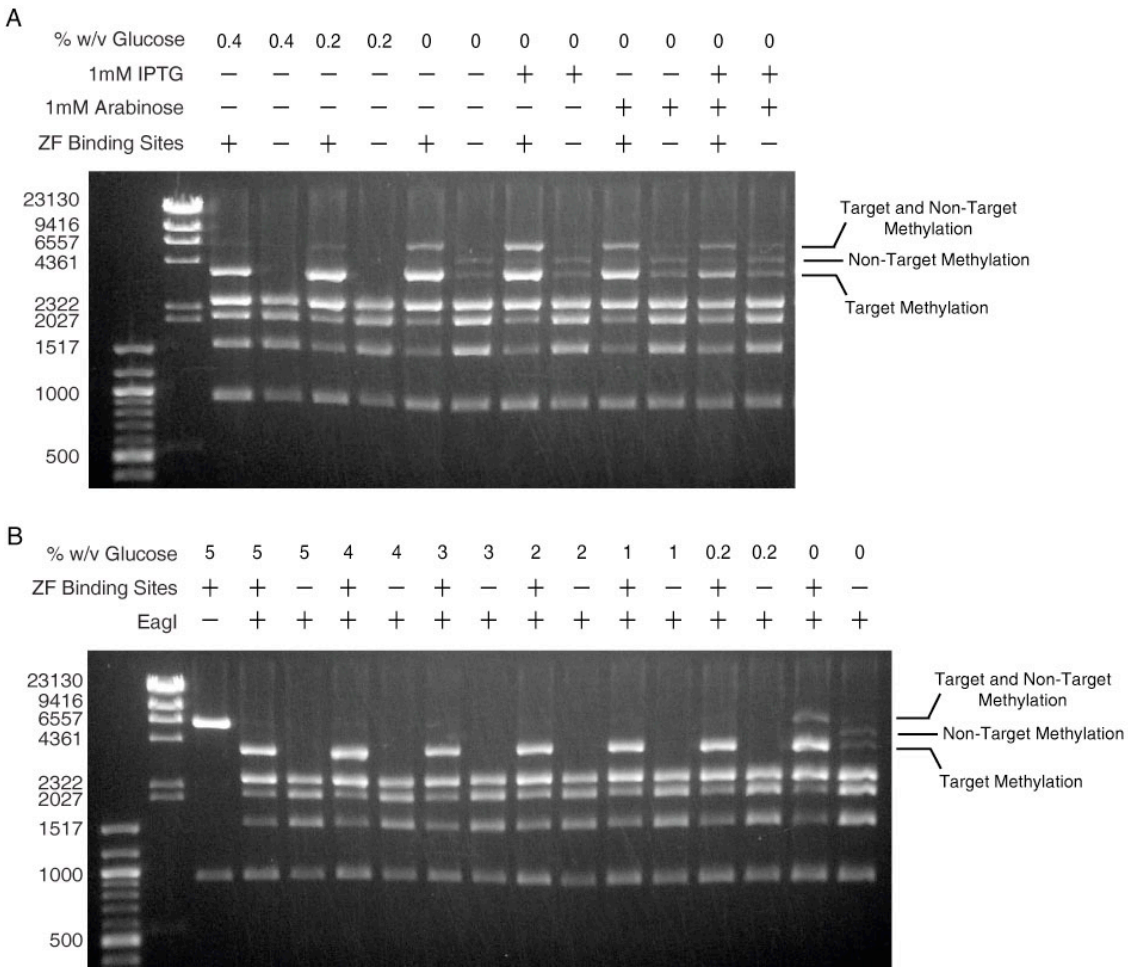
Supplemental Figure S1. DNA and amino acid sequences of M.EcoHK31IB. The M.EcoHK31IB contains only one conserved 5mC methyltransferase motif (motif IX). M.EcoHK31IB was truncated from the N-terminus to create various deletion fragments (β Δxx). The starting locations for various truncations are marked.



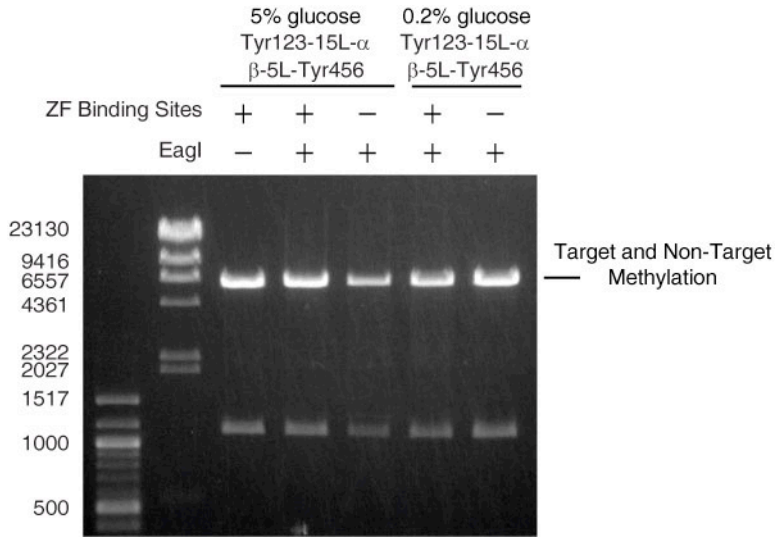
Supplemental Figure S2. The effect of truncation of the N-terminus of M.EcoHK31IB on methylation activity in *E. coli*. A) Schematic depiction of the pDIM-N7 and pAR plasmids containing the M.EcoHK31IA (α) and M.EcoHK31IB (β) fragments (and truncations thereof), respectively. The NotI site used for detection of M.EcoHK31I methylation is marked on the pDIM-N7 plasmid. B) Plasmid DNA from cells expressing the α fragment and the indicated truncations from the N-terminus of the β fragment were digested with NotI and analyzed by agarose gel electrophoresis to test for methylation activity. The NotI enzyme is blocked by M.EcoHK31I methylation. The arrow indicates the location of the expected length of DNA produced when the NotI site is not methylated. The plasmids were prepared from cells grown in LB media.



Supplemental Figure S3. Optimization of the deletion length for M.EcoHK31IB for targeted methylation. A plasmids encoding deletion fragments of M.EcoHK31IB fused to Tyr456 via a 5 amino acid were cotransformed with the pDIM-N7 plasmid encoding Tyr123-15L-M.EcoHK31IA (Tyr123-15L-α). Two versions of the latter plasmid were used, one containing Tyr123 and Tyr456 zinc finger (ZF) binding sites flanking the target site (+) and one in which the flanking sequences lacked affinity for Tyr123 and Tyr 456 (–). The methylation state of target and non-target sites was assessed as in Figure 1. Agarose gel electrophoresis results for A) a set of truncations in the range of 42 to 70 amino acid residues were tested along with B) a set of truncations in one amino acid intervals from 65 to 70 amino acid residues. The locations of bands resulting from only target methylation, only non-target methylation, and methylation at both sites are indicated. The plasmids were prepared from cells grown in LB media.



Supplemental Figure S4. Gel electrophoresis analysis of the effect of growth conditions on methylation specificity. pDIM-N7 plasmids containing Tyr123-15L-α and βΔ67-5L-Tyr456 were isolated from overnight cultures of ER2267 cells and incubated with BsaAI and EagI (+) restriction enzymes. Cells were grown in LB media supplemented with various amounts of glucose (from 5% to 0.2% w/v) to repress expression (A and B), or combinations of 1mM IPTG and/or 1mM arabinose to induce the lac or pBAD promoters (A). Plasmids contained target sites containing the zinc finger binding site (+) or a control target site lacking the zinc finger recognition sites (-).



Supplemental Figure S5. Methylation activity of the pDIM-N7 plasmids encoding a full length β-5L-Tyr456 and Tyr123-15L-α. Plasmids were incubated with BsaAI/EagI (+) or just BsaAI (-). Plasmids contain target sites containing the zinc finger binding site (+) or a control target site lacking the zinc finger recognition sites (-). The band location for plasmids fully methylated at target and non-target sites are indicated.