Supplementary Information for

Pathways of thymidine hypermodification

Yan-Jiun Lee, Nan Dai, Stephanie I. Müller, Chudi Guan, Mackenzie J. Parker, Morgan E. Fraser, Shannon E. Walsh, Janani Sridar, Andrew Mulholland, Krutika Nayak, Zhiyi Sun, YuCheng Lin, Donald G. Comb, Katherine Marks, Reyaz Gonzalez, Daniel P. Dowling, Vahe Bandarian, Lana Saleh, Ivan Corrêa Jr., and Peter R. Weigele

Correspondence to: weigele@neb.com

This PDF file includes:

Figures S1 to S21
Table S1

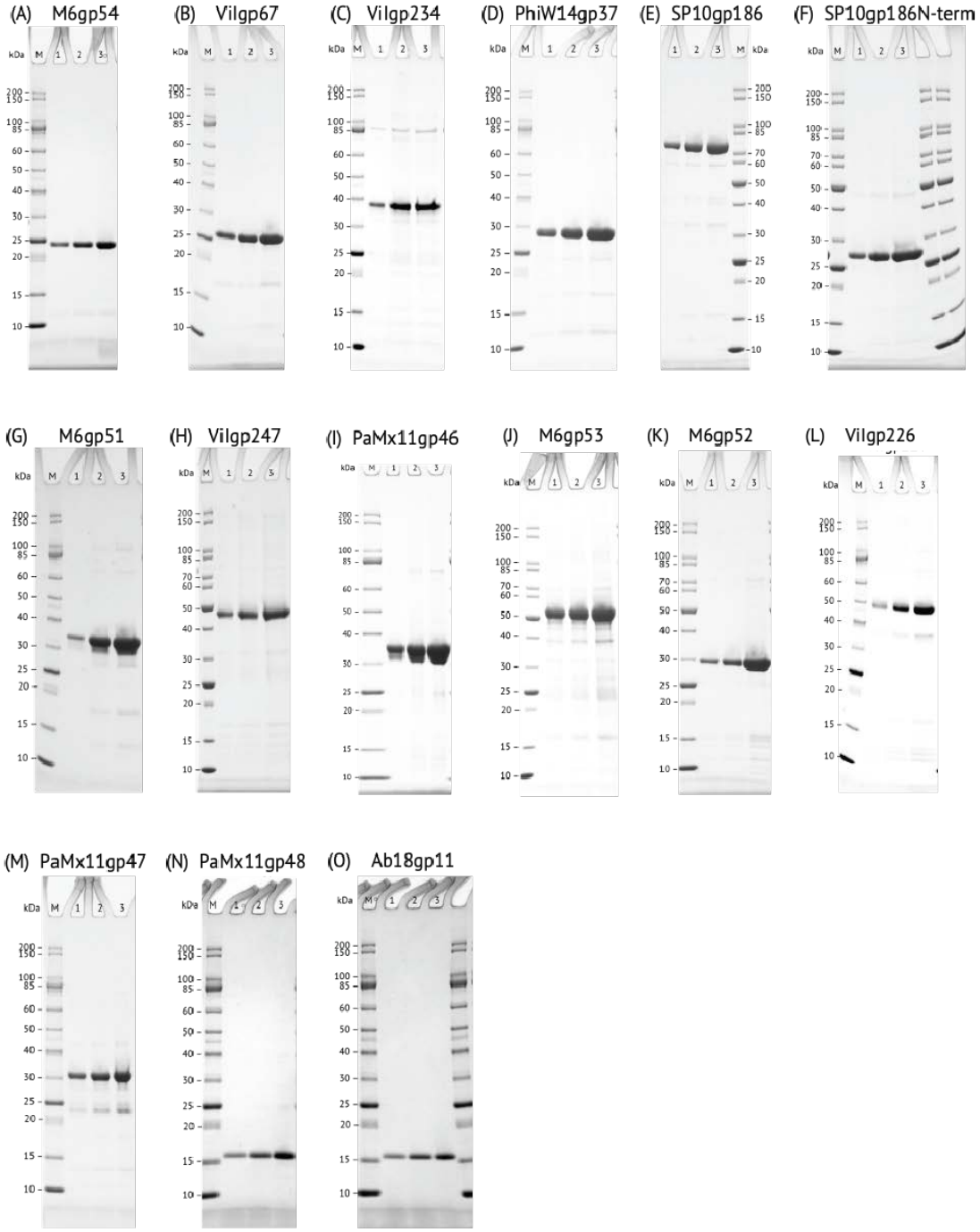


Fig. S1. SDS-PAGE analysis of recombinant proteins purified in this study.
(A) through (O) are images of purified proteins analyzed using SDS-PAGE. After electrophoresis, tris-glycine SDS-PAGE gels (10–20%) were stained with Coomassie G-250 and documented using CCD camera of the digital gel documentation system. M, the protein standard ladder. Lane 1, 2, and 3, the purified protein in 1, 2, and 4 µg protein loading, respectively, in which the protein concentrations were determined by Bradford protein assay.

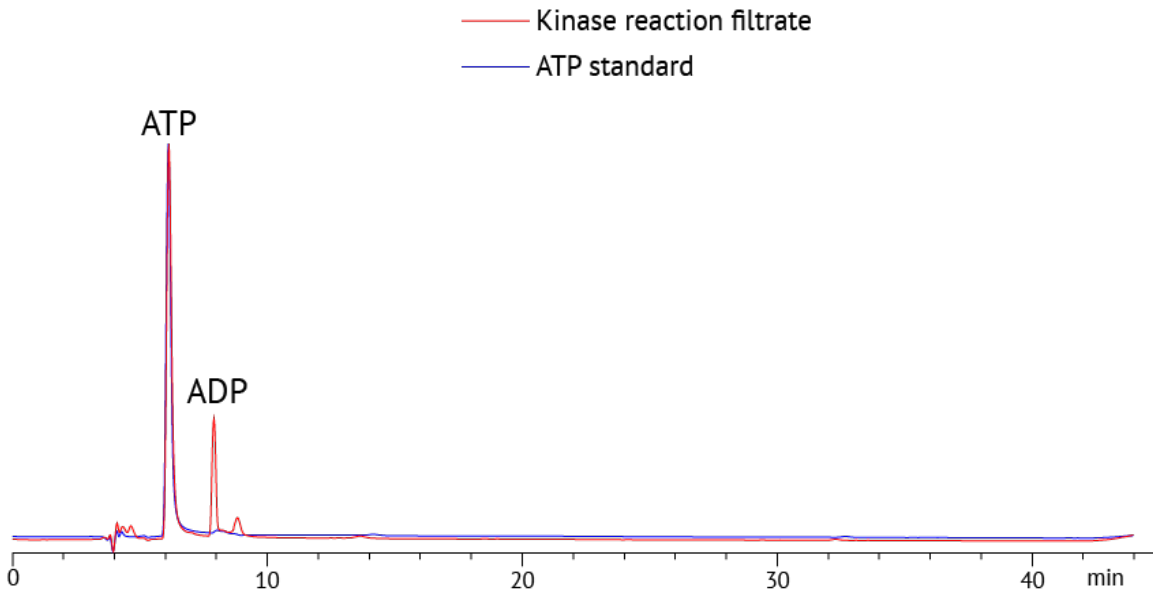


Fig. S2. Production of ADP during 5-HMUDK reaction.
A mixture containing 6.5 µg of purified SP8 genomic DNA, 100 µM ATP, and 300 ng of 5-HMUDK was incubated in 1X T4 DNA ligase buffer in a total volume of 60 µl for 1 hour at 30°C. After incubation, samples were filtered through a 5K MWCO spin concentrator (Millipore) and the effluent was analyzed by LC/MS. Two species, ATP and ADP were detected.

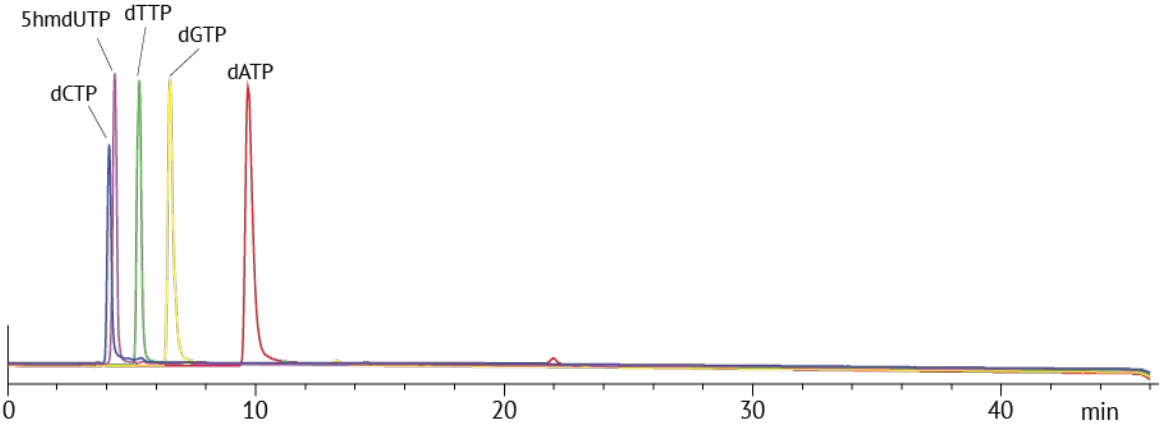


Fig. S3. Resolution of nucleotide triphosphates by the HPLC method used in this study. Each of the nucleotide triphosphate acquired from commercially available source in 100 mM solution was prepared as 1 mM solution in the 1x nucleoside digestion buffer. Then, the 1 mM solutions were incubated at 37°C for 1.5h and approximately 2.5 nmole from each of the solution was subjected to LC/MS analysis using the standard LC method in this study. The resulted chromatography traces were overlayed. From earlier to later retention times, the order of elution is: dCTP, 5-hmdUTP, dTTP, dGTP, and dATP.

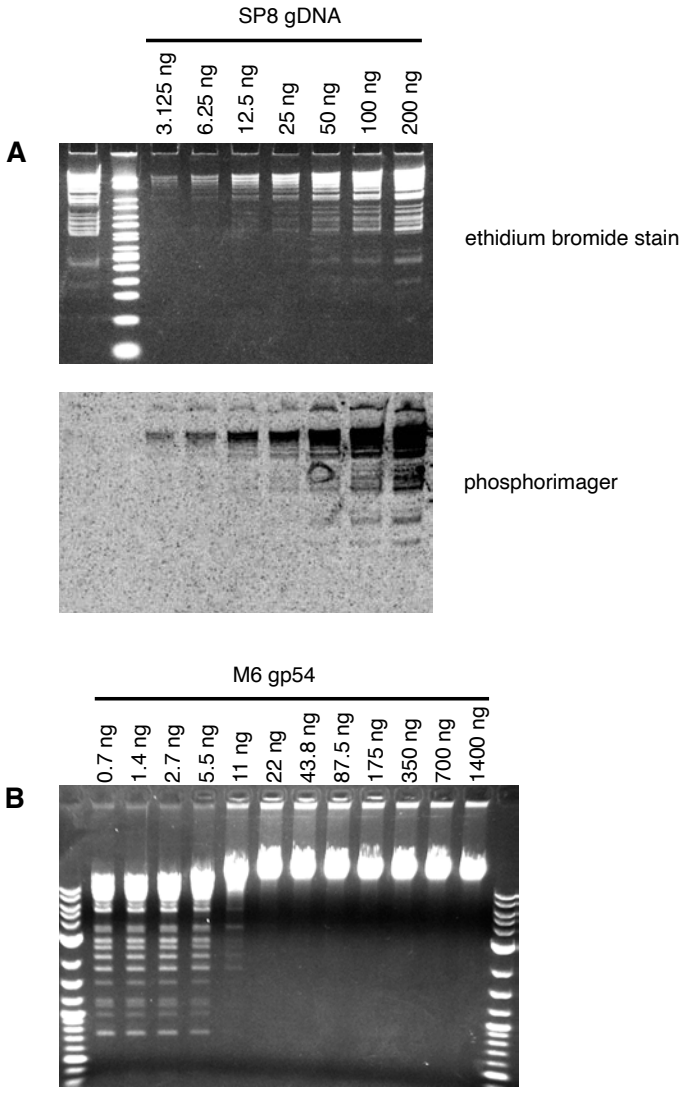


Fig. S4. 5-HMUDK (M6 gp54) transfers a gamma-phosphate from ATP to 5-hmdU in DNA.

(A) Samples from a 2-fold dilution series of XbaI digested SP8 genomic DNA spanning 200 ng to 3.125 ng were incubated with 1 µg of purified M6 gp54 and 25 µCi [γ-^{32}P]ATP in a 50 µl reaction containing 50 mM Tris pH 7.5, 10 mM MCl_2, and 50 µM cold ATP for 1 hour at 30°C. Reactions were quenched with 1X Gel loading dye (NEB #B7024) and electrophoresed through a 1% agarose gel followed by staining with ethidium bromide/UV illumination as well as visualization of radioactivity using a GE Typhoon phosphorimager. **(B)** Reactions containing 200 ng intact SP8 genomic DNA, 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and a 2-fold dilution series of M6 gp54 spanning 1.4 µg to 0.7 ng of enzyme were incubated for one hour at 30°C after which 10 U of NdeI were added to each reaction and incubated for an additional 30 minutes at 37°C before being quenched with 1X NEB Gel loading dye and electrophoresed through 1% agarose, stained and visualized by ethidium bromide/UV illumination.

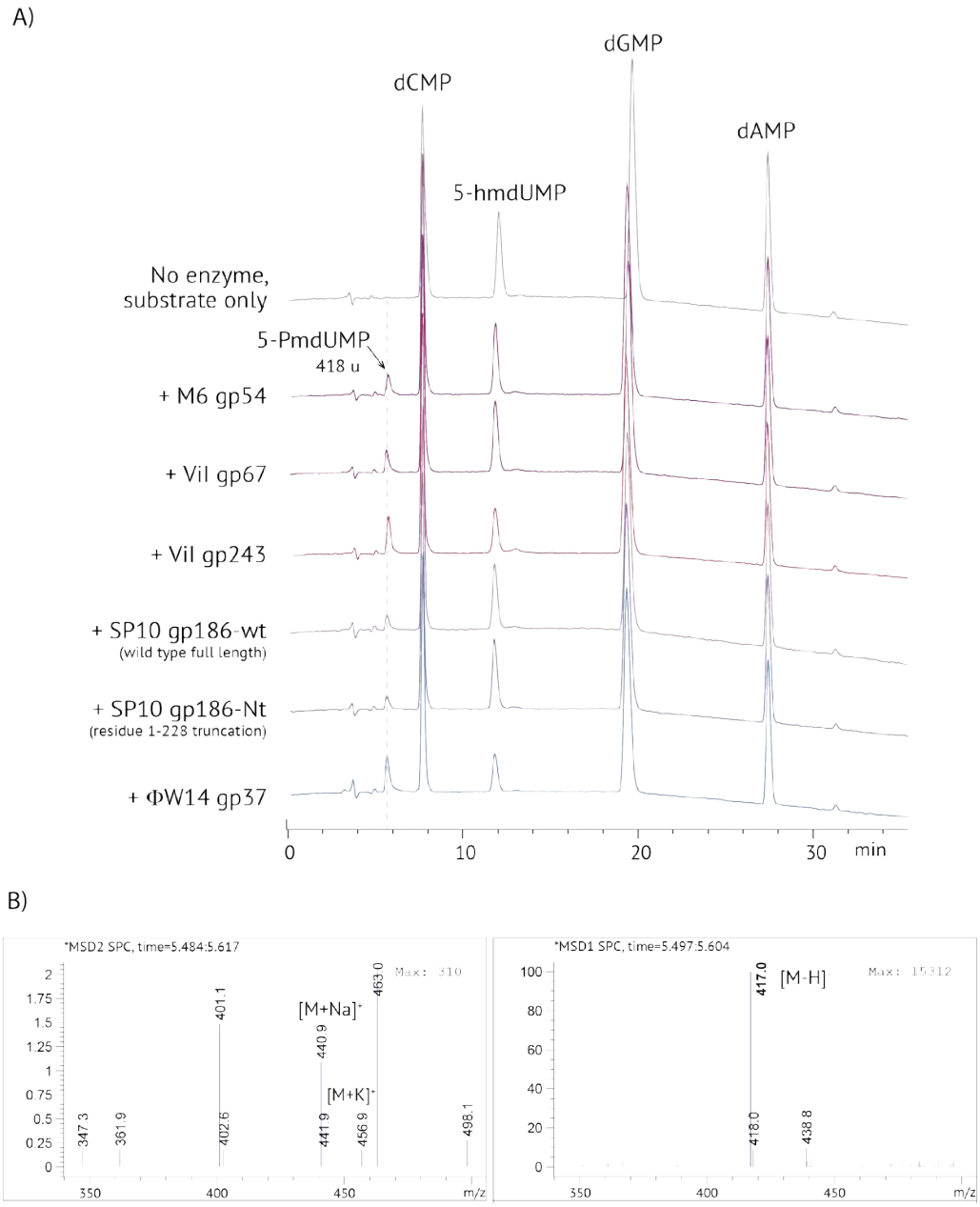


Fig. S5. Production of 5-PmdUMP by phage encoded 5-hmdU DNA kinases (5-HMUDKs). (A) Genomic DNA purified from bacteriophage SP8 (which naturally contains 5-hmdU fully replacing T) was restriction enzyme-digested and incubated with indicated enzyme in the presence of 1 mM ATP for 1 hour at 37 °C. Following incubation, DNA was recovered by spin column purification (Monarch PCR Cleanup Kit, NEB), digested to free nucleotides using phosphatase-free nucleoside digestion mix (NEB), and resolved by LC/MS. The top-most trace derives from substrate incubated in a no-enzyme control reaction. The phosphorylated nucleotide product 5-PmdUMP from the kinase reaction elutes at approximately 5.5 minutes using the LC method described in the method session. (B) The mass signal of 5-PmdUMP nucleotide product from the SP8 DNA reaction with M6 gp54. Positive (+ESI) and negative (-ESI) electrospray ionization modes are presented at left and right panels, respectively.

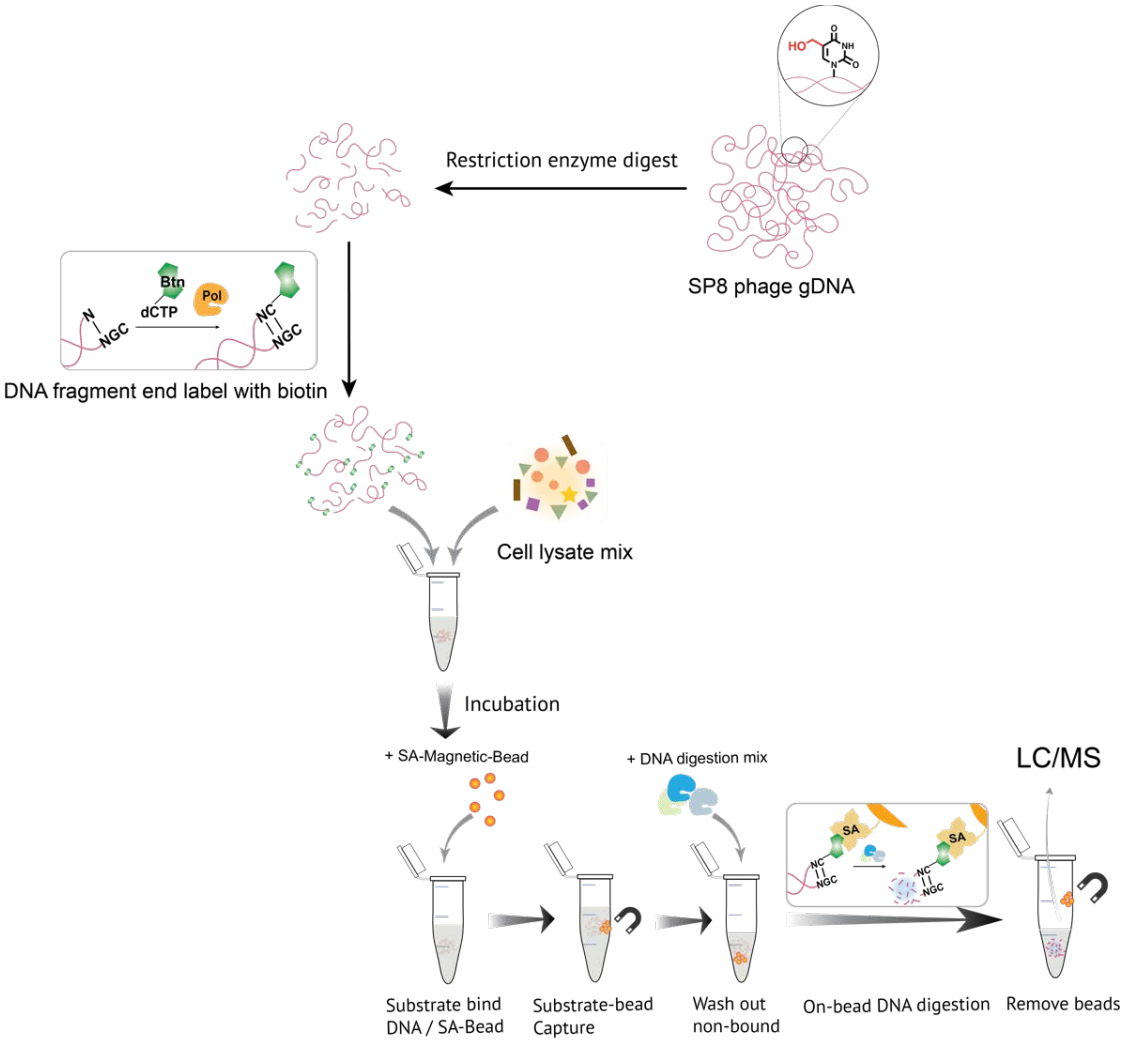


Fig. S6. General scheme of *ex vivo* hypermodifying enzyme activity assay.

SP8 genomic DNA (gDNA) extracted from Bacillus phage SP8 was digested with restriction enzyme. The resulted product was labeled with biotinylated nucleotide through the polymerase and purified by silica-based spin column. To perform the *ex vivo* enzyme modification assay, the biotinylated DNA was challenged with *E. coli* lysate which was obtained from the strain transformed with the plasmid harboring the putative thymidine base modifying enzyme gene. The lysate can be prepared as the mixture of the combination of several lysates each different strain with different base modifying enzyme genes. Necessary cofactors can be provided in the reaction buffer. After the reaction in the lysate, the gDNA product was diluted with bead-binding buffer. The mixture was then transferred to the preconditioned streptavidin magnetic beads (SA-magnet-bead). The gDNA-bead complex was capture by magnet and was further washed with washing buffer, three times. To release the nucleoside from the gDNA-bead complex for HPLC-MS analysis, nucleoside digestion mix (NEB), the cocktail of nucleases and phosphatase reagent, was added to the solution and left for incubation. After digestion step, the beads were captured by magnet, and the supernatant was collected and subjected for HPLC-MS analysis.

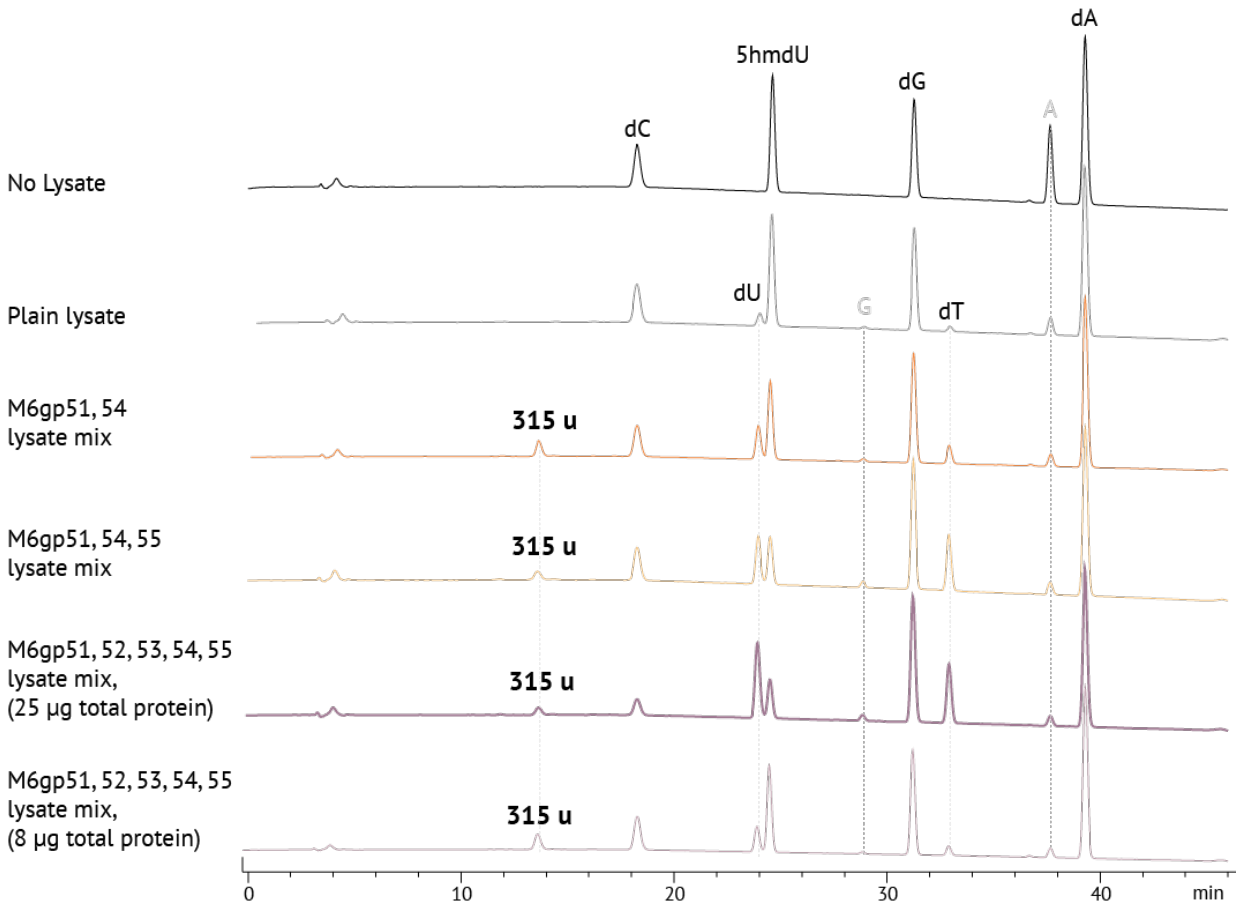


Fig. S7. Thymidine hypermodification *ex vivo* assay on M6 gene products.
Biotinylated fragments of SP8 genomic DNA prepared as described in Methods were incubated in buffer (trace 1), plain lysate (trace 2), or combinations of lysates derived from cells expressing the indicated phage M6 gene products. Incubation of substrate DNA with lysates containing gp51 and gp54 was sufficient to produce a new nucleoside species with 315 u nominal mass.

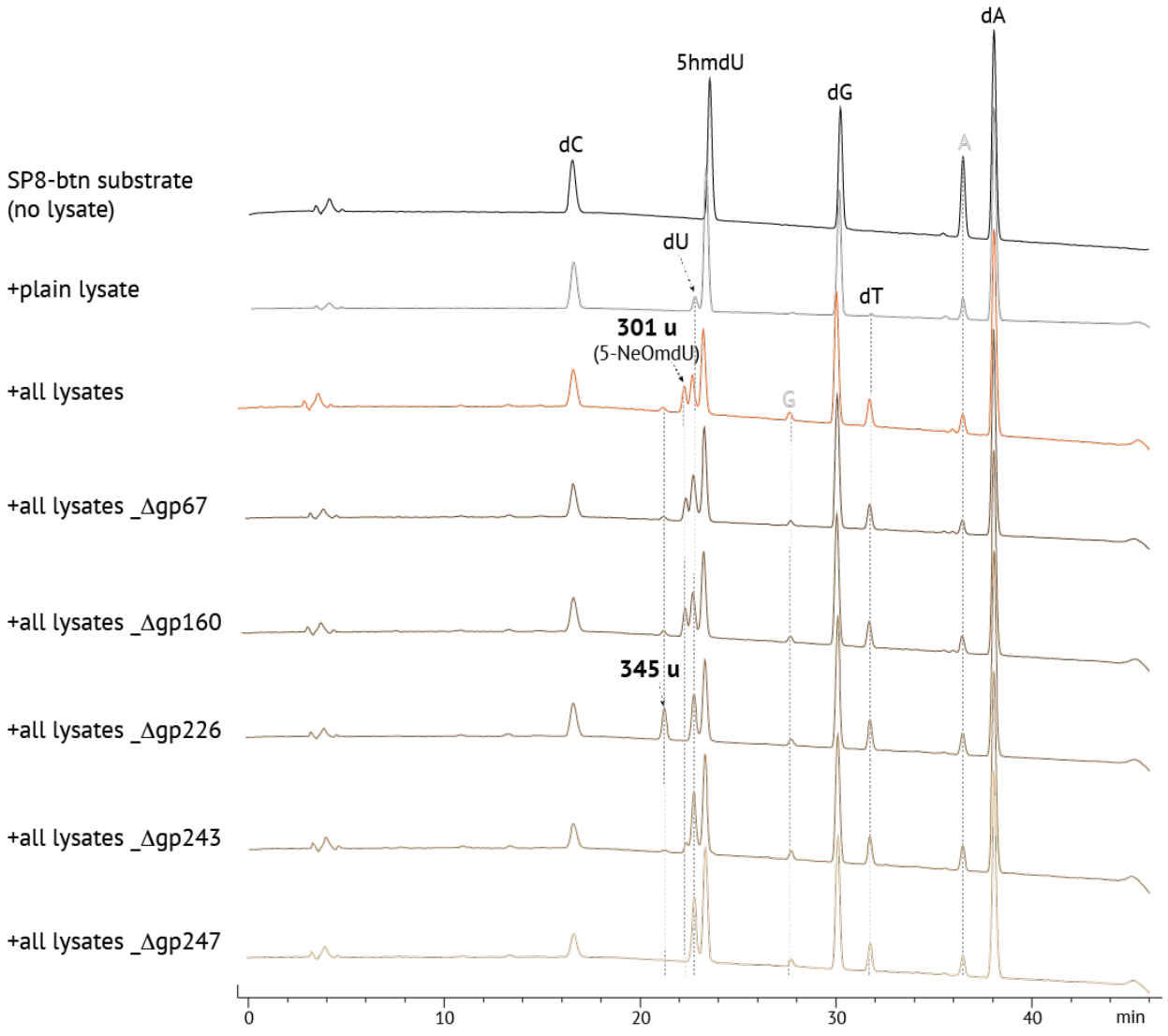


Fig. S8. Thymidine hypermodification *ex vivo* assay on phage ViI gene products.
Biotinylated fragments of SP8 genomic DNA prepared as described in Methods were incubated in buffer (trace 1), plain lysate (trace 2), or combinations of lysates derived from cells expressing the indicated phage ViI gene products. Incubation of substrate DNA with lysates containing five gene products (gp67, gp160, gp226, gp243, and gp247) could form the mature modification 5-*Ne*O*m*dU as seen in the third trace from the top. The presence of either gp67 or gp243, both predicted 5-hmdU DNA kinases, together with gp160, gp226, and gp247 was sufficient to produce 5-*Ne*O*m*dU. The gp226 drop-out reaction produced a nucleoside with nominal mass of 345 u, which is 44 u greater than the mature product, 5-*Ne*O*m*dU.

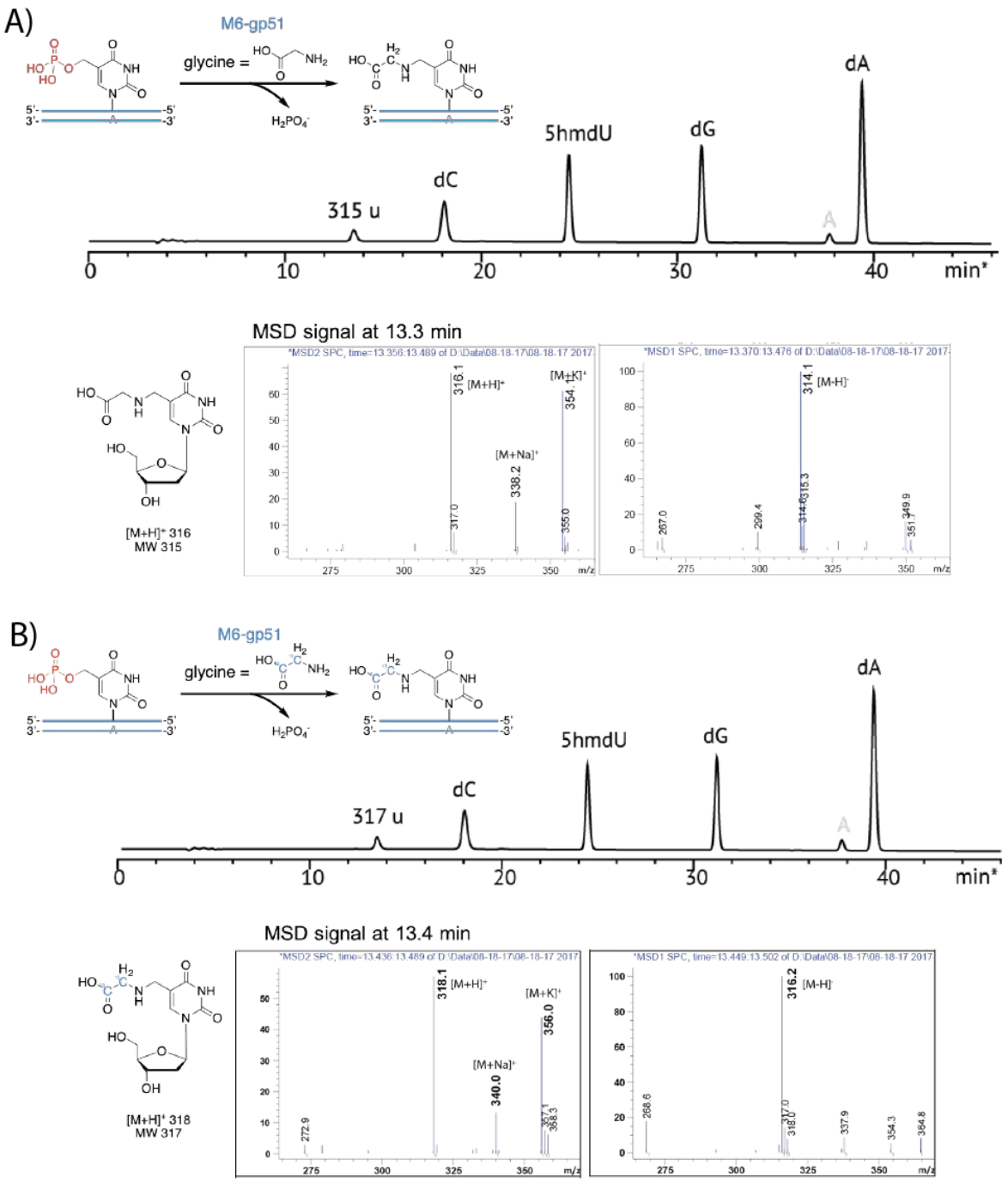


Fig. S9. Stable isotope labeling experiment of M6 gp54, M6 gp51 and ^{13}C labeled glycine.
DNA containing 5-hmdU fully substituting for dT was incubated with M6 gp54 (5-hmdU DNA kinase) and M6 gp51 in the presence of 1 mM glycine (top trace and mass spectra) or 1 mM stable isotope labeled glycine (13C2, 97-99%) (Cambridge Isotope Laboratory, Inc., Tewksbury, MA) (lower trace and spectra). The formation of a 317 u nucleoside in the presence of labeled glycine confirms that the added mass is derived from glycine.

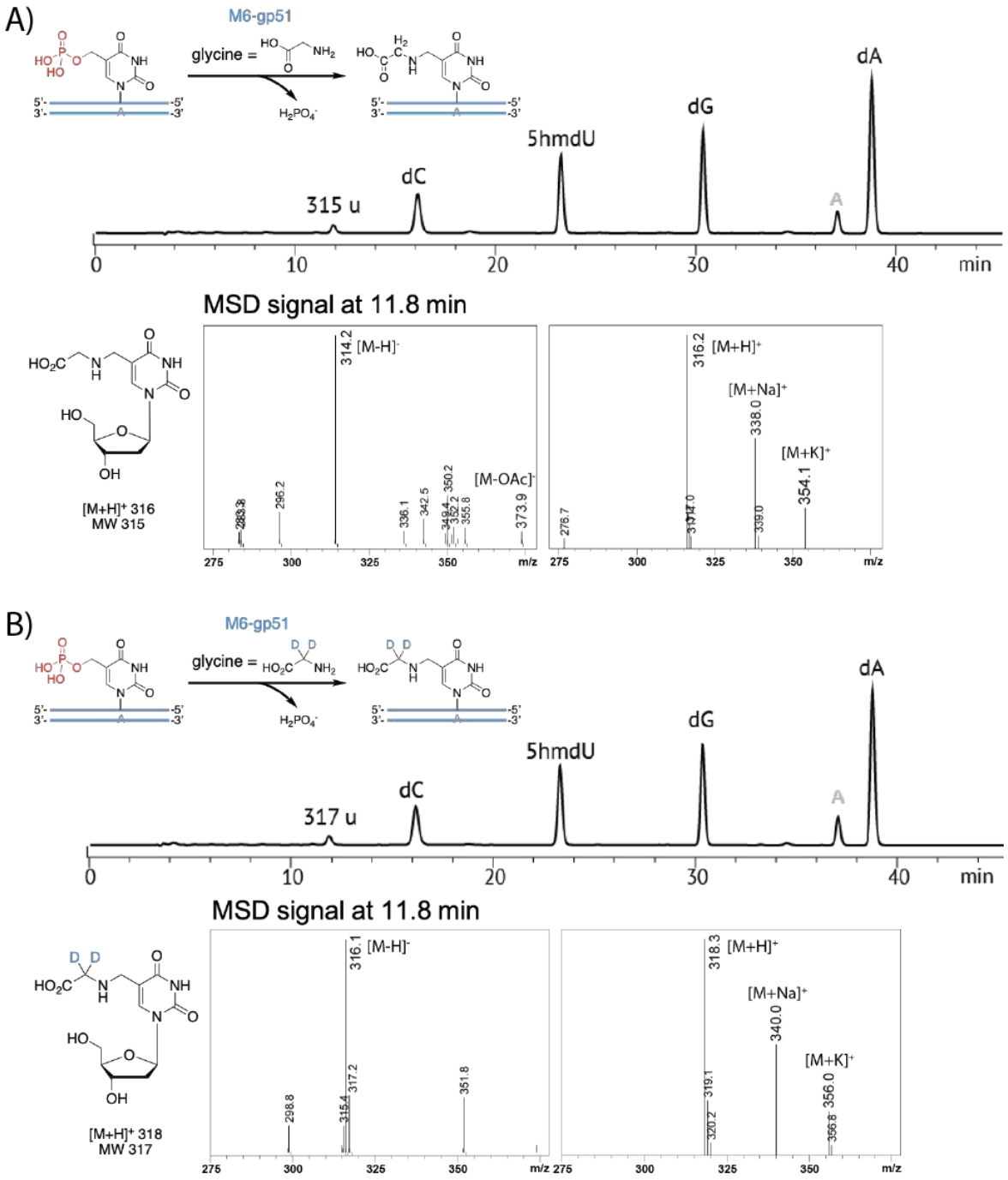


Fig. S10. Incorporation of deuterium labeled substrate into DNA shows retention of both protons at alpha carbon of glycine.

DNA containing 5-hmdU fully substituting for dT was incubated with M6 gp54 and M6 gp51 in the presence of 1 mM glycine (top trace and mass spectra) or 1 mM stable isotope labeled glycine (2,2-D2, 98%) (Cambridge Isotope Laboratory, Inc., Tewksbury, MA) (lower trace and spectra). The formation of a 317 u nucleoside in the presence of the doubly deuterated glycine indicates that both deuteriums are retained on the alpha carbon following group transfer to the base.

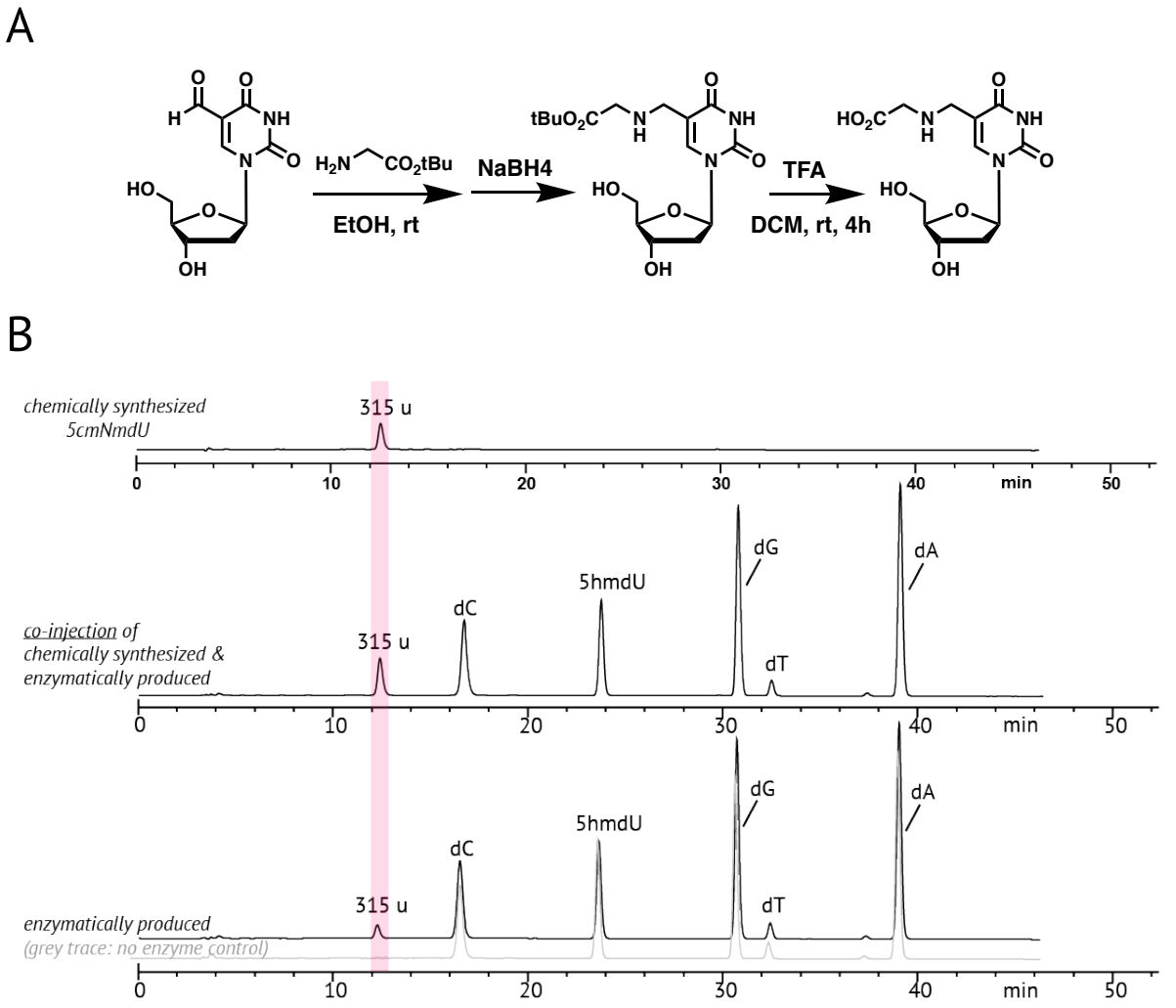


Fig. S11. Confirmation of N^{α}-glycinylthymidine formation by chemical synthesis of nucleoside standard.

N^{α}-GlyT was synthesized by reductive amination of 5-formyluridine as described in Methods and indicated in by the scheme in panel (**A**). As shown in panel (**B**), the synthetic N^{α}-GlyT is indistinguishable from the enzymatically produced nucleoside both in nominal mass and retention time.

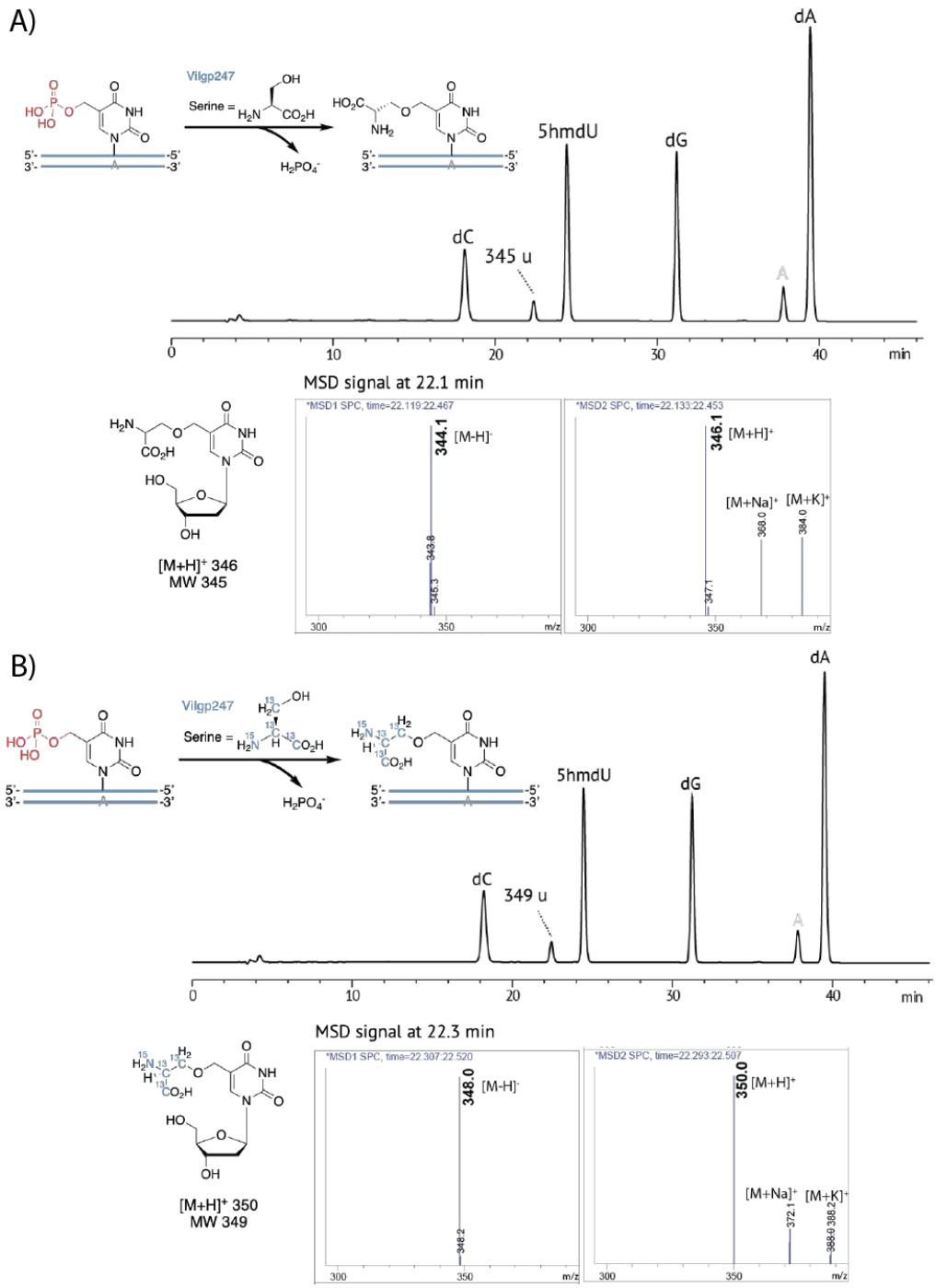
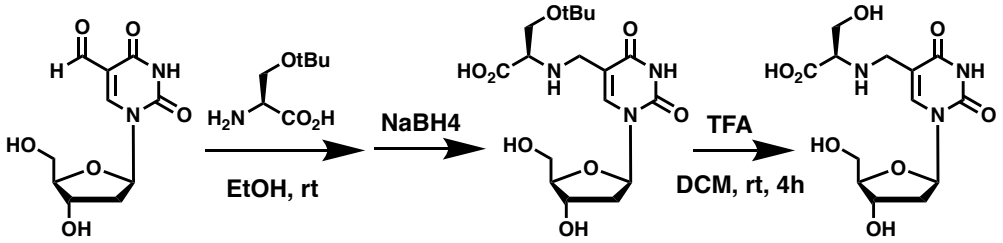


Fig. S12. Stable isotope DNA labeling experiment using recombinantly expressed and purified enzymes from phage ViI and ^{13}C labeled serine.
The DNA substrate containing 5-PmdU was incubated with purified ViI gp247 and 1 mM L-serine (**A**) or 1 mM stable isotope labeled L-serine (13C3, 99%; 15N, 99%) (Cambridge Isotope Laboratory, Inc., Tewksbury, MA) (**B**) and products were digested to free nucleosides and subjected to LC/MS analysis. LC/MS shows production of nucleoside with nominal mass of 345 u for (**A**) but with stable isotope labeled serine (**B**) yielded a nucleoside with nominal mass of 349 u.

A



B

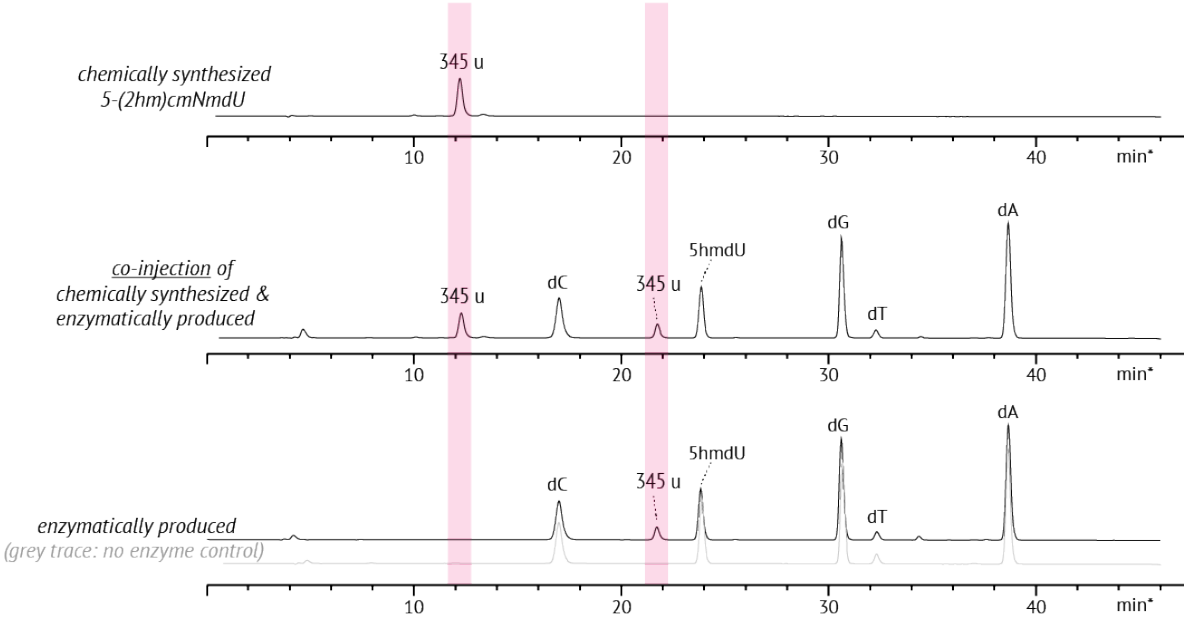


Fig. S13. Chemical synthesis shows N^{α}-serinylthymidine is not the isomer produced by ViI gp247.

N^{α}-SerT was synthesized by reductive amination of 5-formyluridine as described in Methods and depicted in the scheme shown in (**A**). As shown in (**B**), the synthetic N^{α}-SerT has a different retention time, albeit identical nominal mass, compared to the enzymatically produced nucleoside.

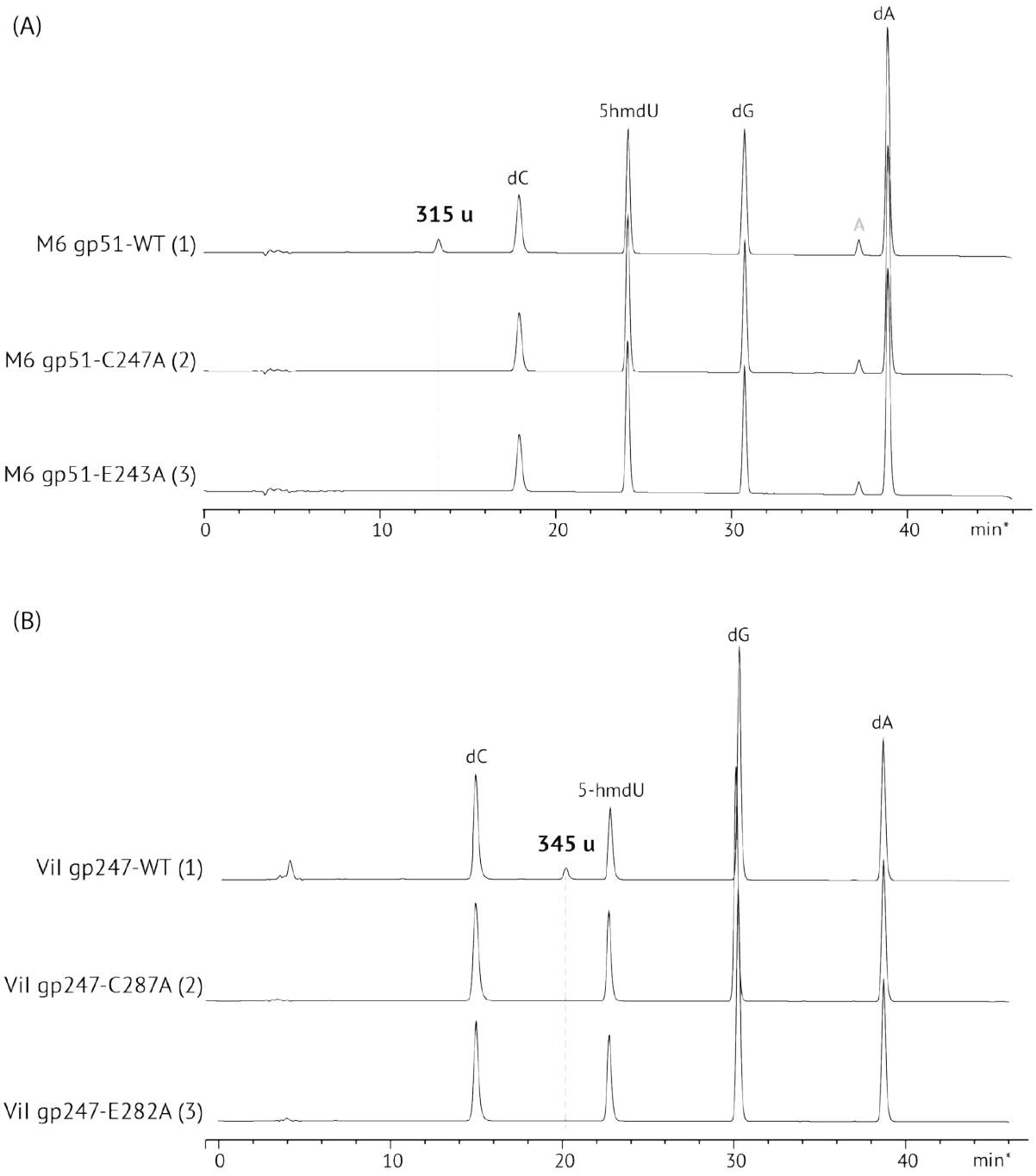


Fig. S14. Mutation of conserved residues in presumptive active abolish activity of M6 gp51 and ViI gp247.

Clade2 aGPT-Pplase proteins contain strictly conserved glutamate and cysteine residues predicted to be involved in enolization of *O*4 and Michael addition at *C*6, respectively. Wild-type and mutant enzymes were incubated with substrate 5-PmdU containing DNA as described in methods and the products analyzed by LC/MS of nucleoside digests. Mutation of conserved presumptive active-site residues in M6 gp51 (**A**) and ViI gp247 (**B**) abrogate enzyme function.

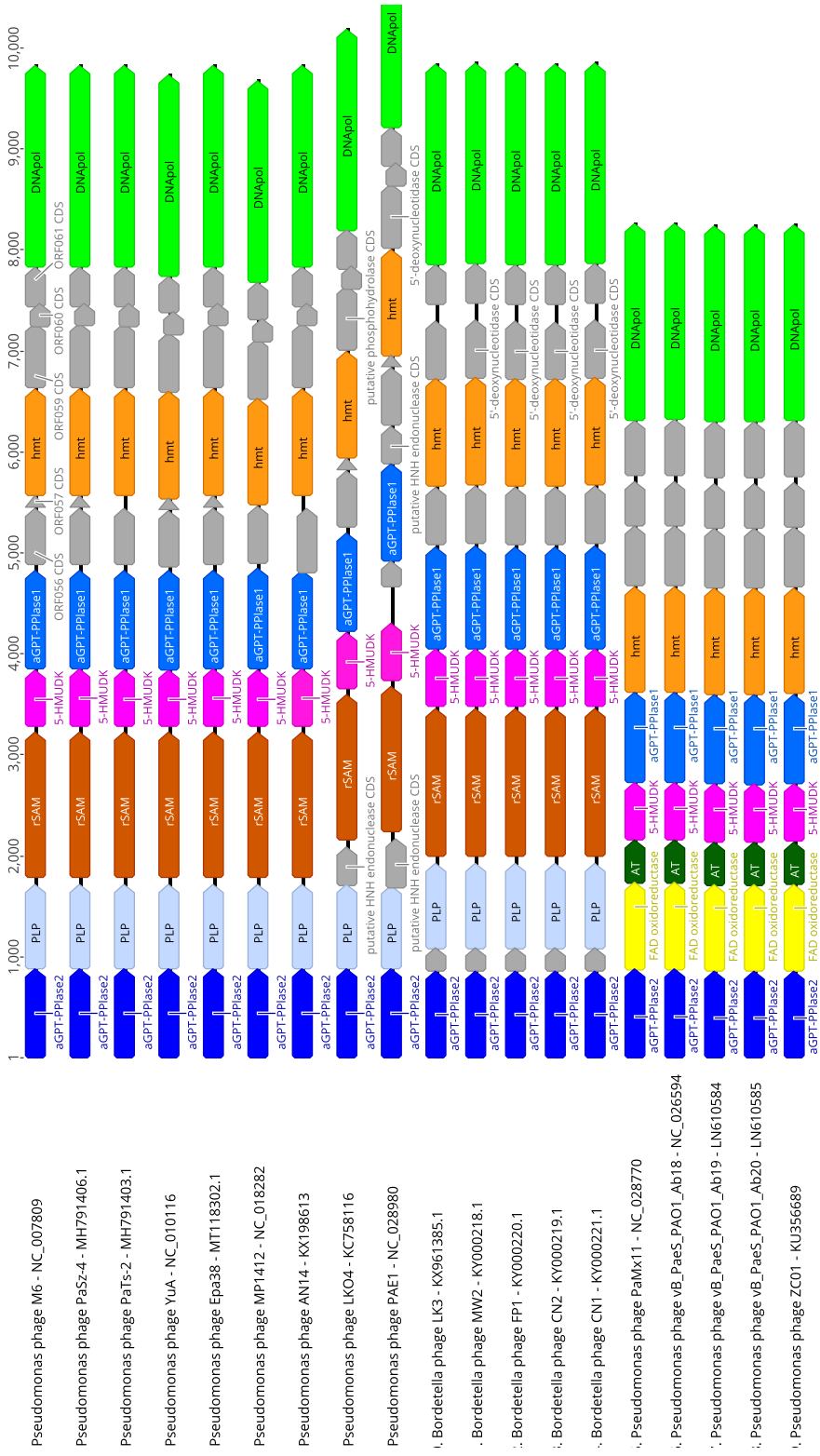


Fig. S15. Thymidine-hypermodification gene clusters of M6-like phages infecting *Pseudomonas* and *Bordetella*.

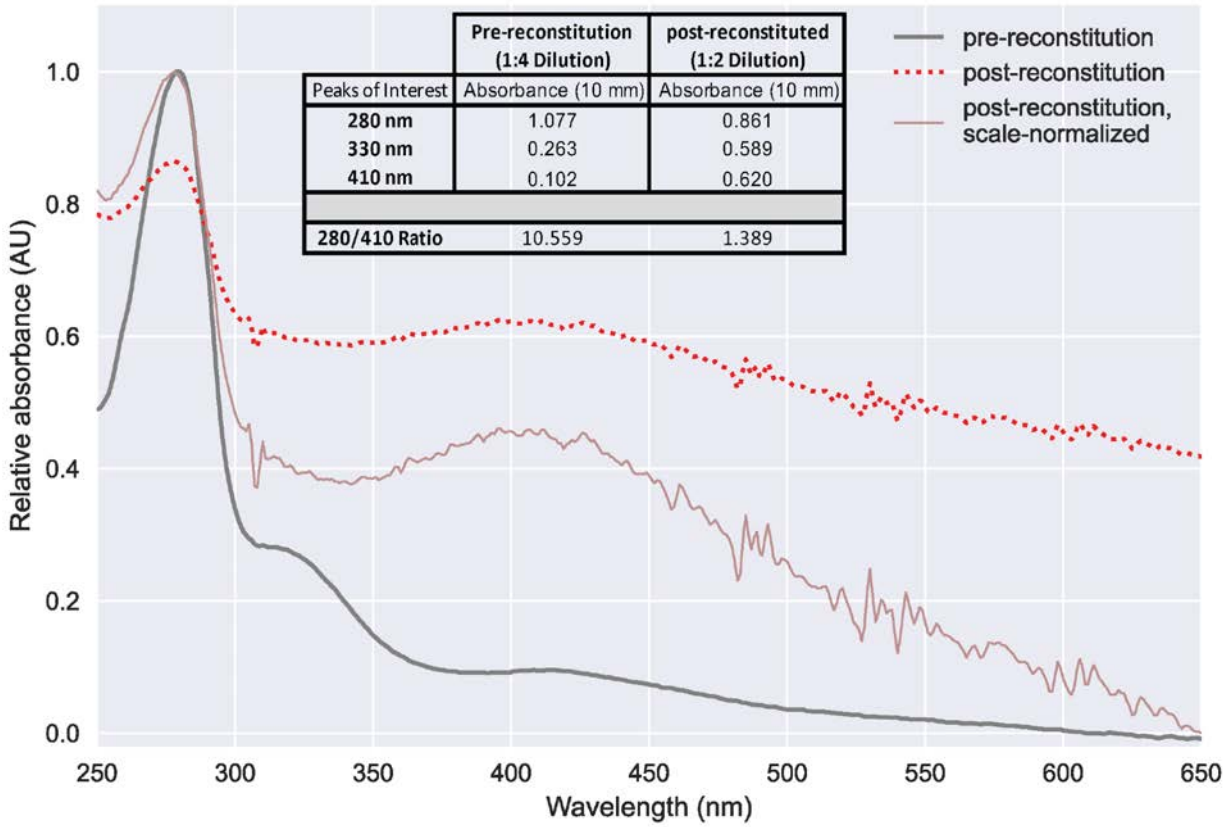


Fig. S16. UV-Vis absorbance spectra of purified and reconstituted M6 gp53.
UV-Vis absorbance spectrum of purified M6 gp53 shows increase in absorbance in ~410 nm region following anaerobic iron-sulfur cluster reconstitution as described in methods. Determination of iron content by ferrozine assay yielded ~9 Fe per protein, suggesting two 4Fe-4S clusters per monomer.

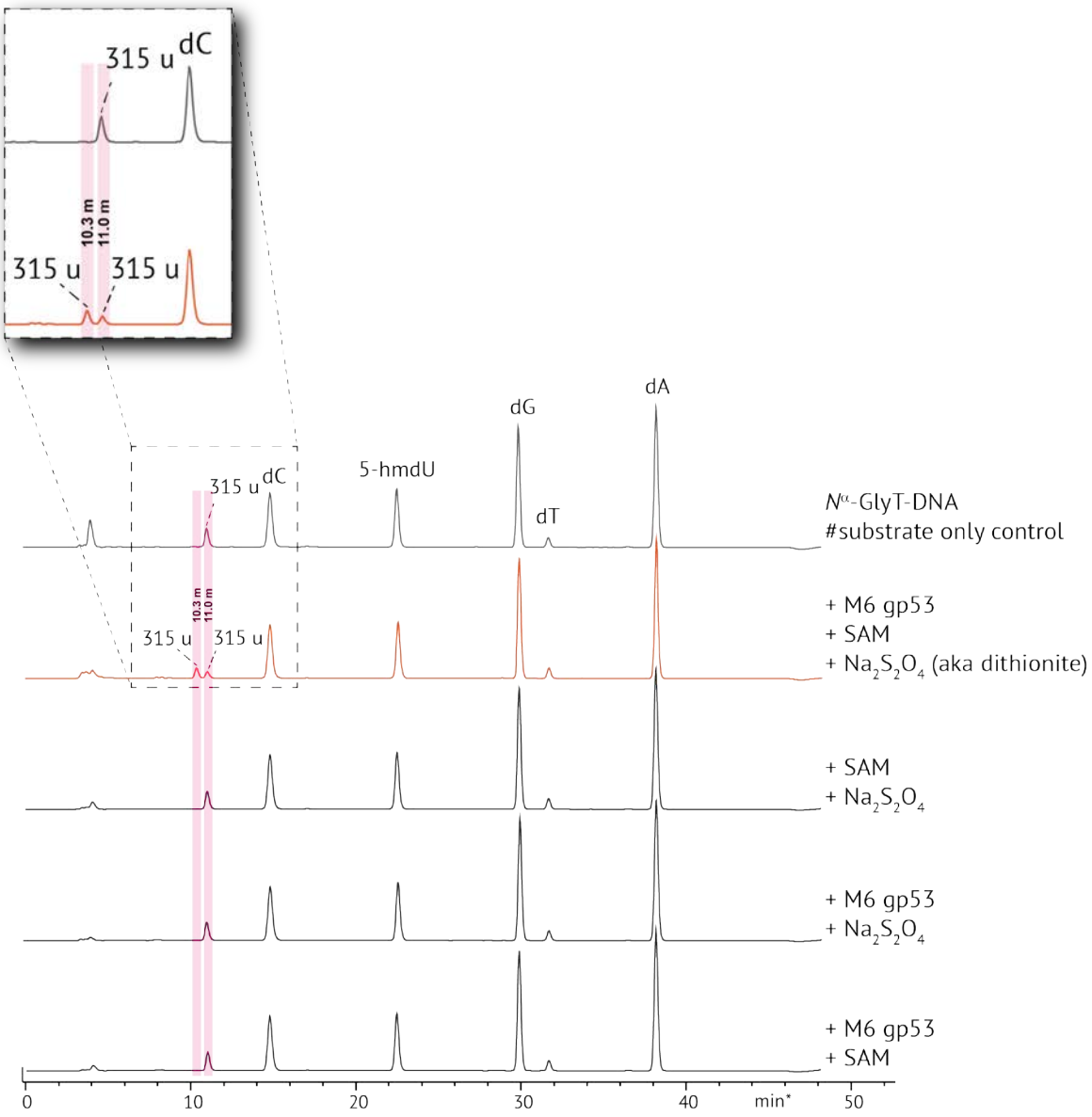
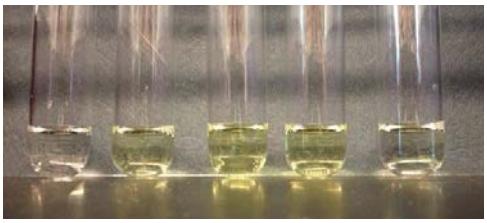


Fig. S17. M6 gp53 catalyzed isomerization of N^{α}-GlyT is dependent on SAM and dithionite in vitro.

Substrate DNA containing N^{α}-GlyT was prepared by incubation of 5-hmdU DNA with M6 gp54 (5-HMUDK) and M6 gp51 (Pfam: aGPT-Pplase2, also coined as AA:DNA transferase) and purified by spin column (Monarch PCR cleanup kit, NEB). N^{α}-GlyT in substrate confirmed by LC/MS analysis of nucleoside digests as shown in the first trace from the top. N^{α}-GlyT substrate was reacted with gp53 in the presence of SAM and sodium dithionite, producing C^{α}-GlyT, as shown in the second trace. Reactions leaving out enzyme, SAM, or sodium dithionite failed to produce C^{α}-GlyT, as shown in the third, fourth, and fifth traces, respectively.



M6gp52 FPLC fractions

Vilgp226 FPLC fractions

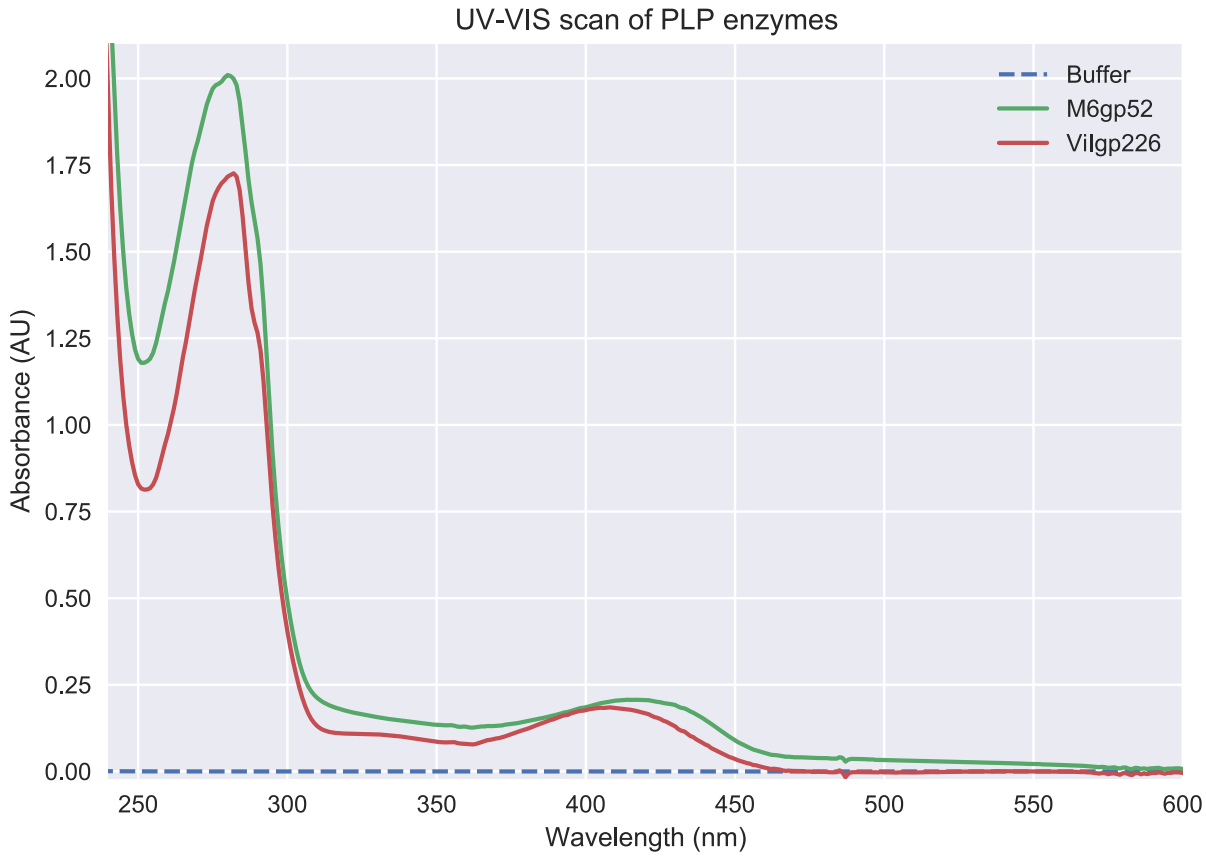


Fig. S18. Visual inspection and UV-Vis Spectra of purified M6 gp52 and ViI gp226 suggest presence of copurified PLP-cofactor.
Chromatographic fractions obtained during purification of M6 gp52 and ViI gp226 have visible yellow hue by naked eye. UV-Vis spectra of the resulting purified proteins show absorbance peaks in the 420-450 nm range characteristic of PLP-dependent enzymes.

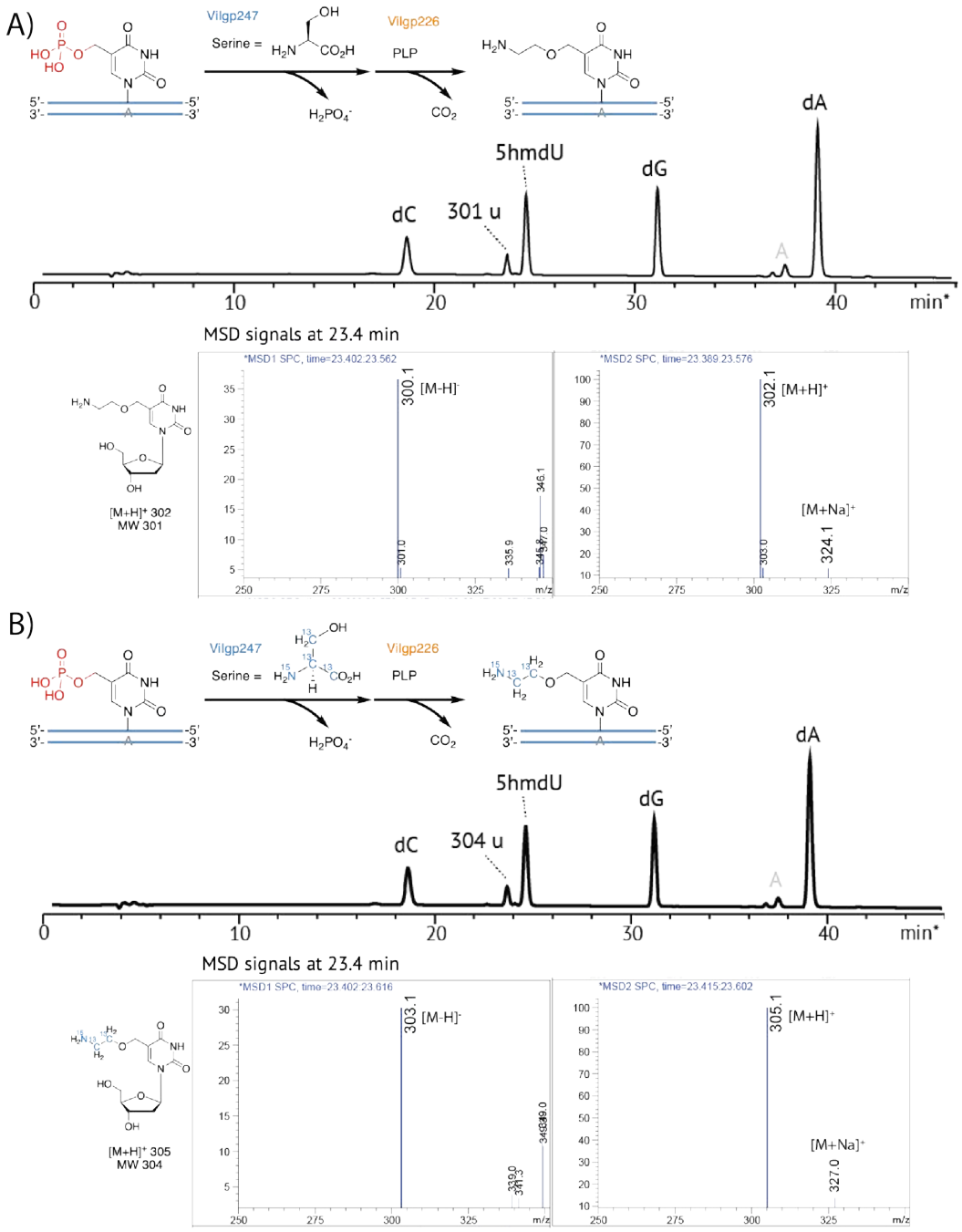


Fig. S19. Decarboxylation of *O*-serinylthymidine formed with stable isotope labeled serine.
The DNA substrate containing 5-PmdU was incubated with purified ViI gp247 and ViI gp226 in the presence of 1 mM L-serine (**A**) or 1 mM stable isotope labeled L-serine (13C3, 99%; 15N, 99%) (Cambridge Isotope Laboratory, Inc., Tewksbury, MA) (**B**) and products were digested to free nucleosides and subjected to LC/MS analysis. LC/MS shows production of nucleoside with nominal mass of 301 u for (**A**) but with stable isotope labeled serine (**B**) yielded a nucleoside with nominal mass of 304 u.

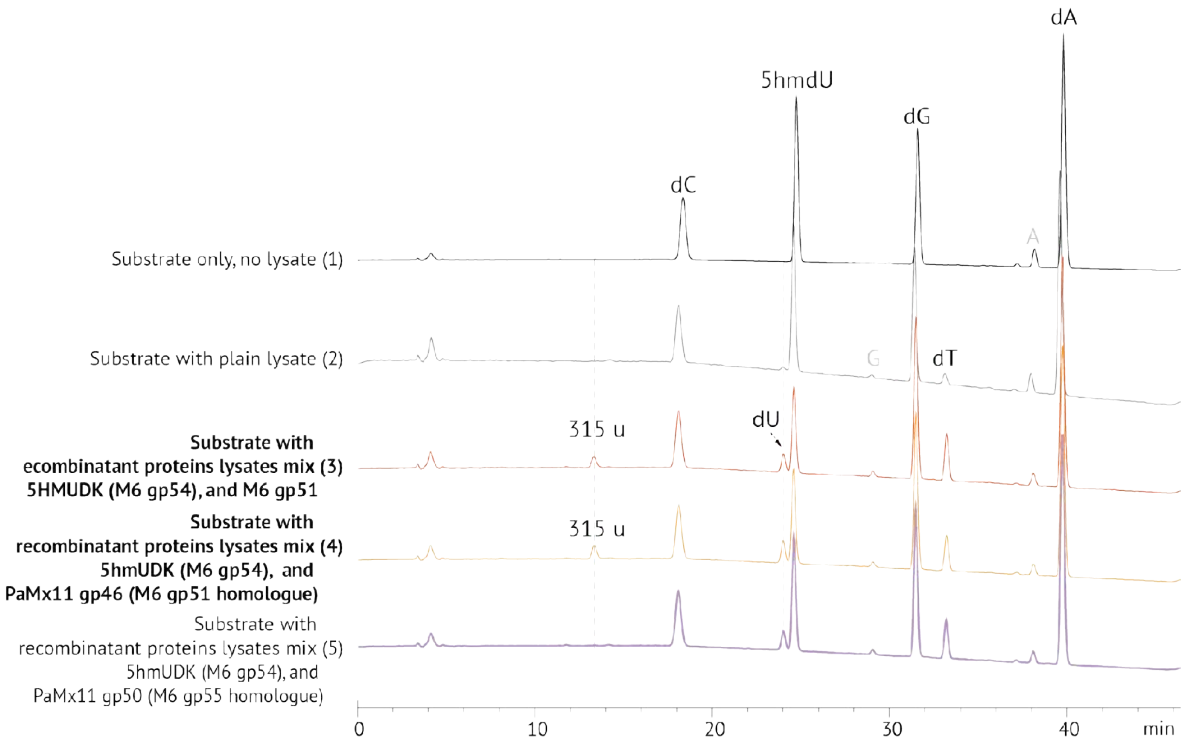
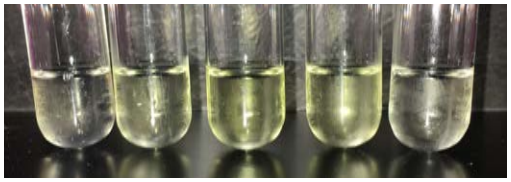
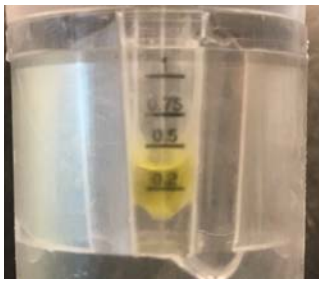


Fig. S20. PaMx11 gp46 produces N^{α}-glyT.
Biotinylated fragments of SP8 genomic DNA prepared as described in Methods were incubated in buffer (trace 1), plain lysate (trace 2), or combinations of lysates derived from cells expressing recombinant M6 gp54 (5-HMUDK) and M6 gp51 (Pfam: aGPT-Pplase2, coined as AA:DNA transferase) (trace3), M6 gp54 and PaMx11 gp46 (an M6 gp51 ortholog) (trace 4), or M6 gp54 and PaMx11 gp50 (an M6 gp55 ortholog). Incubation of substrate DNA with lysates containing M6 gp54 and PaMx11 gp46 was sufficient to produce a nucleoside indistinguishable from the N^{α}-GlyT nucleoside produced by reaction with M6 gp54 and M6 gp51.



PaMx11gp47 FPLC fractions



Ab18gp10 concentrate

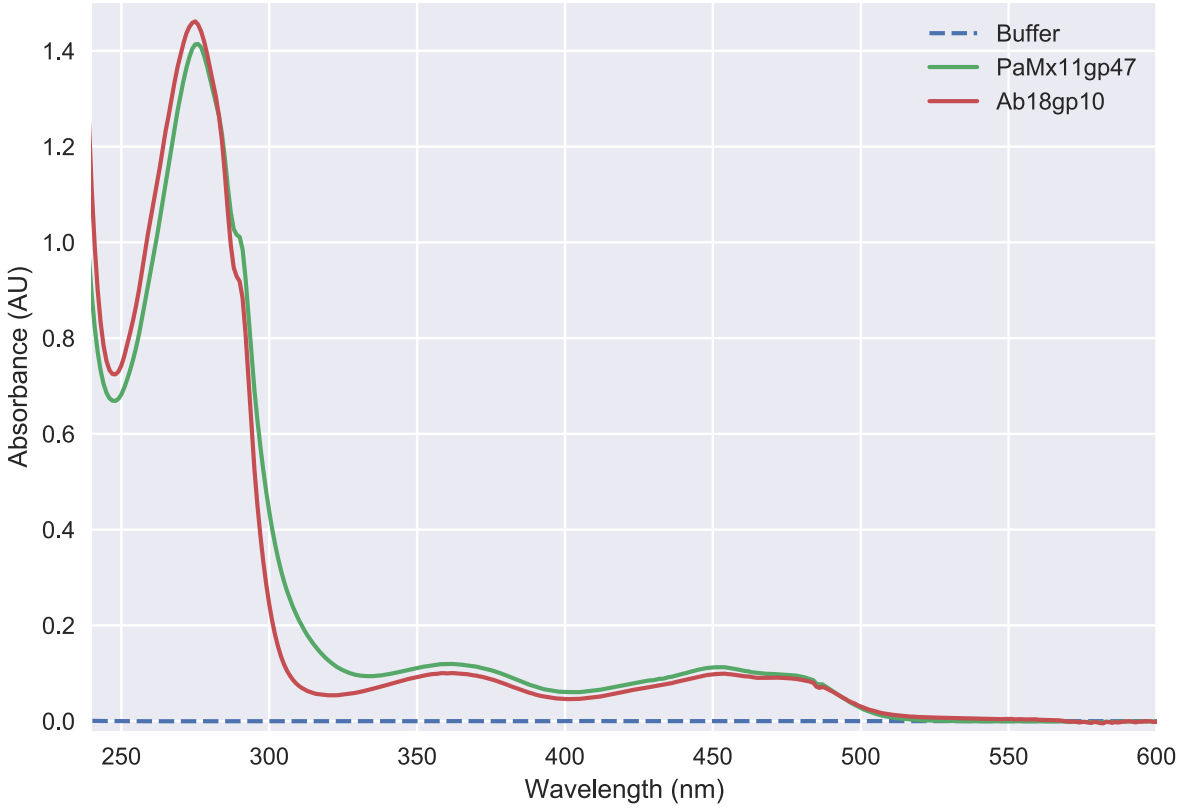


Fig. S21. Visual inspection and UV-Vis Spectra of purified PaMx11 gp47 and Ab18 gp10 suggest they are flavoproteins.

Purified PaMx11 gp47 and Ab18 gp10 have a yellow hue visible to the naked eye. UV-Vis spectra of the resulting purified proteins show two absorbance peaks at ~360 nm and ~455 nm characteristic of proteins containing a flavin cofactor.

Table S1 Plasmids used in this study

Name	Select	Origin	Promoter	Encoding	Prot ID accession	Native/synthetic
pYJL068	Kan	ColE1	T7	M6 gp54	YP_001294562	native gene seq.
pYJL331	Kan	ColE1	T7	PaMx11 gp49	YP_009196302	gene synthesis
pYJL085	Kan	ColE1	T7	ViI gp67	YP_004327432	native gene seq.
pYJL087	Kan	ColE1	T7	ViI gp243	YP_004327565	native gene seq.
pYJL082	Kan	ColE1	T7	ΦW14 gp37	YP_003358891	native gene seq.
pYJL028	Kan	ColE1	T7	SP10 gp186	YP_007003443	native gene seq.
pYJL033	Kan	ColE1	T7	SP10 gp186NT	YP_007003443	native gene seq.
pYJL019	Kan	ColE1	T7	M6 gp51	YP_001294559	native gene seq.
pYJL088	Kan	ColE1	T7	ViI gp247	YP_004327568	native gene seq.
pYJL292	Kan	ColE1	T7	PaMx11 gp46	YP_009196299	gene synthesis
pYJL078	Kan	ColE1	T7	M6 gp53	YP_001294561	native gene seq.
pYJL035	Kan	ColE1	T7	M6 gp52	YP_001294560	native gene seq.
pYJL037	Kan	ColE1	T7	ViI gp226	YP_004327553	native gene seq.
pYJL120	Kan	ColE1	T7	PaMx11 gp47	YP_009196300	gene synthesis
pYJL118	Kan	ColE1	T7	PaMx11 gp48	YP_009196301	gene synthesis
pYJL121	Kan	ColE1	T7	Ab18 gp10	YP_009125113.1	gene synthesis
pYJL121	Kan	ColE1	T7	Ab18 gp11	YP_009125114.1	gene synthesis
pDB1282	Amp		pBAD	*isc* cluster		
pYJL140	Kan	pMB1	Ptac	GFPuv		gene synthesis
pYJL366	Kan	pMB2	Ptac	M6gp51	YP_001294559	native gene seq.
pYJL367	Kan	pMB3	Ptac	ΦW14 gp109	YP_003358963	native gene seq.