Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses

Yan-Jiun Lee1,3, Nan Dai1,3, Shannon E. Walsh3, Stephanie Müller3, Morgan E. Fraser3, Kathryn M. Kauffman3, Chudi Guan4, Ivan R. Corrêa Jr.1,2, and Peter R. Weigele3,2

1Research Department, New England Biolabs, Ipswich, MA 01938; and 2Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Caroline S. Harwood, University of Washington, Seattle, WA, and approved February 23, 2018 (received for review August 23, 2017)

Certain viruses of bacteria (bacteriophages) enzymatically hypermodify their DNA to protect their genetic material from host restriction endonuclease-mediated cleavage. Historically, it has been known that virion DNAs from the Delftia phage ΦW-14 and the Bacillus phage SP10 contain the hypermodified pyrimidines α-potrescyluridine and α-glutamylthymidine, respectively. These bases derive from the modification of 5-hydroxymethyl-2-deoxyuridine (5-hmdU) in newly replicated phage DNA via a pyrophosphorylated intermediate. Like ΦW-14 and SP10, the Pseudomonas phage M6 and the Salmonella phage VII encode kinase homologs predicted to phosphorylate 5-hmdU DNA but have uncharacterized nucleotide content [Iyer et al. (2013) Nucleic Acids Res 41:7635–7655]. We report here the discovery and characterization of two bases, 5-(2-aminoethoxy)methyluridine (5-NeOmdU) and 5-(2-aminoethyl)uridine (5-NedU), in the virion DNA of VII and M6 phages, respectively. Furthermore, we show that recombinant expression of five gene products encoded by phage VII is sufficient to reconstitute the formation of 5-NeOmdU in vitro. These findings point to an unexplored diversity of DNA modifications and the underlying biochemistry of their formation.

DNA modification | bacteriophage | hypermodification | pyrimidine

The tailored viruses of bacteria (bacteriophages or phages) contain the greatest variety of known biologically derived deoxiribonucleotide modifications (1) compared with any group of cellular organisms. In addition to methylated bases (e.g., N6-methyladenosine, N4-methylcytosine, and 5-methylcytosine) commonly found in cellular organisms, bacteriophages can also contain modifications ranging in size from an amino group appended to the N2 position of adenosine (N2-aminoadenine) as found in the Synechocystis phage S-2L (2) to complex hypermodifications, such as glucosyl-phosphoglucuronolactone-dihydropentauracil, found in the Bacillus phage SP-15 (3, 4)—this latter modification is predicted to append a 488-Da side chain to the thymidine nucleotide.

Details of the genetic and biochemical basis for DNA hypermodification have been determined in some bacteriophages, including the Escherichia phage T4, the Delftia phage ΦW-14, and the Bacillus phage SP10 (reviewed in ref. 1), with hypermodifications that are represented in Fig. 1. The phages ΦW-14 and SP10 hypermodify their DNA bases through a combination of pre- and postreplicative mechanisms (schematically represented in Fig. 1) (1). These two phages synthesize 5-hmdUMP from dUMP using a thymidylate synthase homolog—dUMP hydroxymethyltransferase (5); 5-hmdUMP is converted to 5-hmdUTP and incorporated into newly replicated phage DNA by a phage-encoded DNA polymerase. The 5-hmdU bases are then converted in situ to either thymidine or the hypermodified base in a sequence-specific fashion (6, 7). How 5-hmdU is converted to thymidine during phage replication is still unknown. The hypermodification reaction proceeds through a two-step modification of 5-hmdU on the DNA polymer coincident with or immediately after DNA replication (7, 8). A phage-encoded kinase (or pyrophosphotransferase) modifies 5-hmdU at the hydroxymethyl moiety to yield 5-((O-pyrophosphoryl)hydroxymethyl)uridine (5-PPmdU) (7, 8). The pyrophosphorylation of 5-hmdU could activate the 5-methylene carbon for nucleophilic substitution, with the pyrophosphate serving as a leaving group.

Previous work by Aravind and coworkers (9) predicted an enzyme family responsible for the formation of 5-PPmdU from 5-hmdU. In that work, ORFs predicted to encode kinases synthesizing 5-PPmdU were identified in ΦW-14, SP10, the Salmonella phage VII, and the Pseudomonas phage M6 in addition to other bacteriophage genomes. For the purposes of discussion in this work, we will refer to this enzyme as 5-hmdU DNA kinase (5-HMUDK). We reasoned that the bacteriophages M6, VII, and others encoding 5-HMUDK homologs should contain thymidine hypermodifications that would be either identical or similar to those found in ΦW-14 and SP10. In the study reported here, we describe the discovery of two thymidine hypermodifications, 5-(2-aminoethyl)uridine (5-NedU) and 5-(2-aminoethoxy)methyluridine (5-NeOmdU) replacing 40% of thymidine nucleotides in the DNA, respectively. Further, we show that recombinantly expressed five gene products encoded by phage VII is sufficient to reconstitute the formation of 5-NeOmdU in vitro. These findings point to an unexplored diversity of DNA modifications and the underlying biochemistry of their formation.

Significance

Bacterial viruses (bacteriophages) append a variety of molecules, including sugars, amino acids, and polyamines, to the nucleobases of their genomic DNA to circumvent the endonuclease-based defenses of their hosts. These DNA hypermodifications are formed through bacteriophage-encoded biosynthetic pathways, with steps occurring before and after replication of bacteriophage DNA. We report here the discovery of two thymidine hypermodifications: 5-(2-aminoethoxy)methyluridine replacing 40% of thymidine nucleotides in the Salmonella phage VII and 5-(2-aminoethyl)uridine replacing 30% of thymidine in the DNA of the Pseudomonas phage M6. Additionally, we show in vitro reconstitution of 5-(2-aminoethyl)uridine biosynthesis from five recombinantly expressed proteins. These findings reveal an expanded diversity in the types of naturally occurring DNA modifications and their biosynthetic pathways.

Author contributions: I.R.C. and P.R.W. designed research; Y.-J.L., N.D., S.E.W., S.M., M.E.F., C.G., I.R.C., and P.R.W. performed research; Y.-J.L., N.D., K.M.K., I.R.C., and P.R.W. contributed new reagents/analytic tools; Y.-J.L., N.D., K.M.K., I.R.C., and P.R.W. analyzed data; and Y.-J.L., I.R.C., and P.R.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1Y.-J.L. and N.D. contributed equally to this work.

2To whom correspondence may be addressed. Email: correia@neb.com or weigele@neb.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714812115/-/DCSupplemental.

Published online March 19, 2018.
Fig. 1. Examples of phage hypermodified pyrimidines and the generalized DNA thymidine (T) hypermodification pathway of phages SP10 and ΦW-14.

(5-NeOmdU), in virion DNA from M6 and ViI, respectively. The chemical nature of these modifications is identified and confirmed using liquid chromatography-tandem MS (LC-MS/MS), NMR, chemical synthesis, and amine-reactive labeling. Furthermore, we reconstitute the biosynthesis of 5-NeOmdU in vitro using recombinant Escherichia coli extracts expressing candidate DNA hypermodification genes identified here and by Aravind and coworkers (9).

Results
Distinct Groups of Phage Encode a Putative 5-HMUDK. Bacteriophages ΦW-14 and SP10 encode a putative 5-HMUDK, first predicted by Aravind and coworkers (9), to participate in the biosynthesis of hypermodified thymidine derivatives. We hypothesized that other 5-HMUDK–encoding phages would similarly contain thymidine hypermodifications in their DNAs. In an effort to characterize the extent and diversity of the bacteriophage genomes encoding this kinase, homologs of SP10 and ΦW-14 5-HMUDKs were identified using BLAST (10). Fully sequenced bacteriophage genome sequences encoding a putative 5-HMUDK were collected from GenBank (11) (Table 1), and the regions surrounding the kinase gene were visualized as genomic maps. Notably, many genomes showed a high degree of conservation not only in their overall gene content but in their gene organization as well. This synteny is illustrated in Fig. 2 through subgenomic maps detailing the regions surrounding the 5-HMUDK gene of a subset of the phage genomes listed in Table 1.

From this analysis, four distinct groups of phages emerged. At the time that this analysis was performed, the hypermodified phages SP10 and ΦW-14 were essentially unique (i.e., they do not share syntenic regions of nucleotide sequence homology with any other phage encoding a putative 5-HMUDK). Of two other groups of phages encoding 5-HMUDK, one is composed exclusively of phages belonging to the Viunaviridae genus of dsDNA bacteriophages (34, 35). This group is defined by relatedness to the Salmonella phage ViI, a phage originally isolated in 1936 (23) and used by others in phage typing schemes for characterizing Salmonella serotypes (36). The members of the other 5-HMUDK–encoding

Table 1. Selected bacteriophages encoding putative 5-HMUDK

| Phage  | Host                     | Accession no. | Morphotype | Source       | Location         | Ref. |
|--------|--------------------------|---------------|------------|--------------|------------------|-----|
| ΦW-14  | Delftia acidovorans     | NC_013697     | Myoviridae | Sewage       | Vancouver, Canada| 12  |
| SP10   | Bacillus subtilis W23    | AB605730      | Myoviridae | Soil         | Unpublished      | 13  |
| ViI-like |                | NC_002412     | Myoviridae | Wastewater   | Indiana, United States | 14  |
| SJ3    | Salmonella enterica     | NC_022772     | Myoviridae | Wastewater   | Texas, United States | 15  |
| 0507-KN2-1 | Klebsiella pneumoniae    | NC_022343     | Myoviridae | Wastewater   | Taipei City, Taiwan | 16  |
| Marshall | E. coli serovar Enteritidis | NC_019925 | Myoviridae | Crop soils   | Merelbeke, Belgium | 18  |
| SKM-39 | S. enterica              | NC_01910      | Myoviridae | Estuary      | Baltimore, United States | 69  |
| φSH19 | S. typhimurium U288       | NC_019530     | Myoviridae | Sewage       | United Kingdom    | 19  |
| PhaxI  | E. coli O157:H7          | NC_019452     | Myoviridae | Sewage       | Tehran, Iran      | 20  |
| CBA120 | E. coli O157:H7 (NCTC 12900) | NC_016570 | Myoviridae | Cattle feces | Southwest United States | 21  |
| SFP10  | E. coli O157:H7          | NC_016703     | Myoviridae | Slurry       | Seoul Forest, South Korea | 22  |
| ViO1   | S. enterica serovar Typhi | NC_015296     | Myoviridae | Human feces  | Canada            | 23  |
| φSbom-A3 | Shigella boydi           | NC_013693     | Myoviridae | Wastewater   | Guelph, Canada    | 24  |
| SJ2    | Salmonella enteritidis C721 | KJ174317     | Myoviridae | Chicken egg  | Guelph, Canada    | 25  |
| Maynard | S. enterica serovar Typhimurium | KF669654 | Myoviridae | Sewage       | Texas, United States | 26  |
| ECM-4  | E. coli O157:H7          | JK128257      | Myoviridae | Harbor water | Baltimore, United States | 70  |
| 1.255.O | V. splendidus 10N.286.45.F1 | MG592621     | Myoviridae | Seawater     | Massachusetts, United States | 27  |
| CBA6   | E. coli O157:H7 (NCTC 12900) | KM389296     | Myoviridae | Cattle feces | Southwest, United States | 21  |

M6-like

| Phage  | Host                     | Accession no. | Morphotype | Source       | Location         | Ref. |
|--------|--------------------------|---------------|------------|--------------|------------------|-----|
| M6     | P. aeruginosa M6,9       | NC_007809     | Siphoviridae | Unknown      | Unknown          | 28  |
| YuA    | P. aeruginosa PAO1        | NC_010116     | Siphoviridae | Pond water   | Moscow, Russia   | 29  |
| MP1412 | P. aeruginosa PAO1        | NC_018282     | Siphoviridae | Sewage water | South Korea     | 30  |
| PAE1   | P. aeruginosa strain PAO9505 | NC_028980 | Siphoviridae | Sewage water | Bendigo, Australia | 31 |
| LKO4   | P. aeruginosa             | KC758116      | Siphoviridae | Unpublished  | Greece           |     |
| AN14   | Pseudomonas               | KX198613      | Siphoviridae | Freshwater   | Lake Baikal, Siberia |   |
| Ab18   | P. aeruginosa PAO1        | NC_026594     | Siphoviridae | Sewage       | Abidjan, Côte d’Ivoire | 32 |
| PaMx11 | P. aeruginosa PAO1        | NC_028770     | Siphoviridae | Wastewater   | Central Mexico   | 33  |
phage group are all related in sequence and gene organization to the *Pseudomonas* phage M6 (28, 37), a siphovirus (38, 39) also originally isolated for use in phage typing schemes. We chose Salmonella phage ViI and *Pseudomonas* phage M6 (available from the ATCC) (Materials and Methods) as representatives of each of these groups for our studies.

**Phages ViI and M6 Have Noncanonical Bases.** Bacteriophage hypermodified DNAs are often resistant to a range of restriction endonucleases in vitro (40). As seen in Fig. 3, genomic DNA extracted from *Salmonella* phage ViI and *Pseudomonas* phage M6 is resistant to cleavage by certain classical type II restriction enzymes. For example, NcoI does not cleave DNA from phages ViI and M6, although it can cleave canonical DNA from bacteriophage *λ* as well as DNA from the *Bacillus* phage SP8, the latter containing 5-hmdU fully substituting for thymidine (41). Additionally, NcoI does not cleave hypermodified DNAs of *Bacillus* phage SP10 and *Delftia* phage Φ W-14. DNAs of ViI and M6 are resistant to cleavage by EcoRI, although this enzyme is also inhibited by 5-hmdU. Although EcoRI is sensitive to CpG methylation, NcoI is not inhibited by dam, dcm, or cytidine–guanosine sequence dinucleotide (CpG) methylation, further suggesting the presence of noncanonical nucleotides in M6 and ViI.

Interestingly, we find that three RNA guided DNA endonucleases of CRISPR-Cas systems each can cleave genomic DNA from phage ViI to varying degrees. Four of four single guide RNA (sgRNA)-targeted sequences of ViI are specifically cleaved by *Streptococcus pyogenes* Cas9, two of four sites are cleaved by EcoRI, although this enzyme is also inhibited by 5-hmdU. Although EcoRI is sensitive to CpG methylation, NcoI is not inhibited by dam, dcm, or cytidine–guanosine sequence dinucleotide (CpG) methylation, further suggesting the presence of noncanonical nucleotides in M6 and ViI.

Phages ViI and M6 sequences targeted for cleavage contained hypermodified bases or if any bases of the protospacer adjacent motif (PAM) sequences in ViI DNA were hypermodified.

To determine if the restriction enzyme-inhibiting features on ViI and M6 DNAs indeed correspond to noncanonical bases, we performed LC-MS analysis of the nucleoside content obtained from enzymatic digestion of each phage genomic DNA. HPLC traces from each of these samples are shown in Fig. 4. In addition to the peaks corresponding to the canonical nucleosides 2′-deoxyadenosine, 2′-deoxyguanosine, 2′-deoxycytidine, and thymidine, the phages ViI and M6 each have a fifth peak. The retention time
HPLC traces and MS analysis of bacteriophage M6 and V7 nucleosides. The trace in Top was obtained from bacteriophage λ to show the retention of canonical nucleosides. M6 and V7 show a fifth major peak corresponding to the hypermodified base. The protonated molecular ion detected for each hypermodified base is indicated as well as a hypothetical combination of atoms to account for the observed masses, dA, 2'-deoxyadenosine; dG, 2'-deoxyguanosine; dC, 2'-deoxycytidine; dT, thymidine.

(23.2 min) and the mass (301 Da) of the fifth nucleoside peak of V7 are identical to those found for the V7-like phages 1.255.O, CBA6, and CBA120 (SI Appendix, Fig. S1). The 301-Da nucleoside replaces ∼40% of thymidines in V7 and V7-like DNA (SI Appendix, Table S1). The retention time and mass of the fifth nucleoside peak of phage M6 are 15.1 min and 271 Da, respectively. The 271-Da nucleoside replaces ∼30% of thymidines in M6 DNA (SI Appendix, Table S1).

The mass difference between the V7-modified base and thymidine corresponds to two additional carbons, five hydrogens, one nitrogen, and one oxygen. The difference between the M6-modified base and thymidine is one extra carbon, three hydrogens, and one nitrogen. Since the DNA of ΦW-14 and SP-10 contain thymidines hypermodified with putrescyl and glutamyl moieties, respectively, via an N-C linkage (7, 42), we initially hypothesized that V7 and M6 would also feature N-C-bonded thymidine adducts. In the case of phage V7, the adduct could originate from the appendage of a group of formula C3H7NO, such as ethanolamine, to give 5′-(2-hydroxyethyl)amino)methyluridine (5-heNmdU). In the case of phage M6, we envisioned a modified base comprising a methylamine substitution to form 5′-(methylamino)methyluridine (5-mNmdU). The proposed structures of these nucleosides are shown in Fig. 5.

To test whether these hypothetical structures indeed correspond to the native phage modifications, we chemically synthesized 5-mNmdU and 5-heNmdU according to a method adapted from Saavedra (43). Briefly, starting with commercially available 2′-deoxy-5-formyluridine, 5-mNmdU and 5-heNmdU were obtained by reductive alkylation with methyl amine or 2-ethanolamine, respectively, in the presence of NaBH4 (SI Appendix, Fig. S3). These compounds were purified by reverse-phase preparative HPLC and characterized by NMR and high-resolution MS (HRMS).

To our surprise, the synthetic and naturally modified nucleosides of both M6 and V7, other than having identical masses, exhibited markedly different HPLC retention times. Peaks of synthetic 5-mNmdU and 5-heNmdU were detected at 12.0 and 12.8 min, respectively, while those of M6 and V7 were observed at 15.1 and 23.2 min. Per our hypothesis, the hypermodification of phages encoding a 5-HMUDK are derived from further elaboration of the hydroxyl moiety of 5-hmdU in DNA. The fact that our predicted structures did not match the ones obtained from M6 and V7 DNA digests suggested that these phage families may follow a different substitution pattern than the N-C-linked thymidine modifications encountered for phages ΦW-14 and SP-10. Hence, we presumed that the M6 modification could originate from an inverse relationship between the incoming C and N atoms as shown in Fig. 5A. The resulting 5-NedU would feature a terminal primary amine and a C-C linkage between the hypermodification and the thymine base. In the case of the V7, the hypermodification presenting a terminal amine but connected to the thymine via an ether linkage to give 5-NeOmdU seemed the most plausible among several isomeric structures (Fig. 5B). For both 5-NedU (M6) and 5-NeOmdU (V7), the proposed primary amino group could project into the major groove of DNA and interact favorably with the phosphate backbone.

Fragmentation Indicates a C-C Modification in Phage M6. To further infer the atom connectivity of the hypermodified M6 and V7 nucleosides, we then performed a series of collision-induced dissociation (CID) MS/MS experiments. Electrospray ionization (ESI)-MS/MS spectra of 5-hmdU, 5-mNmdU, and the unknown hypermodified nucleoside from phage M6 are shown in SI Appendix, Figs. S4–S6. A generalized fragmentation pathway is shown in Fig. 6. Low-energy CID spectra (SI Appendix, Figs. S4, Upper,

![Fig. 4. HPLC traces and MS analysis of bacteriophage M6 and V7 nucleosides. The trace in Top was obtained from bacteriophage λ to show the retention of canonical nucleosides. M6 and V7 show a fifth major peak corresponding to the hypermodified base. The protonated molecular ion detected for each hypermodified base is indicated as well as a hypothetical combination of atoms to account for the observed masses, dA, 2′-deoxyadenosine; dG, 2′-deoxyguanosine; dC, 2′-deoxycytidine; dT, thymidine.](https://example.com/fig4)

![Fig. 5. Proposed structures of phage M6 (A) and V7 (B) modifications. 5-NedU and 5-NeOmdU are shown to be the actual modifications in this work.](https://example.com/fig5)
Lower Lee et al. ions were detected at 156, 139, 96, and 68, which could be attributed to the cleavage of the glycosidic bond with loss of neutral 2-deoxyribose and yielding protonated nucleobase [B + H]^+ ions were detected at m/z 143, 156, and 156, respectively. Minor peaks resulting from ribose dehydration of [M + H]^+ ions (e.g., at m/z 241, 213, and 205 for 5-hmdU and at m/z 223 and 205 for 5-NmdU) as well as from the cleavage of the glycosidic bond with loss of the protonated 2-deoxyribose ion (m/z 117) and its further dehydration ions (m/z 99 and 81) were also observed. The loss of neutral 2-deoxyribose ion, and, in the same vein, the presence of protonated 2-deoxyribose ion at m/z 117 in the spectra provided conclusive evidence that the M6 modification was present at the nucleobase and not present at the sugar moiety. High-energy CID spectra (SI Appendix, Figs. S4, Lower, S5, row 4, and S6, row 4) indicated losses of water (5-hmdU) or methylamine (5-NmdU) from the protonated nucleobase [B + H]^+ ions to give origin to a common fragment ion at m/z 125. Subsequent losses of HNCO and CO produced ions at m/z 82 and 54, respectively, which are consistent with well-documented fragmentations for uracil derivatives (44-46).

In the case of the M6 modification, MS/MS spectra displayed fragment ions at m/z 156, 139, 96, and 68, which could be attributed similarly to the loss of 2'-deoxyribose to form [B + H]^+ followed by losses of ammonia [B+H-NH_3]^+, HNCO [B+H-NH_2-HNCO]^+, and CO [B+H-NH_2-HNCO-CO]^+, respectively. Comparative analysis suggests the presence of an additional CH_2 group in all main fragments, which is in agreement with the formation of a styrenoid cation (Fig. 6) (47), and supports the occurrence of a C-C-substituted thymine. It is also important to note that the similarity of the UV spectra (SI Appendix, Table S2) of the unknown M6 modification (λ_{max} 266 nm) to that of thymine (λ_{max} 265 nm) rather than isothymine (6-methyluracil, λ_{max} 261 nm) (47) further strengthens our presumption that the uridine substitution is, in fact, at the five position. Finally, since the loss of NH_3 directly from the uracil ring has been reported to be prohibited in 5-hmdU and related structures (46), the facile loss of ammonia observed for the unknown M6 nucleoside substantiated the presence of a terminal primary amine. All of the results taken together led us to conclude that the hypermodification found in phage M6 is 5-NedU.

Next, we compared the CID mass spectra of the unknown VI modification with those of the synthetic 5-heNmdU. For both nucleosides, low-energy CID (SI Appendix, Figs. S7, row 1 and S8, row 1) showed the formation of two abundant ions at m/z 302 and 186, which correspond to [M + H]^+ and [B + H]^+, respectively. Additional fragmentation of protonated 5-heNmdU resulted in loss of ethanolamine to give rise to the fragment ion [B+H-HOCH_2CH_2NH_2]^+ at m/z 125. The peak corresponding to the protonated form of ethanolamine was also found at m/z 62. As was the case with 5-hmdU and 5-NmdU, loss of HNCO via retro Diels–Alder followed by successive loss of CO was the major pathway of dissociation from the common ion at m/z 125, and it produced the fragment ions m/z 82 and 54 (Fig. 6 and SI Appendix, Figs. S7 and S8). In addition to [M + H]^+ and [B + H]^+, the fragmentation of the protonated unknown VI nucleoside generated a series of fragment ions at m/z 125, 82, and 54 that paralleled those of 5-heNmdU. These results led us to surmise that 5-heNmdU and the unknown VI modification shared the same uridine backbone but had an inverse arrangement of the heteroatom substituents at the five position, such as the one found in 5-NeOmdU. Since we could not obtain any further conclusive structural information from these fragmentation studies, we decided to isolate the unknown VI nucleoside and perform NMR analysis in attempt to unambiguously elucidate its structure.

NMR Reveals an Ethoxy Linkage in VI Modification. To obtain sufficient amounts of the unknown nucleoside for NMR analysis, 18 L of early log-phase host Salmonella typhi was infected with freshly prepared plaques of VI at about one plaque per 250 mL. The resultant phage particles were purified, and their genomic
DNA was extracted as described in Materials and Methods. Approximately 0.4 g of DNA was obtained and subsequently processed in four 0.1-g batches. DNA was extensively dialyzed against water, and then, 0.1 g of dialyzed genomic DNA was digested to free nucleosides. After HPLC purification using a gradient of ammonium acetate buffer, pH 4.5, and methanol and pooling the desired fractions, ~9 mg of the unknown nucleoside was isolated from the total starting material. This nucleoside was subjected to 13C NMR analyses. We compared the NMR spectra of the unknown VII nucleoside with those of the synthetic 5-heNmU and 5-mNmU as well as with those of the known compounds 5-hmdU and 5-(2-cyanoethoxy)methyluridine (5-ceOmdU) (48).

As depicted in Table 2, there is a strong correlation between the NMR chemical shifts of 5-ceOmdU and the proposed VII nucleoside 5-NeOmdU and between the chemical shifts of 5-heNmU and 5-mNmU. The presence of an ethoxy modification (X = O) off the 5-methyl group of the thymine ring in 5-ceOmdU and 5-NeOmdU was evidenced by the downfield shift of the methylene hydrogens immediately adjacent to the oxygen atom [4.17 and 4.12 ppm for -OCH2-dU and 3.58 and 3.43 ppm for -CH2-O-mdU, respectively] relative to those of the compounds carrying an ethylamine modification (X = N)]. The cognate signals of 5-heNmU and 5-mNmU were detected at 3.38 and 3.35 ppm (-NCH2-dU) and 2.55 and 2.26 ppm (-CH2-N-mdU), respectively (Table 2) (48). A small but consistent downfield shift was also observed for the thymine H6 of 5-ceOmdU (7.94 ppm) or 5-NeOmdU (8.00 ppm) relative to that of 5-heNmU (7.79 ppm) or 5-mNmU (7.81 ppm). All other 1H NMR signals from these four compounds were essentially identical (SI Appendix, Fig. S9).

A very similar outcome was revealed by 13C NMR analysis (Table 2). The ethoxy modification in 5-ceOmdU and 5-NeOmdU caused an expected downfield shift of signals in the vicinity of the oxygen atom [64.4 and 64.7 ppm (OCH2-dU) and 64.6 and 70.6 ppm (-CH2O-mdU)] relative to the corresponding signals in the ethylamine-modified 5-heNmU and 5-mNmU, 44.9 and 46.3 ppm (-NCH2-dU) and 50.5 and 34.5 ppm (-CH2-N-mdU), respectively (Table 2). Again, a smaller but significant downfield shift was obtained for the thymine C6 of 5-ceOmdU (139.4 ppm) or 5-NeOmdU (138.7 ppm) relative to that of 5-heNmU (137.6 ppm) or 5-mNmU (138.1 ppm). No significant difference was observed for thymine C5 or any other 13C NMR signal from these four compounds (SI Appendix, Fig. S10). Based on these results, we, therefore, confirmed the identity of the unknown VII nucleoside as 5-NeOmdU. It is worth noting that the reported 1H and 13C NMR signals for 5-hmdU [in particular, HOCH2-dU (4.12 ppm) and HOCH2-dU (62.8 ppm)] also corroborated well with the presence of an oxygen immediately after the methyl group of the thymine nucleobase. In support of the determination of the VII modification as 5-NeOmdU (which should contain a terminal primary amino group), we find that VII genomic DNA can be specifically labeled with the amine reactive dye 3-(4-carboxybenzoyl) quinoline-2-carboxaldehyde (SI Appendix, Fig. S11).

Phage Genes Unique to Thymidine Hypermodifying Phages. The finding that, like SP10 and ΦW-14, the VII- and M6-like bacteriophages hypermodify their thymidines offers the possibility to narrow the search space for genes encoding thymidine hypermodification functions according to the following logic: only genes found within some or all of the hypermodifying phage genomes but not in genomes of phages containing nonhypermodified thymidines (λ, T4, and Bacillus phage SPO1) should be directly involved in the formation of hypermodified thymidines. Through pairwise BLAST comparisons and manual curation of lists of annotated reading frames from the thymidine hypermodifying phages M6, SP10, VII, and ΦW-14, we found four classes of genes that appear to be unique to the thymidine hypermodifying phages (Table 3). These genes are predicted to encode the following products: a putative 5-HMUDK, a pyridoxal-phosphate (PLP)-dependent enzyme, a predicted radical S-adenosyl-L-methionine (SAM) enzyme, and an enzyme predicted to have a Helix-hairpin-Helix DNA base glycosylase fold by Phyre2 (49). These findings agree with observations and predictions previously made by Aravind and coworkers (9). The DNA base glycosylase-like gene product has been named by Aravind and coworkers (9) an “alpha-glutamyl/putrescyl thymyidine pyrophosphorylase” (aGPT-PPlase) after its predicted activity toward 5-PPmdU during DNA hypermodification. Within the genomes that we examined (SP10, ΦW-14, VII, M6, λ, T4, and Bacillus phage SPO1), the aGPT-PPlase appeared to be restricted to the hypermodifying phages; however, homologs are also found within cellular organisms and viruses (9).

The locations of these putative hypermodification genes within their respective genomes are schematically shown as colored boxes in Fig. 2 and SI Appendix, Fig. S12. The M6-like family of phages shows the tightest clustering of genes predicted to function

### Table 2. Selected 1H and 13C NMR chemical shift data of 5-substituted uridines (in parts per million)

| Compound | H6 | -X-CH2-dU | -CH2-X-mdU | C5 | C6 | -X-CH2-dU | -CH2-X-mdU |
|----------|----|------------|-------------|----|----|------------|-------------|
| 5-hmdU*  | 7.77 | 4.12 | — | 114.3 | 136.8 | 62.8 | — |
| 5-ceOmdU* | 7.94 | 4.17 | 3.58 | 110.0 | 139.4 | 64.4 | 64.6 |
| 5-NeOmdU† | 8.00 | 4.12 | 3.43 | 110.4 | 138.7 | 64.7 | 70.6 |
| 5-heNmU | 7.79 | 3.38 | 2.55 | 111.5 | 137.6 | 44.9 | 50.5 |
| 5-mNmU | 7.81 | 3.35 | 2.26 | 110.4 | 138.1 | 46.3 | 34.5 |

*From Hansen et al. (48).

†Proposed structure for the unknown VII nucleoside (this work). DMSO-d6 was used as internal standard.
in hypermodification. Of this grouping, the *Pseudomonas* phages PAO1_Ab18 and PaMx11 differ in having predicted flavin adenine dinucleotide (FAD) and acetyltransferase genes instead of PLP-dependent and radical SAM enzyme-encoding genes at the orthologous locus (Fig. 2). Notably, the predicted hypermodification biosynthesis genes are often found near genes for dUMP hydroxymethyl transferase and phage-encoded DNA polymerase (indicated in yellow and green, respectively, in Fig. 2 and SI Appendix, Fig. S12).

**Reconstitution of Base Hypermodification from Recombinant E. coli Lysates.** While working toward the in vitro reconstitution thymidine hypermodifying, we first attempted to clone an apparent gene cluster of phage M6 containing five genes predicted to function in base hypermodification (Fig. 2 and column 5 in Table 3). However, we were unable to clone the entire operon as a single construct, possibly due to toxicity of one or more gene products encoded within the operon. However, we successfully cloned VII genes encoding gene product 67 (gp67), gp160, gp226, gp243, and gp247 (Table 3) into plasmids for expression in *E. coli* under control of T7-inducible promoters. After confirming protein expression, individual lysates from cultures expressing each of these proteins were mixed together and incubated with biotinylated restriction fragments of genomic DNA from the *Bacillus* phage SP8, which contains 5-hmdU fully substituting for thymidine. After incubation, the biotinylated DNA was recovered using streptavidin-coated magnetic beads, substituted for thymidine. After incubation, the biotinylated substrate DNA treated with a mixture of the five lysates showed an identical retention time and mass for an equivocal 5-hmdU. Given that 5-hmdU DNA + recombinant lysates

Based on our findings, we predict here that *Pseudomonas* phages YuA, MP1412, PAE1, LKO4, and AN14 similarly contain 5-NedU. It should be noted that, within the M6 family of *Pseudomonas* phages discussed here, PAO1_Ab18 and PaMx11 have a predicted FAD-dependent oxidoreductase and an acetyltransferase gene replacing the predicted PLP-dependent enzyme and radical SAM enzyme-encoding genes found at the same location within the genomes of the other M6-like family members listed in Table 1 and diagrammed in Fig. 2. We hypothesize that these two phages contain hypermodifications that differ from the 5-NedU of phage M6.

Each of the four phage groups described here encodes a putative 5-HMUDK implicated in the hypermodification process (9). Despite the lack of similarity between SP10 and ΦW-14 gene arrangements, both of their hypermodifications are composed of thymidines modified via an N-C linkage of the 5-methyl group (7, 42). Although the attached glutamyl and putrescyl moieties are different in their structure, the phages nonetheless use the same intermediate in the formation of the hypermodification (i.e., a pyrophosphorylated 5-hmdU) (7, 8). The occurrence of a 5-PPmU intermediate would suggest a mechanism in which the hypermodifications are formed via displacement at a pyrophosphate-activated 5-hmdU. Given that 5-NedOdu contains an ether bond, it is tempting to imagine cytoplasmic metabolites bearing free amine or hydroxyl groups as donors for nucleophilic substitution at pyrophosphorylated 5-hmdU for an even greater variety of possible modifications. However, phage hypermodifications are not limited to N-C or O-C bonds. The 5-NedU modification of M6 contains a C-C bond. This hypermodification is more difficult to rationalize via a simple nucleophilic substitution. It remains to be determined if 5-NedU forms through a charged or radical intermediate, such as might be suggested by the presence of a radical SAM enzyme-encoding gene in the M6-like phages. Collectively, the modifications observed to date in 5-HMUDK-encoding phages are now known to encompass N-C, O-C, and C-C hypermodified pyrimidines.

**Discussion**

The hypermodified thymidines described in this work imply an even wider variety of nucleotide modifications synthesized by bacteriophages and cells. Bacteriophage VII and its closely related phages CBA6, CBA120, and 1.255.O contain a C-O bond linking the aminoethanol moiety to the 5-methyl group of thymidine (5-NeOmdU). The high degree of sequence similarity and synteny of these phages to other members of the VII-like group listed in Table 1 and illustrated in Fig. 2 indicate that they likely contain 5-NeOmdU as well. The M6-modified base contains an aminomethyl group attached via a C-C linkage to the 5-methyl group of thymidine (5-NedU).

![Table 3. Genes unique to thymidine hypermodifying bacteriophages](image)

| Predicted | ΦW-14 | SP10 | VII | M6 |
|-----------|-------|------|-----|----|
| Kinase     | gp37  | 863983 | gp186N* | 14007383 | gp67 | 10351655 | gp54 | 5237077 |
| Kinase2    | N/d   | gp218 | 14007415 | gp243 | 10351581 | N/d |
| PLP enzyme | N/d   | gp226 | 10351570 | gp52 | 5237094 |
| Radical SAM | N/d   | gp185 | 14007382 | N/d | gp53 | 5237052 |
| aGPT-Piase1 | gp72 | 8684019 | gp186C* | 14007383 | gp160 | 10351543 | gp51 | 5237089 |
| aGPT-Piase2 | gp109 | 8684057 | N/d | gp247 | 10351584 | gp55 | 5237058 |

A National Center for Biotechnology Information Gene ID number is indicated in parentheses after the gene product number. N/d, none detected.

*Domain fusion.
We show here that five recombinantly expressed genes (VII gp67, gp160, gp226, gp243, and gp247) are sufficient to reconstitute 5-NeOmU synthesis in vitro using a crude lysate system. It is not yet known which of these five are necessary. Of the five genes, two pairs are paralogs (gp67 and gp243 are predicted 5-HMUDKs, and gp160 and gp247 are predicted to have a DNA glycosylase fold) (Table 3). This redundancy suggests that three classes of enzymatic activity are minimally necessary for thymidine hypermodification. The metabolic origin of the additional atoms constituting the VII hypermodifying moiety has not yet been determined, although given that we can reconstitute 5-NeOmU in lysates, it is likely a metabolite normally present in cytoplasm of E. coli. For ΦW-14, the putrescine moiety has been shown to derive from ornithine (50). Glutamyl thymidine is likely derived from glutamate, which is the most abundant solute in bacterial cells (51). Less obvious is how 5-NeOmU and 5-NeDU are derived. To the best of our knowledge, ethanolamine and methylethylamine are not present at above-micromolar levels in bacterial cells (51) and therefore, are unlikely to be the direct precursors to the hypermodifying moieties. Although PLP-dependent enzymes and radical SAM enzymes encompass very diverse functionalities, they are often found in amino acid metabolic pathways. The known modifications of ΦW-14 and SP10 utilize components of amino acid metabolism. Together with enzymes having similarity to DNA glycosylases, perhaps these enzymes utilize amino acids as precursors. For example, an ethanolamine group could be derived from the decarboxylation of serine, and a methylethylamine could be derived from the decarboxylation of glycine. These hypotheses await experimental testing.

An obvious possible function of thymidine hypermodification is to counteract bacterial immunity systems. For example, bacteriophage T4 contains glucosylated hydroxymethylcytidines that allow T4 DNA to resist cleavage by a variety of restriction endonucleases both in vitro (40) and in vivo (52). Presumably, the presence of glucose in the major groove of DNA sterically blocks either the recognition of DNA by endonucleases or the mechanism of cleavage. As seen in Fig. 3, thymidine hypermodifications in VII and M6 also inhibit cleavage by certain restriction endonucleases. Resistance to host restriction would be expected to confer a fitness advantage to phages encoding thymidine hypermodification machinery, and the global distribution of phages encoding thymidine hypermodifications (Table 1) suggests that this is the case.

Less clear is the effect that the thymidine hypermodification might have on CRISPR-based adaptive immunity systems. We observed cleavage of VII DNA by each of three Cas RNA guided DNA endonucleases in vitro (SI Appendix, Fig. S1). Engineered Cas9 systems in E. coli have been shown to restrict glucosylated hydroxymethylcytidine of phage T4 (53) in vivo, but under slightly different experimental conditions, T4 could bypass Cas9 cleavage (54). Thymidine hypermodifications could potentially interfere with at least three aspects of Cas endonuclease function: contacts between Cas endonuclease and DNA backbone during R-loop formation, pairing between the sgRNA and DNA, and Cas-endonuclease–DNA interactions at the PAM site. Although the exact sequence motifs of thymidine hypermodifications are not yet known, the length of the target sequences (~20 nt) and the fact that at least 8 of 12 sites targeted in our experiment are nonetheless cleaved (SI Appendix, Fig. S1) suggest that hypermodification does not interfere with Cas endonuclease binding and R-loop formation. However, it is not known if a hypermodification occurs within any of the PAM sequences of the DNA sites that we targeted. Therefore, we hypothesize that, if any inhibition of Cas endonuclease function by thymidine hypermodification was to occur, it would be mediated by steric interference between modified PAM sequences and the Cas endonuclease.

Bacteria are in a constant arms race with their viral predators (55), and the existence of the 5-NeOmU and 5-NeDU modifications points to the possibility of host mechanisms targeting these modifications. Certain strains of E. coli, such as CT596 and UT189, encode a hypermodification-dependent restriction endonuclease called GmrSD (56, 57). GmrSD specifically recognizes and cleaves glucosylated 5-hmdC of bacteriophage T4 but leaves canonical cytidine DNA intact. GmrSD likely contains distinct cleavage and reader domains that specifically recognize the modified base and hydrolyze the phosphodiester backbone, respectively (56). Some bacteria encode up to five GmrSD homologs, suggesting that paralogous members of this family may recognize different types of hypermodified pyrimidines (56), including the kinds described in this work.

In addition to potential roles in virus-host interactions, the thymidine hypermodifications may have structural effects on DNA stability of both the free polymer and that of DNA tightly packaged in the phage capsid. Putrescyl thymidine, 5-NeOmU, and 5-NeDU each terminate in a solvent-exposed positively charged amino group likely located in the major groove. This positive charge might serve the function of countering the negative charge of the phosphodiester backbone to overcome repulsion and facilitate packaging of DNA in the virion capsid, a stabilizing function ordinarily provided by divalent metal cations, such as Mg2+ and Ca2+. Indeed, although the genome of ΦW-14 is ~7% shorter than T4, it is packaged into a 28% smaller volume (58). Additionally, the Tm of DNA from ΦW-14 is ~6°C higher than expected based on the presence of guanosine and cytidine (GC) content (42). Thymidine modifications of VII and M6 might have similar physicochemical effects on their capsid-packaged genomic DNAs, potentially impacting phage viability under different environmental conditions.

Among the set of genes identified in Table 3, the putative kinase (5-HMUDK) and the DNA glycosylase-like genes are the only genes absolutely shared between SP10, ΦW-14, VII, and M6. Homologs to these DNA glycosylase-like genes are found not only among bacteriophages but in eubacterial genomes as well (SI Appendix, Fig. S12). Recently, through a combination of bioinformatic analyses and chemical methods, a diverse class of naturally occurring queuosine-like 2'-deoxy-7-deazaguanosine–derived DNA modifications was discovered in the DNA of the E. coli phage 9g as well as in pathogenic enteric bacteria (59). These guanosine hypermodifications were shown to inhibit cleavage of phage 9g DNA by restriction endonucleases in vitro (60), and a genomic island including 2'-deoxy-7-deazaguanosine biosynthetic genes was shown to be a barrier to transformation of Salmonella by unmodified DNA in vivo (59), showing that viruses and cells can co-opt the same complex DNA modifications, each for their own survival. We observe homologs of VII gp160 and gp247 distributed among viruses, basidiomycete fungi, and diverse eubacteria, including soil-dwelling actinomycetes and medically relevant enteric gamma-proteobacteria (SI Appendix, Fig. S13), suggesting a much larger, unexplored diversity of nucleotide modifications and biological roles. The approaches described here show that thymidine hypermodifications are experimentally tractable, and additional investigation of their biosynthesis may reveal nucleic acid biochemistries, uncover biological functions, and lead to innovative experimental tools for nucleic acid manipulation.

Materials and Methods

*Chemicals, Media, Enzymes, and Microbes.*** Unless otherwise specified, all chemicals were obtained from Sigma-Aldrich and used without additional purification. Media and molecular components were sourced from Beckton Dickinson-Difco. All enzymes used in this study were obtained from New England Biolabs (NEB). The following strains were purchased from the American Type Culture Collection (ATCC): bacteriophage VII (ATCC 27870-B1), the VII host Salmonella enterica ss. enterica serovar Typhi (ATCC 27870), bacteriophage M6 (ATCC BAA-9-B1), and the M6 host Pseudomonas aeruginosa (ATCC BAA-318). VII phage gp226 was referred to as 1.255.O and its host, Vibrio splendidus 1ON.286.45.F1, were isolated from littoral zone seawater at Canoe Cove (Nahant, MA) on October 13, 2010 (27). Bacteriophages CBA6 and CBA120 as well as their host E. coli
Seroype O157:H7 National Collection of Type Cultures (NCTC) strain 12900 (also known as ATCC 700728) were provided by Elizabeth Kushner, Evergreen State College, Olympia, WA.

Growth and Purification of Bacteriophages. *Escherichia* and *Salmonella* were grown in Rich Broth composed of 10 g tryptone, 5 g NaCl, and 5 g yeast extract per 1 L. *Pseudomonas* was grown on trypticase soy broth (61). *Vibrio* was cultured in Zobell Marine Broth 2216 (61). For solid medium, bactoagar was added at 1.5% to the liquid medium base before autoclaving. Large-scale bacteriophage lysates were prepared as described previously (62), purified by CSE density ultracentrifugation, and purified according to previous methods (63). Additional details of phase purification and preparation of phage genomic DNA are given in SI Appendix.

Isolation and Characterization of Bacteriophage-Modified Nucleosides.

**Phage VII.** Dialyzed phage VII DNA (100 mg) was digested at 0.1 mg/mL final concentration to nucleosides at 37 °C with the Nucleoside Digestion Mix in 1 mL of 1× Nucleoside Digestion Mix Buffer (NEB). After passage through a 3-kDa molecular mass cutoff Vivaspin 20 spin column (GE Healthcare Bio-Sciences) to remove proteins, the flow through was collected and passed through a 0.22-μm membrane (Corning) to remove any particles before HPLC purification. The filtrate was lyophilized to dryness and then redissolved in water. The phage VII hypermodified nucleoside was purified on a preparative HPLC (Waters Atlantis WHRP 19×150 mm). A semipreparative column (19 × 250 mm, 5 μm) as described in the SI Appendix. A white powder was obtained after lyophilization. HRMS m/z calculated for C$_{11}$H$_{14}$N$_{2}$O$_{2}$ (M + H$^+$) = 231.1104, 232.1166; observed, 231.1098, 232.1163, respectively. Nucleosides at 0.1 μg were analyzed by LC/MS where 5 μL of the Nucleoside Digestion Mix in 5 μL of DMSO (NEB) was added to the bead magnet, and the supernatant was subjected to LC/MS analysis.

**Chemical Synthesis of Nucleoside Standards.** 5-mMdmU and 5-hemdU were synthesized following a method adapted from Saaavedra (43). Additional details of their synthesis are given in SI Appendix.

**LC-MS/MS and Fragmentation Analysis.** LC-MS/MS was performed on an Agilent 1290 ultra-high performance liquid chromatograph equipped with a 5-mm × 150 mm, 1.8-μm particle size) at a flow rate 0.2 mL/min with a binary gradient from 1% solvent A (10 mM ammonium formate, pH 4.4) to 100% B (methanol). Absorbance was monitored at 260 nm. MS/MS fragmentation spectra were recorded in the positive product-ion mode with the following parameters: gas temperature 200 °C, gas flow 14 L/min, nebulizer 45 psi, sheath gas temperature 350 °C, sheath gas flow 11 L/min, capillary voltage 2 kV, nozzle voltage 1.5 kV, fragmentor voltage 380 V, and collision energy 5–6 V.

**Computational Methods.** Homology searches were performed using web-based local application-based implementations of BLAST and position-specific iterated BLAST (10, 64). The multiple sequence alignments were generated using MAFFT (65) with the JTT100 scoring matrix (66). Phylogenetic trees were generated using PHYML with the LG substitution model (67). Most bioinformatic operations described were performed within the Geneious software package (68). Protein function and fold were predicted using the Phyre2 structure prediction server (49).

**Preparation and Assay of Hypermodification Reactions.** SP8 genomic DNA was digested to completion with restriction endonuclease HpyCH4IV according to the manufacturer’s guidelines. After heat inactivation at 80 °C for 20 min, the digest was purified using a Monarch nucleic acid purification kit (NEB). Fragmented SP8 genomic DNA at a final concentration of 200 ng/μL was biotinylated by incubating 1 U/μL of Taq DNA polymerase containing 50 μM biotin-16-DCtP at 68 °C for 3 h in 1× ThermoPol buffer (NEB). After cooling, reaction to room temperature and biotinylated DNA was purified using a QiAquick nucleotide removal kit (Qiagen).

Codon-optimized phage VII candidate hypermodification genes were synthesized and cloned into pET28b vectors by Genscript’s custom gene synthesis services (Genscript). These plasmids were used to transform *E. coli* strain NEB T7 Express. Proteins were expressed in 5 mL liquid medium incubated 1:50 overnight starting culture. Cultures were incubated at 37 °C with agitation until they reached an OD$_{600}$ = 0.6 and cooled to room temperature before isolation. The biotinylated genomic DNA was fragmented and then redissolved in lysis buffer (10 mM Tris·HCl, pH 8, 100 μM NaCl, 10 mM KCl, 1 mM PMSF) and then collected. Cells were harvested by centrifugation and suspended in lysis buffer (10 mM Tris·HCl, pH 7.6, 200 mM NaCl, 2 mM EDTA, 15% PEG 6000, 20 mM EDTA). Streptavidin-coated magnetic beads (NEB) were used to capture and separate the biotinylated SP8 DNA from endogenous nucleic acids. The beads (30 μL slurry) were conditioned by washing in 20 mM Tris·HCl, pH 7.5, 0.3 M NaCl, and 1 mM EDTA. Beads were added to the DNA lysate quenching buffer and incubated at ambient temperature for 10 min. The DNA-bound beads were captured by a magnet, and the supernatant was discarded. After two wash cycles, the DNA-bound beads were suspended in 43 μL of deionized water. To release the nucleosides for analysis, 5 μL of Nucleoside Digestion Mix Buffer (50 mM NaOAc, pH 5.4, 1 mM ZnCl$_2$) and 2 μL of Nucleoside Digestion Mix were added to the bead suspension and incubated at 37 °C for 2 h. The beads were removed with a magnet, and the supernatant was subjected to LC/MS analysis.

**Acknowledgments.** We thank Prof. Betty Kutter (Evergreen State College) for sharing microbial strains. Dr. Lana Saleh (NEB) provided a critical reading of the manuscript throughout its preparation. Y.-J.L., N.D., S.E.W., S.M., M.E.F., C.G., I.R.C., and P.R.W. acknowledge guidance from Christopher Noren, William Jack, and Richard J. Roberts and financial support from NEB.

1. Weigle P, Raleigh EA (2016) Biosynthesis and function of modified bases in bacteria and archaea. *Annu Rev Genet* 50:1265–1287.
2. Kirno MS, Khudyakov YI, Alexandrushka NI, Vanyushin BF (1977) 2′-aminoadenine is an adenine substituting base for a 5′-L-cyanoapadine. *Nature* 270:369–370.
3. Walker MS, Mandel M (1978) Biosynthesis of 5-(4′-desoxythymidylyl) uracil as a nucleotide triphosphate in bacteriophage SP15-infected Bacillus subtilis. *J Biol Chem* 253:500–509.
4. Ehrlich M, Ehrlich KC (1981) A novel, highly modified, bacteriophage DNA in which deoxyribo nucleic acid of bacteriophage phi W-14. *Biochemistry* 20:3586–3591.
5. Iyer LM, Zhang D, Burroughs AM, Aravind L (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res* 41:7635–7655.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
7. Benson DA, et al. (2013) GenBank. *Nucleic Acids Res* 41:D36–D42.
8. Kropinski AMB (1965) Isolation and characterization of a bacteriophage active against pseudomonas acidovorans. *Milb. Tidskr. vet. Med.* 11:149–151.
9. Wall SK, Zhang J, Rostagno MH, Ebner PD (2013) *Nature* 497:198–202.
10. Venter J (1990) *Science* 245:397–398.
11. Larson AL, Wood TL, Chamakura KR, Puterwatt EF (2013) Complete genome of Salmonella enterica serovar enteriditis myophagia Marshall. *Genome Announce* 1: e00867-13.
33. Sepúlveda-Robles O, Kameyama L, Guarneros G (2012) High diversity and novel typing phages of the Lindberg set. J Bacteriol 194(23):6203-6212.

19. Hooton SPT, Atterbury RJ, Connerton IF (2011) Application of a bacteriophage in feedlot cattle feces. Lett Appl Microbiol 53:203-209.

16. Hsu C-R, Lin T-L, Pan Y-J, Hsieh P-F, Wang J-T (2013) Isolation of a bacteriophage that efficiently infects Salmonella enterica serovar Typhimurium. J Virol 87(8):4593-4597.

23. Craigie J, Brandon KF (1936) Bacteriophage specific for the O antigens of Salmonella typhimurium. J Bacteriol 31(1):1-10.

38. Bradley DE, Pitt TL (1974) Pilus-dependence of four Pseudomonas aeruginosa bacteriophages. J Gen Virol 25(2):181-188.

37. Kwan T, Liu J, Dubow M, Gros P, Pelletier J (2006) Comparative genomic analysis of Pseudomonas aeruginosa bacteriophages. J Mol Biol 359(3):619-642.

36. Felix A, Callow BR (1943) Typing of paratyphoid B bacilli by Vi bacteriophage. BMJ 2:127-130.

15. Kwan T, Liu J, Dubow M, Gros P, Pelletier J (2006) Comparative genomic analysis of 18 Pseudomonas aeruginosa bacteriophages. J Bacteriol 188(21):6519-6528.

34. Saavedra JE (1985) Reductive alkylation of beta-alcanolamines with carbonyl compounds and sodium borohydride. J Org Chem 50:2271-2277.

43. Saavedra JE (1994) Collision-induced dissociation of uracil and its derivatives. J Am Soc Mass Spectrom 5:339-349.

Dudley E (2010) Analysis of free phage DNA by mass spectrometry. Mass Spectrometry of Nucleic Acids and Nucleic Acids, Ed Banouh JD (Limbach, Pittsburgh), pp 435-444.

Cao H, Wang Y (2006) Collisionally activated dissociation of protonated 2-deoxy-cytidine, 2′-deoxyuridine, and their oxidatively damaged derivatives. J Am Soc Mass Spectrom 17:1335-1341.

59. Thiaville JJ, et al. (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. J Virol 90:8451-8464.

56. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

58. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.

67. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 59:403-410.

69. Pasternack G, Sulakvelidze A (2014) US Patent 8,685,696 B2.

66. Pasternack G, Sulakvelidze A, Pasternak GR (2009) US Patent 7,635,584 B2.

63. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

62. Lech K (2001) Growing lambda-derived vectors. Curr Protoc Mol Biol 1:1181-1188.

44. Saavedra JE (1985) Reductive alkylation of beta-alcanolamines with carbonyl compounds and sodium borohydride. J Org Chem 50:2271-2277.

45. Saavedra JE (1994) Collision-induced dissociation of uracil and its derivatives. J Am Soc Mass Spectrom 5:339-349.

46. Dudley E (2010) Analysis of free phage DNA by mass spectrometry. Mass Spectrometry of Nucleic Acids and Nucleic Acids, Ed Banouh JD (Limbach, Pittsburgh), pp 435-444.

47. Hayashi H, Nakanishi K, Brandon C, Marmur J (1973) Structure and synthesis of dihydro-putrescinyluracil from bacteriophage SP15 desoxyribonucleic acid. J Am Chem Soc 95:8749-8757.

48. Hansen AS, Thalhammer A, El-Sagheer AH, Brown T, Schofield CJ (2011) Improved synthesis of 5-hydroxymethyl-2′-deoxyctydine phosphoramidite using a 2′-deoxyuridine to 2′-deoxyctydine conversion without temporary protecting groups. Bioorg Med Chem 19:1890-1900.

49. Kung L, Meulis S, Yates CM, Wass MN, Sternberg NJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845-858.

50. Maltman KL, Neuhard J, Lewis HA, Warren RA (1980) Synthesis of thymine and alpha-pyrimethenyluridine in bacteriophage phi W-14-infected Pseudomonas acidovorans. J Virol 34:354-359.

51. Bennett BD, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol 5:593-599.

52. Revel HR, Luria SE (1970) DNA-glucosylation in T-even phage: Genetic determination and role in phage-host interaction. Annu Rev Genet 4:177-192.

53. Yang SJ, Eswert KM, Church GM (2014) CRISPR-Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. PLoS One 9:e98811.

54. Blyson AL, et al. (2015) Covalent modification of bacteriophage T4 DNA inhibits CRISPR-Cas9. mBio 6:e00645-18.

55. Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317-327.

56. Machnik MA, Kaminska KH, Dunin-Horkawicz S, Bujnicki JM (2015) Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes. BMC Bioinformatics 16:336.

57. Bœije D, Black LW (2007) A type IV modification-dependent restriction nuclease that targets glycosylated hydroxymethyl cytosine modified DNAs. J Mol Biol 366:768-778.

58. Scraba DG, Bradley RD, Lerytze-Wills M, Warren RA (1983) Bacteriophage phi W-14: The contribution of covalently bound putrescine to DNA packing in the phage head. Virology 124:152-160.

59. Thivalier L, et al. (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. Proc Natl Acad Sci USA 113:14312-14317.

60. Tsai R, Corrêa IR, Xu MY, Xu S-Y (2017) Restriction and modification of deoxyuridine to 2′-deoxyuridine phosphoramidite using a 2′-deoxyuridine to 2′-deoxyctydine conversion without temporary protecting groups. Bioorg Med Chem 25:1890-1900.

61. Atlas RM (1993) Atlas of Bacterial Artificial Chromosomes. Cold Spring Harbor, New York.

62. Lech K (2001) Growing lambda-derived vectors. Curr Protoc Mol Biol 1:1-11.

63. Schmid G, et al. (2001) Preparation of lambda DNA from phage lysates. Curr Protoc Mol Biol 1:1-11.

64. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

65. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.

66. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of multiple sequence alignments from protein sequences. Comput Appl Biosci 8:275-282.

67. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 59:307-321.

68. Kearse M, et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

69. Pasternack G, Sulakvelidze A (2014) US Patent 8,685,696 B2.

70. Sulakvelidze A, Pasternack GR (2009) US Patent 7,635,584 B2.