

# Novel nonspecific DNA adenine methyltransferases

**Marek Drozdz, Andrzej Piekarowicz, Janusz M. Bujnicki & Monika Radlinska**

## SUPPLEMENTARY DATA

**Table 1S.** Oligonucleotide primers used for cloning and mutagenesis. Introduced restriction enzyme sites are indicated in bold.

Name of the primer	Primer sequence (5' – 3')	Restriction site created or mutagenic substitution
HindNco	<b>CCATGGCAAACAAAAAAACTTTAACAGC</b>	NcoI,
HindXho	<b>CTCGAGGAATTTATAAATCATATTATCCTCATAAATTCC</b>	XhoI
NmeNde	<b>CATATGATGCAAAAATACCACTCAACGGCC-</b>	NdeI
NmeXho	<b>CTCGAG ATCTATTAAATCATATTGCCTGATAC</b>	XhoI
8Hf	GTTGGCGGTTAGTCGAATCAACTTC	
8Hr	ACCGTTTATTGTTACTTCCG	
8HfNde	<b>CATATGGCAAATCAAACACCTTAAACAAGC</b>	NdeI
8HXho	<b>CTCGAGGAATTATAGACTAAATTGTCTCATACGCCACTTG</b>	XhoI
Hia5APPYf	cgCCACCTTACCTTGACCAAGCAG*	D194A
Hia5APPY	ccAATACAAACAACGCTTCGGATCATTGC	D194A
NmeAPPf	gcgCCGCCTTATGTGTCCACTGCTCAGGG	D191A
NmeAPPr	cAATACCAGCAGAGTATTGGGGTTATGCTGATG	D191A
HinAPPf	gCCACCTTATTATGCACTCGCCAAGAGAGTTAC	D194A
HinAPPr	cgCAAGTAATAGTAGCACCTATCTTGATTTC	D194A
Pvu1AT	GACTAGTAGATCTAAGAAGGAGATATAACATGTCTGCAGCGAT	PvuI**
Pst1TGCA	CGCTGCAGACATGTTATCTCCTCTTAAGATCTACTAGTCTGCA	PstII**

\* Lowercase letters represent nucleotides not complementary to the template

\*\*The oligonucleotide duplex PstTGCA/Pvu1AT has sticky ends to create PstI and PvuI sites.

**Table 2S.** Oligonucleotide duplex DNAs used as substrates for *in vitro* methylation assays.

Duplex*	Length (bp)	Number of adenine residues	Sequence (5' – 3')
CA10	21	10	CACACACACACACACAC GTGTGTGTGTGTGTGTGTG
GA10	21	10	GAGAGAGAGAGAGAGAGAG CTCTCTCTCTCTCTCTCTC
TA50	50	50	AT TA
AA21	21	21	AAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTT
Cm2A1	21	3(1)**	CCCCCCCCCaAaCCCCCCCC*** GGGGGGGGGGTTGGGGGGGG
Cm2A3	23	5(3)**	CCCCCCCCCaAAAaCCCCCCCC*** GGGGGGGGGGTTGGGGGGGG

\* To anneal duplexes, the complementary oligonucleotides were mixed, heated to 90°C for 5 min and slowly cooled to 25°C.

\*\*() number of presumed substrate adenines in the duplex

\*\*\*a - the marked adenine (lower case) is m<sup>6</sup>A

**Table 3S.** Summary of the sensitivity of different restriction endonucleases to modifications introduced by the Nme1821 enzyme.

REase	Recognition site	Number of sites in pNmeET	Cleavage <sup>a</sup>
AdeI	CACNNNGTG	1	N
BglII	AGaTCT	1	N
Bsh1285I	CGRYCG R=A/G	5	N
Bsp119I	TTCGAA	0(2) <sup>b</sup>	N
Bsu15I	ATCGAT	2	N
CaiI	CAGNNNCTG	1	N
CseI	GACGC	12	N
CviQI	GTAC	6	N
DraI	TTTAAA	2	N
Eam1104I	GAAGAG	3	N
Eco147I	AGGCCT	0(1) <sup>b</sup>	N
Eco47III	AGCGCT	3	N
Eco91I	GGTNACC	1	N
Eco32I	GATATC	0(1) <sup>b</sup>	P, L
EcoRI	GAATTTC	1	N
HindIII	AaGCTT	0(1) <sup>b</sup>	N
HinfI	GANTC	20	N
HphI	GGTGA/TCACC	17	N
LweI	GCATC	27	N
MboI*	GATC	23	N
MluI	ACGCGT	1	N
MlyI	GAGTC/GACTC	9	N
Mph1103I	ATGCAT	2	N
MunI	CAATTG	2	N
NdeI	CaTATG	1	N
NlaIII	CATG	26	N
PaeI	GCATGC	1	N
PfeI	GAWTC	11	N
PscI	ACATGT	1	N
PstI	CTGCAG	1	N
PvuII	CAGCTG	3	N
SchI	GAGTC/GACTC	9	N
XbaI	TCTGAG	1	N
XhoI	CTCGAG	1	N
XmiI	GTMKAC	3	N
BpiI	GAAGAC	5	P
AluI	AGCT	24	P,
HincII	GTYRAC	2	P, L
KspAI	GAATTTC	1	P, L
SspI	aATATT	3	P
TaiI	ACGT	18	P, L
TasI	AATT	27	P
TaqI	TCGA	18	P, L
Tru1I	TTAA	28	P
VspI	ATTAAT	5	P,L
XapI	RAATTY	6	P

REase	Recognition site	Number of sites in pNmeET	Cleavage <sup>a</sup>
BcnI	CCSGG	13	Y
BglII	GGCN5GCC	2	Y
Bme1390I	CCNGG	23	Y
Bpu1102I	GCTNAGC	2	Y
Bsh1236I	CGCG	37	Y
Bsp143I*	GATC	23	Y
Bsp120I	GGGCC	1	Y
Bsp68I	TCGCGA	1	Y
BspLI	GGNNCC	22	Y
Bst1107I	GTATAAC	1	Y
BsuRI	GGCC	28	Y
Csp6I	GTAC	6	Y
DpnI*	Gm <sup>6</sup> ATC	23	Y
Eco130I	CCWWGG	2	Y
Eco88I	CYCGRG	2	Y
EcoO109I	RGGNCCY	3	Y
FspBI	CTAG	8	Y
Hin1I	GRCGYC	5	Y
Hin6I	GCGC	53	Y
HpyF10VI	GCN7GC	46	Y
MbiI	GAGCGG/CCGCTC	4	Y
MspI	CCGG	32	Y
PvuI	CGATCG	1	Y
SmaI	CCCGGG	1	Y

<sup>a</sup>**Abbreviations:** Y, complete cleavage; N, no digestion; P, partial digestion; L, plasmid linearized.  
(b – number of recognition sites in plasmid pAltOKI)

\*tested in  $\lambda$  DNA dam- *in vitro* methylated with the Nme1821 enzyme.

**Table 4S.** Summary of the sensitivity of different restriction endonucleases to modifications introduced by the Hin1523 enzyme.

REase	Recognition site	Number of sites in pHinET	Cleavage <sup>a</sup>
AdeI	CACNNNGTG	1	N
Bsh1285I	CGRYCG R=A/G	4	N
Bsu15I	ATCGAT	2	N
CaiI	CAGNNNCTG	1	N
Eco147I	AGGCCT	0 (1) <sup>b</sup>	N
Eco32I	GATATC	0 (1) <sup>b</sup>	N
Mph1103I	ATGCAT	2	N
MunI	CAATTG	1	N
PaeI	GCATGC	1	N
PscI	ACATGT	1	N
PstI	CTGCAG	0 (2) <sup>b</sup>	N
XhoI	CTCGAG	1	N
XmiI	GTMKAC	1	N
TaiI	ACGT	16	P, L
AluI	AGCT	23	P
BpiI	GAAGAC	4	P
BglII	AGaTCT	1	P
CseI	GACGC	12	P
CviQI	GTAC	4	P,L
DpnI*	Gm <sup>6</sup> ATC	23	Y
Eam1104I	GAAGAG	3	P
Eco47III	AGCGCT	3	P
Eco91I	GGTNACC	1	P
EcoRI	GAATTTC	1	P
HindIII	AaGCTT	0 (1) <sup>b</sup>	P
HinfI	GANTC	19	P
HphI	GGTGA/TCACC	18	P
LweI	GCATC	23	P
MboI*	GATC	23	N
MluI	ACGCGT	1	P
MlyI	GAGTC/GACTC	10	P
NdeI	CaTATG	1	P
NlaIII	CATG	27	P
PfeI	GAWTC	9	N
PvuII	CAGCTG	3	P
SchI	GAGTC/GACTC	10	P
TaqI	TCGA	15	P, L
TasI	AATT	30	P
Tru1I	TTAA	32	P
VspI	ATTAAT	5	P
XapI	RAATTY	9	P
XbaI	TCTGAG	1	P
Bsp119I	TTCGAA	0 (1) <sup>b</sup>	P
HincII	GYRAC	1 (GTTAAC)	P
KspAI	GAATTTC	1	P
SspI	aATATT	3	P

REase	Recognition site	Number of sites in pHinET	Cleavage <sup>a</sup>
BcnI	CCSGG	12	Y
BglII	GGCN5GCC	1	Y
Bme1390I	CCNGG	21	Y
Bpu1102I	GCTNAGC	2	Y
Bsh1236I	CGCG	37	Y
Bsp143I*	GATC	23	Y
Bsp120I	GGGCC	1	Y
Bsp68I	TCGCGA	1	Y
BsplI	GGNNCC	20	Y
Bst1107I	GTATAC	1	Y
BsuRI	GGCC	22	Y
Csp6I	GTAC	3	Y
DraI	TTTAAA	2	Y
Eco130I	CCWWGG	1	Y
Eco88I	CYCGRG	2	Y
EcoO109I	RGGNCCY	3	Y
FspBI	CTAG	6	Y
Hin1I	GRCGYC	5	Y
Hin6I	GCGC	32	Y
HpyF10VI	GCN7GC	41	Y
MbiI	GAGCGG/CCGCTC	4	Y
MspI	CCGG	29	Y
PvuI	CGATCG	1	Y

<sup>a</sup>**Abbreviations:** Y, complete cleavage; N, no digestion, P, partial digestion, L, plasmid linearized.  
(b –number of recognition sites in plasmid pAltOKI

\*tested in λ DNA dam- *in vitro* methylated with the Hin1523 enzyme.

Hia5Com AY647244	1 GTTGGCGGTTAGTCGAATCACTGCCATAATGTTCAATAAA-AGAAGAATTAAAGGAGAGTTACCC TGCAAAAGTTAATT CAGTAAGAAAAGACGGCGATAACAACGGCACTAGGAATGCTC 1 GTTGGCGGTTAGTCGAATCAACTAGCGGTATATTAAAGAGAAAAGTTAAAGGAGAGTTACCC TGCAAAAGTTAATT CAGTAAGAAAAGACGGCGATAACAACGGCACTAGGAATGCTC
Hia5Com AY647244	130 GTTGT TAC CAG CTAC GCG AGAGTACCTGCATATAGCCATACGCCCTACCTTGCGCAAGGCGGGGATTGT ACAAATCTTTGATTAGGAGAAATATG CAGTCAATTAAAGCAATCGTTGCAC 131 GTTGT TAC CAG CTAC GCG AGAATAAGCCTGCATATAGCCATACGCCCTACCTTGCGCAAGGCGGGGATTGT ACAAATCTTTGATTAGGAGAAATATG CAGTCAATTAAAGTAATCGTTGCAC
Hia5Com AY647244	260 ATTTGTAACAAATTATTGGCAAAGTAGGGATGGTTAGTAACTACAGCTTAATTGATTGAGTGT CAGAATGCCCTGAGCATCGAACGC 261 ATTTGTAACAAATTATTGGCAAAGTAGGGATGGTTAGTAACTACAGCTTAATTGATTGAGTGT CAGAATGCCCTGAGCATCGAACGC
Hia5Com AY647244	390 CATAGAATAGAAAGGAAAAACTATGGCAAATCAAACACCTTTAACAAAGCCCCCTGCCATTATCGAACAAAAGCAATGTTCTCAAACAAATTGAGCAGATTAAATGAGAATATTCGATAAC 391 C----ATAGAAAGGAAAAACTATGGCAAATCAAACACCTTTAACAAAGCTCCATTGCCATTATCGAACAAAAGCAATGTTCTAAACATTGAAACAGTTAAATGAGAATATAAAAGCGAT
Hia5Com AY647244	520 GGC GAAGGCTGGACAATTCTTGACACTTTGGCGTT CAGGCTTACTCAGTCACACGCCAACGGTAAACCGGAAAGCCCGTCAATTACATGATTGATGGCTATGCCGAGCATTGGCACACA 516 GGC GAAGGCTGGACGATTATTGATACATT CGCGGTT CGGGCTATTAGCCACGCAGCCAAGTAATTAAACCAAAGCACCGTACGTTACATGATTGATAGCTATGCCGAAAGATTGGCATATA
Hia5Com AY647244	650 TTGATGATATTAAACCAGTTGGAGCCGAGCTTACTCTGTA GTGGTAACTGCTACGTCAAAAATAACGTATGACGAAGGATTGTAAGCAGAATGCACTAGAATTATTCAAAGGATATAA 646 TCAACGACACTAACGCCCTACGACACAAATCTTGCAAAATTGGTAACGCTACGCCAACAGATAAGCGTTACCGTAAAGCAGAAATCATTAAACATCGACCAATTCAAAGGTACAA
Hia5Com AY647244	780 AGATCTAAATTGCTTAGCGAGTTGGTATTGTCAGTGGCAACAAAGTGGCAACCGCTTGATGACTTATTCCAACATAATTCTGGCATTGTATTGTCAGTCTGATTATCCTAAAGGCTGATGGCTATTG 776 AGATTAAACTGCTGACGAGTTGGCTATTGTTAGTGGACAACAAGTAAGCTCATTAGACGAACTGTATAAGAAAGATTGGCATTGCGTTCGATTAAGCAGATTATCCTAGTGCAGAGGGTATTG
Hia5Com AY647244	910 GACGGCGTAGAGATTGTGAAAGAATCATTCCACACGCTTTGCTAAGTTAGCAATGATCGAACCGTTGTTGTATTAGATCCACCTTACCTTGCAACAGCAGGAAAGCTACAAACAAGCTACCT 906 GATGGCGTAGGTTATCCGTGAATCATTCCACACACTTTGCGGAAGTTAGCGATAATCGAACCGCTTACCTTGCAACAGCAGGAAAGCTACAAACAAGCCACCT
Hia5Com AY647244	1040 ATTTTGATTTGATTGATTTCTTGCAGTGGTCAATATCAGCGACGCCGTATGTTCTTAGCTGACGAAGTGGAGTTATTGCTTTGTGAATTATGCTGAAAGATAAGGTGGATAATTGGCA 1036 ATTTTGATTTGATAGATTCTTGCATTGGTCAATATTACGCTACCCACCTTATATATTCTCAGTTCAACAAAGTCGGAAATTGCGTTATTGAGTATGGTCAATGATAAAAGTGCATAATTGGCA
Hia5Com AY647244	1170 GGC GTTTGAAAACGCCAACCGGATTACAGTCAATGCCAAACTGAACTATCAAGTGGCTATGAAGACAATTAGTCTATAAATTCTAGCAGTAACAAAGGCTTCGAGTAATCAGAACGCCTT-GTTTA 1166 GGC GTTTGAAAACGCCAACCGGATTACAGTCAATGCCAAACTGAACTACCAAGTGGCTATGAAGACAATTAGTCTATAAATTCTAGCAGTAACAAAGGCTTCGAGTAATCAGAACGCCTTGTGTTA
Hia5Com AY647244	1299 GTCTTCTAATT CGGAAAGTAAACAATAAAACGGT 1296 GTCTTCTAATT CGGAAAGTAAACAATAAAACGGT

**Figure 1S.** DNA sequence comparison of a 1332-bp fragment of *H. influenzae* biotype aegyptius strain ATCC 11116 genomic DNA and the homologous segment of the *H. influenzae* biotype aegyptius strain 3031 genomic island (Acc. no. AY647244). Identical nucleotides are shaded green.

```

1 GTTGGCGGTT TAGTCCGAAT CAACTTGCCA TAATGTTCAA TAAAAGAAGA ATTTAAAGGA GAGTTTACCC TGACAAAGT TTAATTCACT AAAAGAAAAGA CGGCGATAAC AACGGCACTA GGAATGCTCG
131 TTGTTACAG CTACGAGAG AGTACCTGCA TATAGCCATA CGCCGCCTAC CTTGCAGCGAG GCGGGCGGAT TGTAACAAAT CTTTGATTA GGAGAAATAT ATGCAGTCAA TTAAAGCAAT CCGTTGCACA
>>.....ComHia.....>
M Q S I K A I R C T

261 TTTTGTAACA AATTATTGGC GAAAGTAGGG ATGGTTGGTT ATTTAGAAAT CAAATGCCCT CGTTGAAAAA CCGTTAACAC TACACGTTAA TTTGATTGA GTGTCAGAAT GCCTTGAGCA TCGGAACGCC
>.....ComHia.....>>
F C N K L L A K V G M V G Y L E I K C P R C K T V N T T R -
M A N Q N T F K Q A P L P F I G Q K R M F L K Q F E Q I L N E N I S D N

391 ATAGAATAGA AAGGAAAAAC TATGGCAAAT CAAAACACCT TTAAACAAGC CCCCTGCCA TTTATCGGAC AAAAACGAAT GTTCTCAAA CAATTGAGC AGATTTAAA TGAGAATATT TCCGATAACG
>>.....Hia5.....>
M A N Q N T F K Q A P L P F I G Q K R M F L K Q F E Q I L N E N I S D N

521 GCGAAGGCTG GACAATTCTT GACACTTTG GCGGTTCAAGG CTTACTCACT CACACCGCCA AACGGTTAAA ACCGAAAGCC CGCGTCATTT ACAATGATT TGATGGCTAT GCGGAGCGAT TGGCACACAT
>.....Hia5.....>
G E G W T I L D T F G G S G L L S H T A K R L K P K A R V I Y N D F D G Y A E R L A H

651 TGATGATATT AACCAAGTTGC GAGCCGAGCT TTACTCTGTA GTTGGTAATG CTACGTCAA AAATAAACGT ATGACGAAGG ATTGTAAGC AGAATGCATC AGAATTATTCA AAAACTTCAA AGGATATAAA
>.....Hia5.....>
I D D I N Q L R A E L Y S V V G N A T S K N K R M T K D C K A E C I R I I Q N F K G Y K

781 GATCTAAATT GCTTAGCGAG TTGGTTATTG TTCAGTGGC AACAAAGTGGC AACGCTTGAT GACTTATTCC AACATAATT CTGGCATTGT ATTCTGTCAGT CTGATTATCC AAAGGCTGAT GGCTATTTGG
>.....Hia5.....>
D L N C L A S W L L F S G Q Q V A T L D D L F Q H N F W H C I R Q S D Y P K A D G Y L

911 ACGGCGTAGA GATTGTGAAA GAATCATTCC ACACGCTTT GCCTAAGTTT AGCAATGATC CGAAAGCGT GTTGTATTA GATCCACCTT ACCTTGCAC CAAGCAGGAA AGCTACAAAC AAGCTACCTA
>.....Hia5.....>
D G V E I V K E S F H T L L P K F S N D P K A L F V L D P P Y L C T K Q E S Y K Q A T

1041 TTTTGATTTG ATTGATTCT TGCGACTGGT CAATATCACG CGACCGCCGT ATGTTTCTT TAGCTCGACG AAGTCGGAGT TTATTCGCTT TGTGAATTAT ATGCTGGAAG ATAAGGTGGA TAATTGGCAG
>.....Hia5.....>
Y F D L I D F L R L V N I T R P P Y V F F S S T K S E F I R F V N Y M L E D K V D N W Q

1171 GCGTTGAAA ACGCCAAACG GATTACAGTC AATGCCAAAC TGAACTATCA AGTGGCGTAT GAAGACAATT TAGTCTATAA ATTCTAGCAG TAACAAAGGC TTGAGTAAT CACGAAGCCT TTGTTTTAGT
>.....Hia5.....>>
A F E N A K R I T V N A K L N Y Q V A Y E D N L V Y K F -

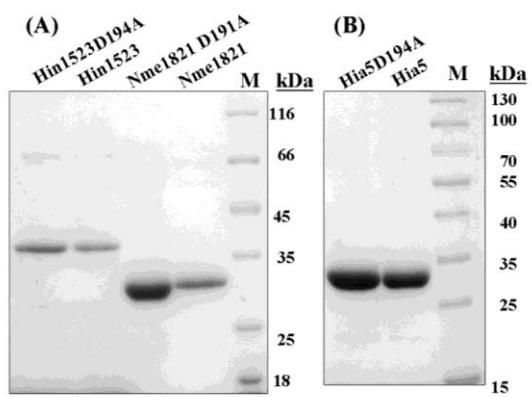
```

1301 CTTCTAATTC GGAAAGTAAA CAATAAAACG GT

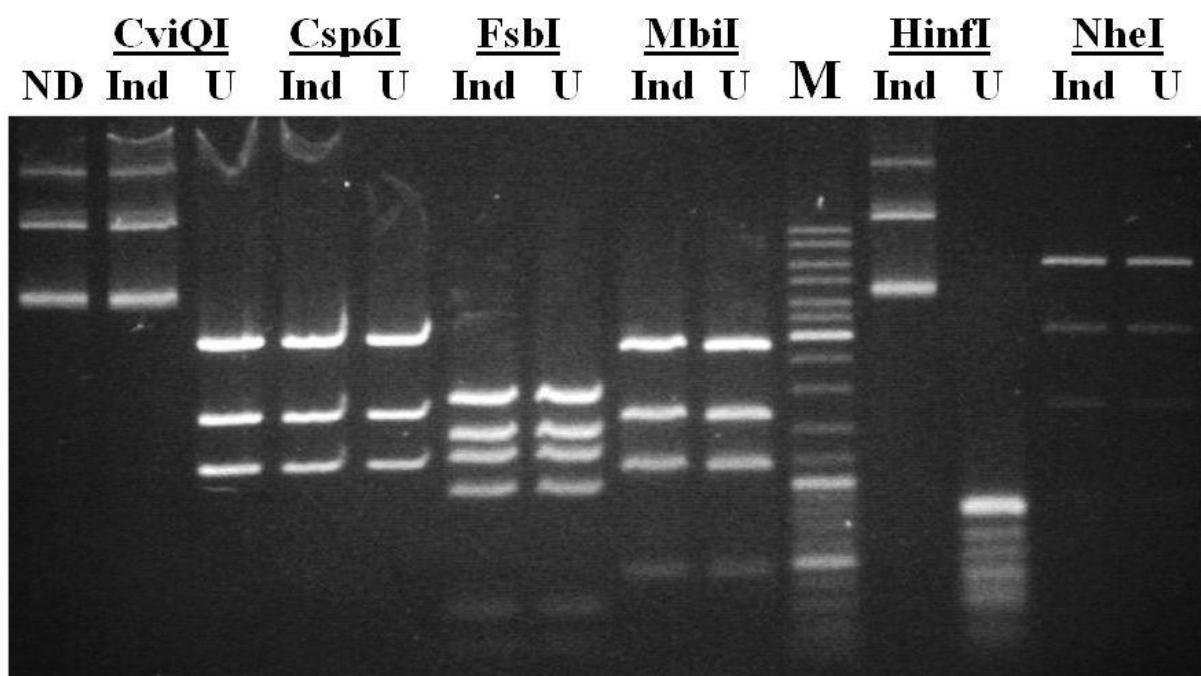
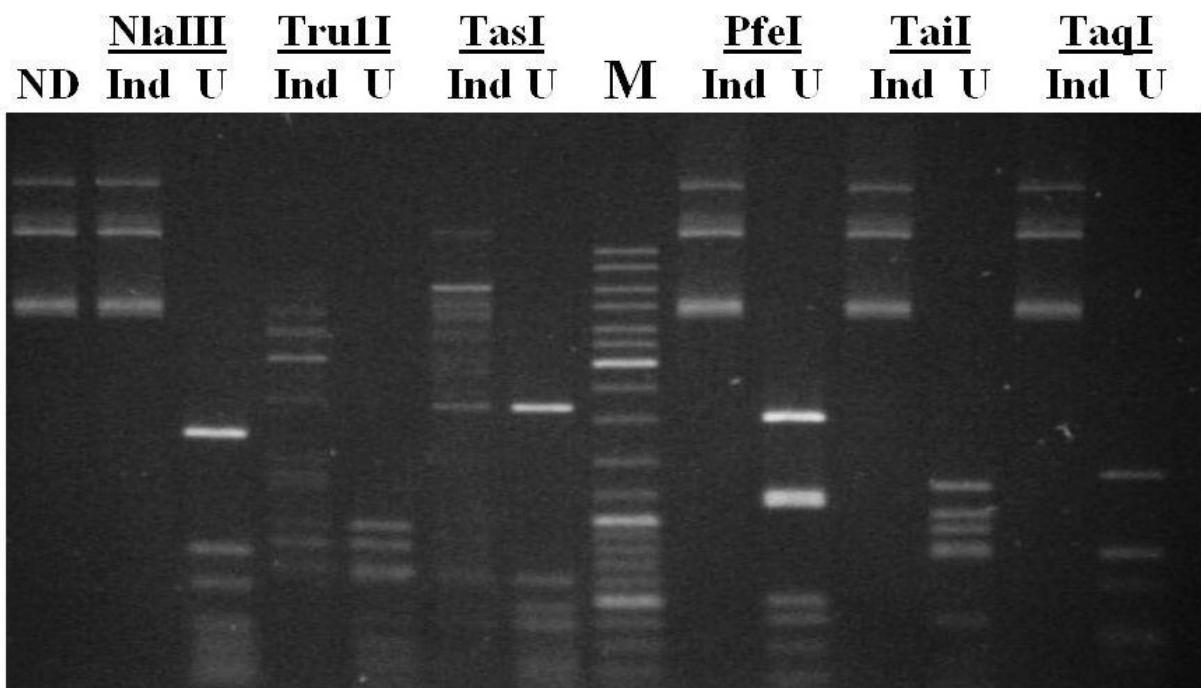
**Figure 2S.** Nucleotide sequence of a 1332-bp fragment of *H. influenzae* biotype aegyptius ATCC 11116 genomic DNA containing the *comHia* and *hia5* genes. The derived amino acid sequences of the ComHia and Hia5 proteins are shown below the DNA sequence.

ComHia5	1 MQSIKAI <u>RCTFCN</u> KLLAKVGMVGYLEIKC <u>PRCK</u> TVNTTR	39
HI1522.1	1 MQSIKVIR <u>CTFCN</u> KLLAKVGIVGYLEIKC <u>PRCK</u> TVNTTR	39
MuCom	1 ---MKS <u>IRCKNC</u> NKLLFKADSF <u>DHIEIRC</u> PRCKRHIIMLNACEHPTEKHCGKREKITHSDET <u>VRY</u>	62

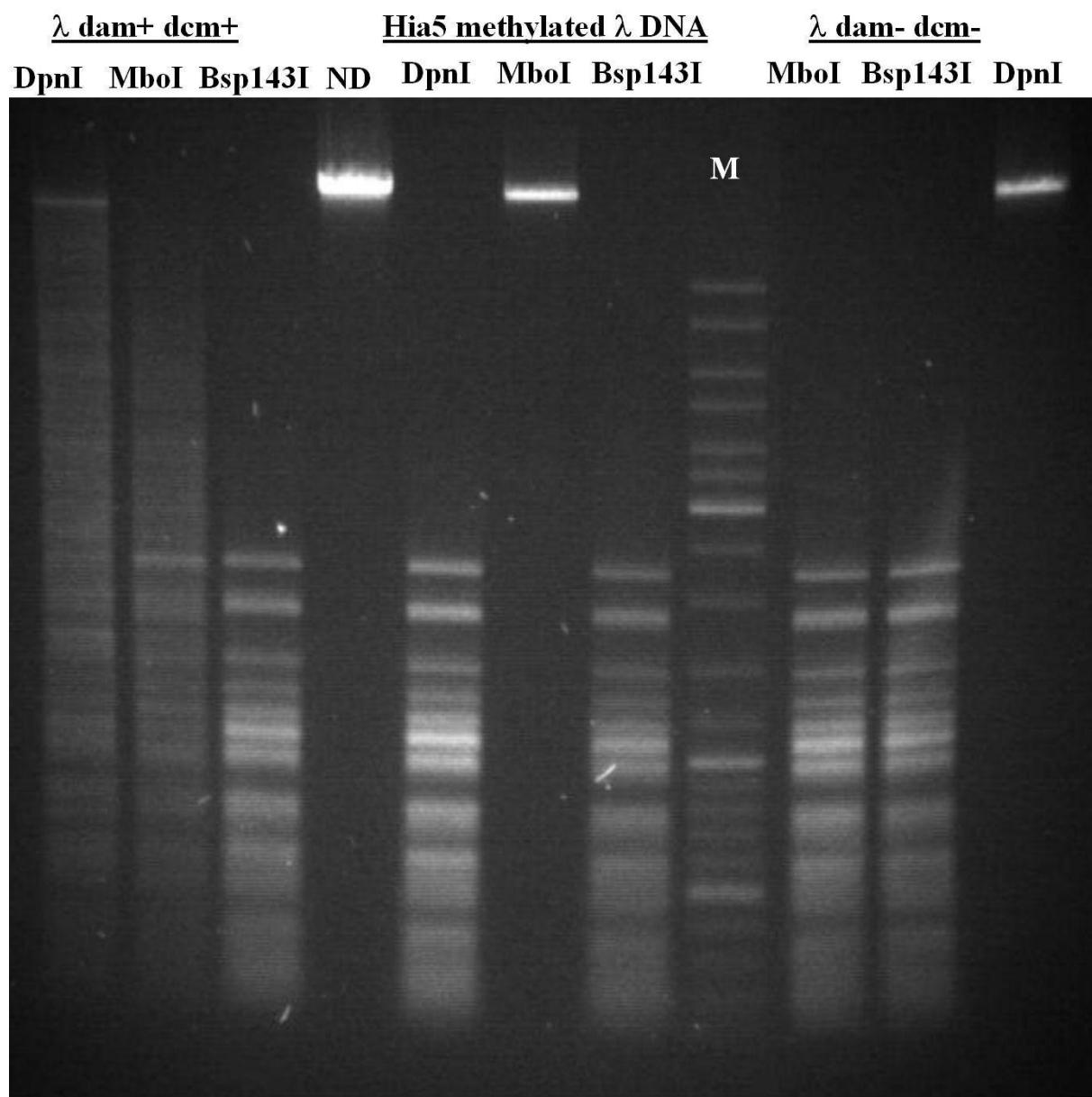
**Figure 3S.** Alignment of the putative Com-like amino acid sequence of *H. influenzae* biotype aegyptius ATCC 11116 with the predicted HI1522.1 Com-like protein of *H. influenzae* Rd and Mu Com. Conserved amino acid residues are shaded green. Four conserved cysteine residues that form the zinc finger are underlined.



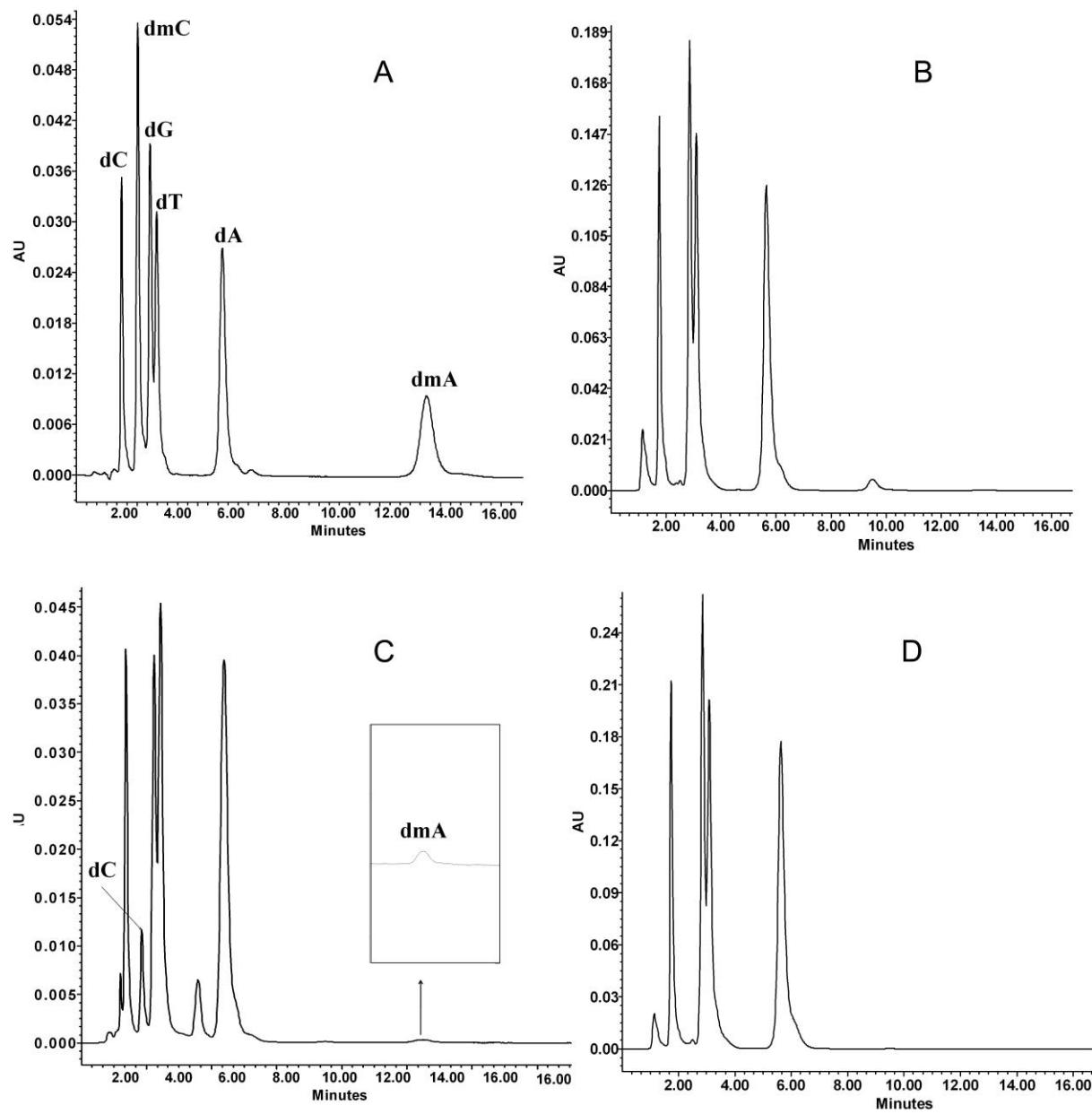
**Figure 4S.** Coomassie-stained SDS-12% PAGE gels of purified recombinant proteins: Hin1523, Nme1821 and their respective mutant variants Hin1523D194A and Nme1821D191A (A), and Hia5 and its mutant variant Hia5D194A (B). M – Molecular mass marker.



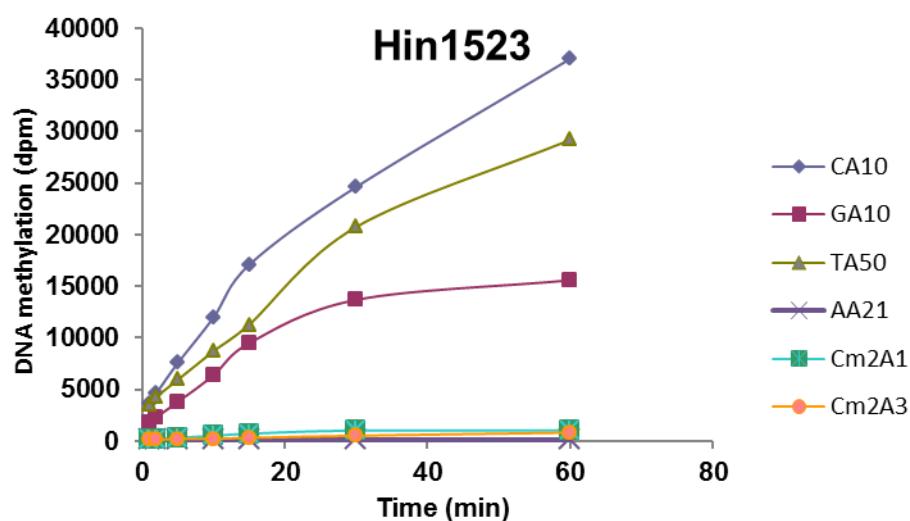
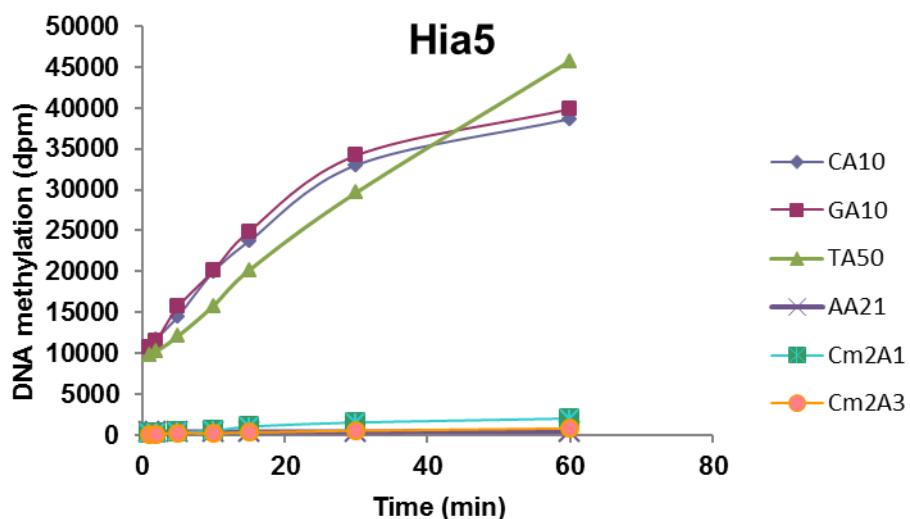
**Figure 5S.** Comparative restriction patterns of plasmid pHia5ET DNA prepared from *E. coli* ER2566 cells grown in the presence (Ind = pHia5ETi) or absence (U = pHia5ET) of inducer IPTG and cleaved with selected REases. Digest mixtures were electrophoresed on 0.8% agarose gels and stained with ethidium bromide. M – GeneRuler 100-10,000 bp size marker.



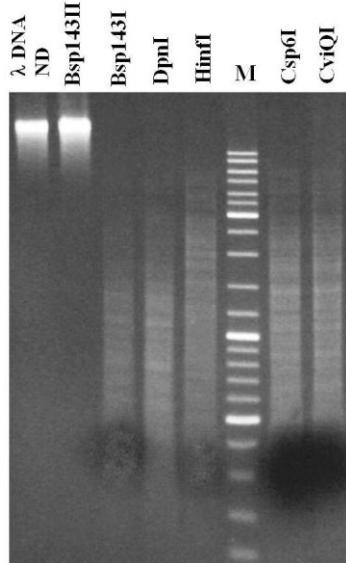
**Figure 6S.** Comparative restriction patterns of  $\lambda$  DNA dam<sup>+</sup> dcm<sup>+</sup>,  $\lambda$  DNA dam- dcm- (controls) and Hia5 *in vitro*-methylated  $\lambda$  DNA cleaved with REases MboI, DpnI and Bsp143I. Digest mixtures were electrophoresed on 0.8% agarose gels and stained with ethidium bromide. ND- undigested  $\lambda$  DNA. M – GeneRuler 100-10,000 bp size marker.



**Figure 7S.** HPLC profiles of deoxynucleosides in the standards mixture (A), and enzymatic digests of  $\lambda$  DNA dam+ dcm+ (B), the genomic DNA of *H. influenzae* biotype aegyptius ATCC 11116 (C) and  $\lambda$  DNA dam- dcm- (D).



**Figure 8S.** Example of DNA methylation kinetics with the Hia5 or Hin1523 enzymes and synthetic oligonucleotides (**Table 2S**).



**Figure 9S.** Restriction patterns of *H. influenzae* biotype aegyptius ATCC 11116 genomic DNA cleaved with selected REases: Bsp143II (RGCGCY), Bsp143I (GATC), DpnI ( $\text{Gm}^6\text{ATC}$ ), Hinfl (GANTC), Csp6I (GTAC) and CviQI (GTAC). Digest mixtures were electrophoresed on 0.8% agarose gels and stained with ethidium bromide. ND- undigested  $\lambda$  DNA. M – GeneRuler 100-10,000 bp size marker.