Published on Dec 19, 2022 by Arcadia Science **DOI**: 10.57844/arcadia-1ey9-j808

A workflow to isolate phage DNA and identify nucleosides by HPLC and mass spectrometry

This pub details a process for phage amplification and concentration, DNA extraction, and HPLC and MS analysis of phage nucleosides. We optimized the approach with model phages known to use non-canonical nucleosides in their DNA, but plan to apply it for other phages.

Contributors (A-Z)

Januka Athukoralage, Adair L. Borges, Feridun Mert Celebi, Megan L. Hochstrasser, Atanas Radkov, Taylor Reiter, Peter S. Thuy-Boun

Version 4 · Mar 31, 2025

Purpose

DNA extraction, high performance liquid chromatography (HPLC) analysis and mass spectrometry (MS) are bread-and-butter techniques for the chemical analysis of nucleic acids. We optimized this set of protocols to enable such analysis for phage genomes with modified nucleosides, and ultimately hope to use it to discover new DNA modifications from bacteriophages that we isolate from microbial communities.

We're sharing our detailed protocols to help others tackling similar problems. This pub may be useful to anyone studying phage nucleic acids or searching for novel DNA chemistries.

- This pub is part of the **project**, "Exploring bacteriophage nucleic acid chemistries." Visit the project narrative for more background and context.
- All associated **code** is available in this GitHub repository.
- Step-by-step **protocols** are available as a collection on protocols.io.
- Our mass spec **data** is on Zenodo.

Background and goals

Bacteriophages (or phages) are the viruses that infect bacteria. Some phages use DNA modifications to protect their genome from degradation by bacterial immune systems **[1][2][3][4][5][6]**. At Arcadia, we are broadly exploring the distribution and diversity of phage nucleic acid chemistries. One way to do this is to isolate phages from microbial communities and screen them for non-standard DNA chemistries. To do this, we needed a set of protocols that would allow us to quickly determine if a phage we've isolated uses a non-standard nucleoside.

In this pub, we share techniques for chemical analysis of modified phage DNA. We optimized these protocols using two phages with well-studied DNA modifications: phage T4, which has modified cytosines with glucosyl-methyl moieties **[7][8]**, and phage SPO1, which has replaced thymine with hydroxy-methyl uracil **[9][10]**. In future experiments, we will use these protocols to characterize nucleic acids from new phages that we isolate.

The strategy

The phage community developed and routinely uses the approaches that we describe here **[11]**. We're sharing our implementation of these existing methods as part of a straightforward workflow, optimized around detecting modified phage nucleosides. We

will apply this approach to perform chemical analysis of uncharacterized phage genomes in future work.

We are sharing a collection of five protocols (view them all on protocols.io or click below to jump to the corresponding pub section):

1. Phage amplification and concentration
2. Phage DNA extraction with Monarch kit and digestion to single nucleosides
3. Phage DNA extraction with phenol-chloroform and digestion to single nucleosides
4. Nucleoside analysis with high-performance liquid chromatography (HPLC)
5. Nucleoside analysis with liquid chromatography–tandem mass spectrometry (LC–MS/MS)

These methods should be applicable to any laboratory-cultivated phage that can be grown to sufficiently high concentration to enable successful nucleic acid extraction.

The method

The following is a high-level overview of our approach, also visually summarized in Figure 1*. You can view detailed, step-by-step protocols in this collection on protocols.io.*

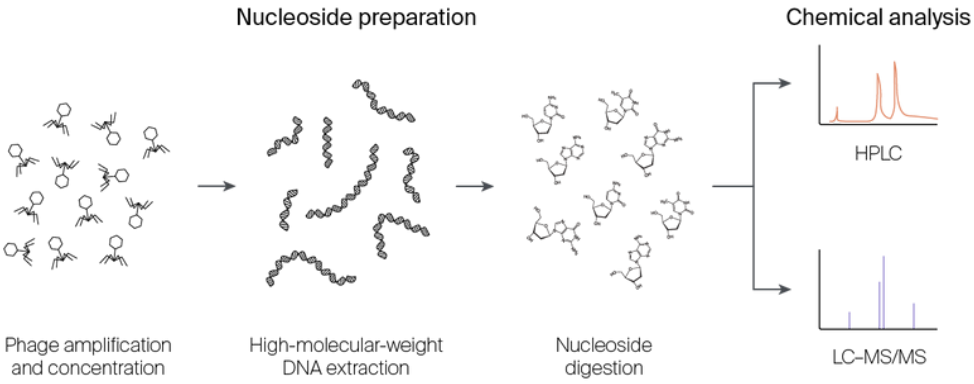

Figure 1

Overview of our general workflow for chemical analysis of phage genomes.

Starting with a pure culture of phage, these protocols detail phage amplification, concentration, DNA extraction, nucleoside digestion, and chemical analysis of phage nucleosides with HPLC LC–MS/MS. We optimized these steps using model dsDNA phages with known genome modifications. Phage T4 infects *Escherichia coli* and has modified cytosines with glucosyl-methyl moieties **[7][8]**. Phage SPO1 infects *Bacillus subtilis* and has replaced thymine with hydroxy-methyl uracil **[9][10]**. These phages and their hosts are easy to work with, and have well-characterized nucleic acid chemistries. This makes them an ideal starting point for researchers looking to establish methods to study phage nucleic acid chemistry.

Below, we detail our protocols and results from analyzing phage T4 and SPO1 genomes. While we developed the protocols using these model lytic dsDNA phages, we anticipate that they can be tweaked to enable chemical analysis of phages that have different growth conditions or ssDNA or RNA genomes.

Step 1: Phage amplification and concentration

This approach to phage genome analysis begins with amplifying the phage to a high titer. Both T4 and SPO1 are lytic phages that grow well in liquid culture, and so we chose to amplify the phage in 30 mL of broth media. We supplemented the media with 1 mM $MgSO_4$ and 1 mM $CaCl_2$ to enhance phage adsorption. This worked well for our model phages — in 30 mL, we obtained a concentration of 10^{10} PFU/mL for T4 and 10^9 PFU/mL for SPO1.

We anticipate that in the future, some of our newly isolated phages may need to be propagated using slightly different techniques. Temperate phages should be amplified using the double-agar overlay method **[12]**, and some large diffusion-limited phages may benefit from using in-gel techniques **[13]**. Also, the identities and levels of cations may need to be adjusted depending on the individual biology of the phage.

After amplification, we concentrated the 30 mL of phage lysate down to 300 µL for DNA extraction. To concentrate the phage, we found that both PEG precipitation and filtration-based concentration worked well. PEG precipitation requires less hands-on time, but is overall longer as it requires an overnight incubation step. We also suspect that individual phages will be differentially sensitive to these concentration methods, so one should select a concentration protocol that works best for their phage of interest.

TRY IT: The full protocol, "Phage amplification and concentration," is available on protocols.io (DOI: 10.17504/protocols.io.yxmvmnb86g3p/v1).

Step 2: Phage DNA extraction and digestion to single nucleosides

After amplification and concentration, the phages are ready for DNA extraction. Initially, we chose to use the NEB Monarch kit to extract high-molecular-weight (HMW) DNA. While any approach that can harvest high-purity phage DNA would be appropriate here, we chose a method that would generate HMW DNA compatible with long-read Nanopore sequencing. We started with the Monarch kit because it can be performed on a benchtop.

Using the Monarch kit, we obtained high concentrations of high-purity T4 and SPO1 DNA. We used a Nanodrop spectrophotometer to quickly check the concentration and purity, and downstream chemical analyses (HPLC and LC–MS/MS) also confirmed the purity of the DNA (Table 1). Note that SPO1 has a high 260/280 ratio: this is because it contains uracil, and thus has an "RNA-like" 260/280 value.

TRY IT: The full protocol, "Phage DNA extraction with Monarch kit and digestion to single nucleosides," is available on protocols.io (DOI: 10.17504/protocols.io.3byl4j2p8lo5/v1).

Phage	Phage input (PFU/mL)	DNA concentration (ng/µL)	Total DNA (µg)	260/230	260/280
T4	3×10^{11}	52.7	5.27	1.79	1.90
SPO1	3×10^{10}	181.3	18.13	1.99	2.18

Table 1. DNA yields.

In further iterations of this experiment, we switched to using phenol-chloroform extraction to harvest HMW phage DNA. Phenol-chloroform extraction cannot be

performed on a benchtop, and generates substantial chemical waste. However, we found that for some phages, phenol-chloroform succeeded when the Monarch kit prep failed to yield DNA. When harvesting DNA for new phages, we now routinely use phenol-chloroform as it appears to be a more robust method.

After DNA isolation, we digested 1 µg of DNA from each phage sample down to single nucleosides using the NEB Nucleoside Digestion Mix. We chose this kit because it is directly compatible with HPLC and LC–MS/MS.

> **TRY IT**: The full protocol, "Phage DNA extraction with phenol-chloroform and digestion to single nucleosides," is available on protocols.io (DOI: 10.17504/protocols.io.8epv5jrxnl1b/v1).

Step 3: Phage nucleoside analysis with high-performance liquid chromatography (HPLC)

Once the DNA is broken down into single nucleosides, those nucleosides can be analyzed using HPLC. We developed a 30-minute binary gradient using a reverse-phase column, which provided great peak resolution (Figure 2). In addition, we developed a short 10-minute isocratic gradient that we may use for higher-throughput analysis of nucleosides.

To analyze phage nucleosides, we first ran a set of standard deoxynucleosides (dA, dT, dG, dC, dU — each at 1 mg per mL) to obtain retention times for unmodified nucleosides (Figure 2, A). These standards should be included in each HPLC run. To analyze the samples for modified nucleosides, we injected 100 ng into the HPLC and compared the retention times of the sample nucleosides to the standards. We also plotted the A_{260} values to see the full sample content.

Some nucleoside modifications are easy to spot visually by looking at A_{260} absorbance plotted over time. T4 phage has two small peaks that correspond to alpha and beta glucosylmethyl deoxycytidine, and is missing a canonical deoxycytidine peak (Figure 2, B). Similarly, SPO1 is obviously missing a thymidine peak, and instead has a new peak that corresponds to hydroxymethyl deoxyuridine (Figure 2, C). However, the difference in retention time between the deoxyuridine standard and the hydroxymethyl

deoxyuridine peak in SPO1 is very small, and easily missed. We interpret this to mean that HPLC analysis is good for quickly flagging large-scale changes to nucleic acid composition, but less sensitive to other changes.

TRY IT: The full protocol, "Nucleoside analysis with high performance liquid chromatography (HPLC)," is available on protocols.io (DOI: 10.17504/protocols.io.5jyl8jn39g2w/v1).

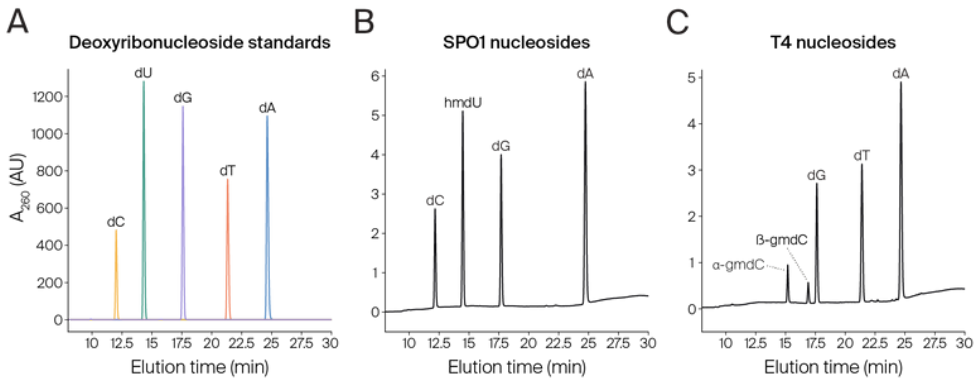

Figure 2

HPLC elution profiles.

Nucleoside elution profiles plotted by absorbance at 260 nanometers (A260, AU: arbitrary units) over time in minutes (min). Each nucleoside peak is labeled with its corresponding identity.

A) Elution profiles of deoxyribonucleoside standards.

B) Elution profile of digested SPO1 phage nucleosides.

C) Elution profiles of digested T4 phage nucleosides.

dA: deoxyadenosine, dG: deoxyguanosine, dT: thymidine, dC: deoxycytidine, hmdU: hydroxymethyl-deoxyuridine, gmdC: glucosylmethyl-deoxycytidine

Step 4: Nucleoside analysis with liquid chromatography–tandem mass spectrometry (LC–MS/MS)

LC–MS/MS is our most sensitive tool for analyzing nucleosides. We analyzed nucleosides derived from 500 ng of DNA, digested with the NEB Nucleoside Digestion Mix. This kit is directly compatible with LC–MS/MS. In our LC–MS/MS run, we first separated nucleosides using a binary solvent gradient on a C18 column. This gradient is not optimized, but generated usable data and works as a starting point for further optimization. We acquired data in positive mode with an MS1 scan targeting ions in the 200–800 m/z range, and followed each MS1 scan with seven data-dependent MS2 scans. In this experiment, we used a Thermo LTQ Orbitrap XL at the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley.

> **TRY IT**: The full protocol, "Nucleoside analysis with liquid chromatography–tandem mass spectrometry (LC–MS/MS)," is available on protocols.io (DOI: 10.17504/protocols.io.q26g7yrq1gwz/v1).

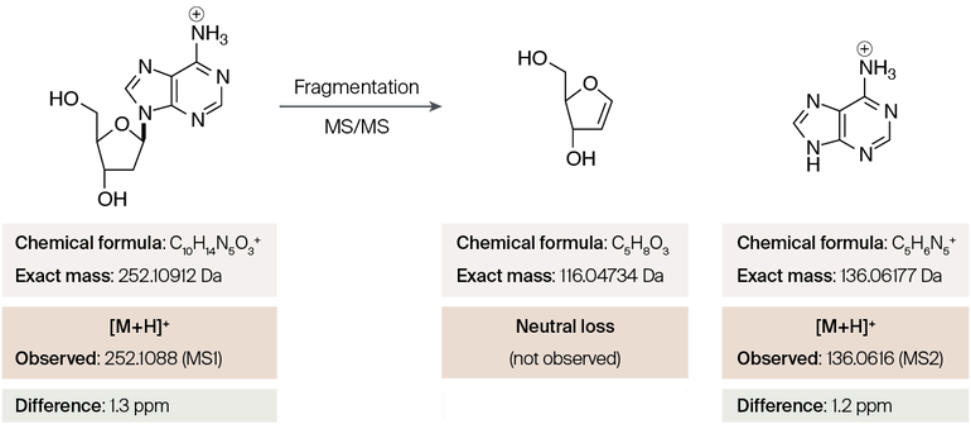

Figure 3

Fragmentation patterns of nucleosides.

Nucleosides fragment via neutral loss of the deoxyribose sugar, while the charged nitrogenous base can be detected directly. [M+H]+ indicates a detected positively charged ion, which we can identify by comparing its observed mass to the expected masses of different nucleoside components.

We manually inspected mass spectrometry data and noticed a consistent pattern of –116 m/z differences between probable nucleoside precursor ions and their most prominent fragmentation product ions, suggesting a pattern of deoxyribose neutral mass loss during fragmentation (Figure 3). Based on this pattern, we wrote Python scripts in Jupyter notebooks to automate nucleoside identification within our accurate mass high-resolution dataset.

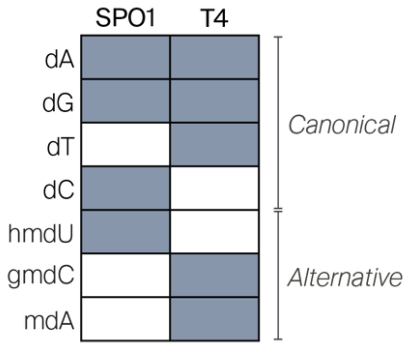

Figure 4

Detection of canonical and alternative nucleosides in phage genomes with mass spectrometry.

This presence/absence chart reflects nucleosides observed in LC–MS/MS analysis of SPO1 and T4 phage genomes. Grey indicates that we detected the nucleoside using LC–MS/MS, while white indicates that we did not detect the nucleoside.

dA: deoxyadenosine, dG: deoxyguanosine, dT: thymidine, dC: deoxycytidine, hmdU: hydroxymethyl-deoxyuridine, gmdC: glucosylmethyl-deoxycytidine, mdA: methyl-deoxyadenosine.

Taking advantage of this consistent fragmentation pattern for nucleosides, we identified ions that corresponded to the nucleosides known to be in phage T4 and SPO1 (Figure 4). We also identified an ion in the T4 sample that corresponds to methylated deoxyadenosine, which the HPLC analysis missed, highlighting the increased sensitivity of LC–MS/MS (Figure 4). This methylation mark was likely added by the *E. coli* strain B Dam methylase **[14]** or the T4 Dam methylase **[15]**, which methylate adenine at GATC motifs **[16]**.

All **code** generated and used for the pub is available in this GitHub repository (DOI: 10.5281/zenodo.7447542), including a Jupyter notebook to find nucleosides in mass spec data; mass lists for nucleosides, charged adducts, and neutral adducts; and outputs.

SHOW ME THE DATA: Access our raw and processed **mass spec data** on Zenodo (DOI: 10.5281/zenodo.7319990).

Challenges identifying nucleosides in complex community samples

We developed this set of protocols using phages with known genome modifications, ultimately aiming to apply them to uncultured phages with potentially novel modifications in microbial community samples. We've chosen to shift away from these scientific directions, but we're sharing our data sets and the issues we encountered to help others working on similar questions.

LC–MS/MS

We tried applying the LC–MS/MS assay to analyze DNA extracted from microbial communities and viromes to see if we could detect nucleoside modification without first individually isolating bacteriophages, but were largely unsuccessful.

We worked with the CRO Arome to use LC–MS/MS to profile the nucleoside content of cheese microbial communities. We chose this CRO because they have a highly sensitive Orbitrap Exploris 480 machine that can take high-resolution measurements, which we thought would be necessary for analyzing potentially complex nucleoside samples from natural communities. We used phenol-chloroform extraction to harvest DNA from cheese microbial communities and their paired viromes (see this protocol

collection for methods details) and analyzed the digested nucleosides via LC–MS/MS with a HILIC column in positive ion mode under neutral pH.

Unfortunately, we didn't achieve the sensitivity that we would need to detect rare, non-standard nucleotides using this approach. For example, we did not see any signal for the nucleoside thymidine (dT) in the MS1, meaning our approach was not even sensitive enough to detect one of the four most abundant nucleosides in the community. If we were going to follow up on this, we would need to put a lot more work into methods development to increase the sensitivity and dynamic range of the assay.

Another issue we saw was a high level of background from RNA nucleosides in our sample, despite the DNA samples having gone through an RNase treatment. We hypothesize that trace RNA nucleosides must have persisted after the digestion, and then were more ionizable than the DNA nucleosides, leading to their enhanced detection in LC–MS/MS. If we were to do this again, we would run the samples through a DNA cleanup column to remove small RNA oligos and/or lingering nucleosides. If anyone wants to explore the raw data, we've shared it on Zenodo.

> **SHOW ME THE DATA**: Our raw LC–MS/MS data from cheese communities and paired viromes, first-pass analysis, and methods details are available on Zenodo (DOI: 10.5281/zenodo.7996414).

Nanopore sequencing

We also hoped to complement these chemical methods with Nanopore-based modification discovery to directly link phage genome sequences to their chemical composition **[17]**. Briefly, we generated paired WGA:native R10 chemistry data sets of cheese microbial communities using Nanopore sequencing (read more about this in **[18]**). Unfortunately, we found that the *de novo* modification prediction tools only worked well with R9 chemistries. We have shared the FAST5 files through the European Nucleotide Archive (ENA) for others to use in tool development, and encourage others to reuse the data.

Acknowledgements

Thank you to the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley (NIH grant number 1S10OD020062-01) for mass spectrometry of isolated phage nucleosides. Thank you to Arome for mass spectrometry of cheese community nucleosides. Thanks also to the Guertin lab for sharing DNA helix paint brushes for Adobe Illustrator.

1 Samson JE, Magadán AH, Sabri M, Moineau S. (2013). Revenge of the phages: defeating bacterial defences. https://doi.org/10.1038/nrmicro3096

2 Weigele P, Raleigh EA. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. https://doi.org/10.1021/acs.chemrev.6b00114

3 Bryson AL, Hwang Y, Sherrill-Mix S, Wu GD, Lewis JD, Black L, Clark TA, Bushman FD. (2015). Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. https://doi.org/10.1128/mbio.00648-15

4 Flodman K, Corrêa IR Jr, Dai N, Weigele P, Xu S. (2020). In vitro Type II Restriction of Bacteriophage DNA With Modified Pyrimidines. https://doi.org/10.3389/fmicb.2020.604618

5 Wang S, Sun E, Liu Y, Yin B, Zhang X, Li M, Huang Q, Tan C, Qian P, Rao VB, Tao P. (2022). The complex roles of genomic DNA modifications of bacteriophage T4 in resistance to nuclease-based defense systems of *E. coli*. https://doi.org/10.1101/2022.06.16.496414

6 Hutinet G, Kot W, Cui L, Hillebrand R, Balamkundu S, Gnanakalai S, Neelakandan R, Carstens AB, Fa Lui C, Tremblay D, Jacobs-Sera D, Sassanfar M, Lee Y-J, Weigele P, Moineau S, Hatfull GF, Dedon PC, Hansen LH, de Crécy-Lagard V.

(2019). 7-Deazaguanine modifications protect phage DNA from host restriction systems. https://doi.org/10.1038/s41467-019-13384-y

7 Sinsheimer RL. (1954). Nucleotides from T2r + Bacteriophage. https://doi.org/10.1126/science.120.3119.551

8 Lehman IR, Pratt EA. (1960). On the Structure of the Glucosylated Hydroxymethylcytosine Nucleotides of Coliphages T2, T4, and T6. https://doi.org/10.1016/s0021-9258(20)81347-7

9 Kallen RG, Simon M, Marmur J. (1962). The occurrence of a new pyrimidine base replacing thymine in a bacteriophage DNA: 5-hydroxymethyl uracil. https://doi.org/10.1016/s0022-2836(62)80087-4

10 Hoet PP, Coene MM, Cocito CG. (1992). REPLICATION CYCLE OF *BACILLUS SUBTILIS* HYDROXYMETHYLURACIL-CONTAINING PHAGES. https://doi.org/10.1146/annurev.mi.46.100192.000523

11 Lee Y-J, Weigele PR. (2020). Detection of Modified Bases in Bacteriophage Genomic DNA. https://doi.org/10.1007/978-1-0716-0876-0_5

12 Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. (2009). Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. https://doi.org/10.1007/978-1-60327-164-6_7

13 Serwer P, Wright ET. (2020). In-Gel Isolation and Characterization of Large (and Other) Phages. https://doi.org/10.3390/v12040410

14 Marinus MG, Morris NR. (1973). Isolation of Deoxyribonucleic Acid Methylase Mutants of *Escherichia coli* K-12. https://doi.org/10.1128/jb.114.3.1143-1150.1973

15 Kossykh VG, Schlagman SL, Hattman S. (1995). Phage T4 DNA [N]-adenine6Methyltransferase. OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION. https://doi.org/10.1074/jbc.270.24.14389

16 Geier GE, Modrich P. (1979). Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. https://doi.org/10.1016/s0021-9258(17)34217-5

17 Kot W, Olsen NS, Nielsen TK, Hutinet G, de Crécy-Lagard V, Cui L, Dedon PC, Carstens AB, Moineau S, Swairjo MA, Hansen LH. (2020). Detection of preQ0 deazaguanine modifications in bacteriophage CAjan DNA using Nanopore sequencing reveals same hypermodification at two distinct DNA motifs. https://doi.org/10.1093/nar/gkaa735

18 Borges AL, Dutton RJ, McDaniel EA, Reiter T, Weiss ECP. (2024). Paired long- and short-read metagenomics of cheese rind microbial communities at multiple time points. https://doi.org/10.57844/ARCADIA-0ZVP-XZ86