ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

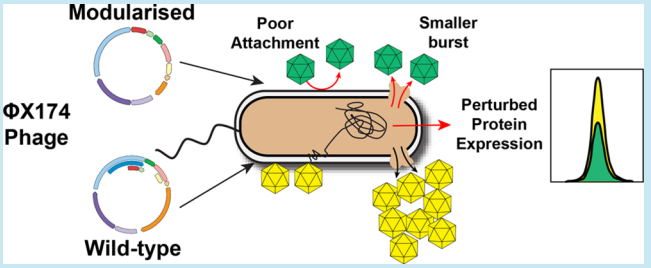
Genome Modularization Reveals Overlapped Gene Topology Is Necessary for Efficient Viral Reproduction

Bradley W. Wright, Juanfang Ruan, Mark P. Molloy,* and Paul R. Jaschke*

Cite This: *ACS Synth. Biol.* 2020, 9, 3079–3090 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Sequence overlap between two genes is common across all genomes, with viruses having high proportions of these gene overlaps. Genome modularization and refactoring is the process of disrupting natural gene overlaps to separate coding sequences to enable their individual manipulation. The biological function and fitness effects of gene overlaps are not fully understood, and their effects on gene cluster and genome-level refactoring are unknown. The bacteriophage φX174 genome has ~26% of nucleotides involved in encoding more than one gene. In this study we use an engineered φX174 phage containing a genome with all gene overlaps removed to show that gene overlap is critical to maintaining optimal viral fecundity. Through detailed phenotypic measurements we reveal that genome modularization in φX174 causes virion replication, stability, and attachment deficiencies. Quantitation of the complete phage proteome across an infection cycle reveals 30% of proteins display abnormal expression patterns. Taken together, we have for the first time comprehensively demonstrated that gene modularization severely perturbs the coordinated functioning of a bacteriophage replication cycle. This work highlights the biological importance of gene overlap in natural genomes and that reducing gene overlap disruption should be an integral part of future genome engineering projects.



KEYWORDS: *bacteriophage, genome engineering, proteomics, synthetic biology, virus structure, refactoring*

The arrangement of genes within a genome is well recognized to be an important feature in the regulation of gene expression,[1−7] and therefore, gene order tends to be evolutionarily conserved.[8−14] Studies have investigated implications of ectopic gene expression revealing functional significance to gene order.[15−18] For example, studies in bacteriophage T7 have highlighted the criticality of gene ordering in achieving the upper limits of fitness,[16,19] and significantly, this fitness attenuating gene reordering is difficult to purge even after long-adaptation cycles.[20] Therefore, despite important functional significance revealed by these studies, rules surrounding the relationship of genome architecture and phenotypic outcome are still only loosely defined.

In addition to gene order, another related feature of genome topology has been gene sequence overlap, which is common across viral,[21,22] prokaryotic,[23,24] and eukaryotic[25−27] genomes, and for that reason, is increasingly thought to be a critical component of genome architecture.[23,28−31] In the model eukaryote yeast, gene overlaps are prevalent, with estimates suggesting 13.7% are involved in different strand overlap, and 4.2% in same-strand overlap.[32] In viruses, gene overlaps are more prevalent and seem to be a result of hard constraints put on genome size as a result of capsid volume and stability, restricting genome length and nucleotide composition.[22,33] Once in place, gene overlaps also seem to have some role in gene expression regulation.[23,34−36]

In prokaryotes and viruses, whereby gene overlap is often found within open reading frames,[23,28] the overlap is thought to arise from mutations resulting in start codon creation within existing genes, or between adjacent genes by mutations removing start or stop codons.[37,38] This results in 5′ or 3′ extensions of the existing gene into the adjacent gene.[37,38] This phenomenon is often called overprinting[39] and may also occur on the antisense strand,[40] typically resulting in the generation of nonprotein encoding but functionally relevant cis or trans-acting RNA.[41]

Genome modularization and refactoring is the process of disrupting natural gene overlaps and coding sequences to enable their individual manipulation and the removal of cryptic regulation.[42] Only a few studies have investigated changes to fitness from uncoupling (modularizing) gene overlap.[43−47] With one exception, in which the desired pathway product was increased,[43] modularization has resulted in fitness loss,[44−47] but

Received: June 17, 2020
Published: October 12, 2020

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079–3090

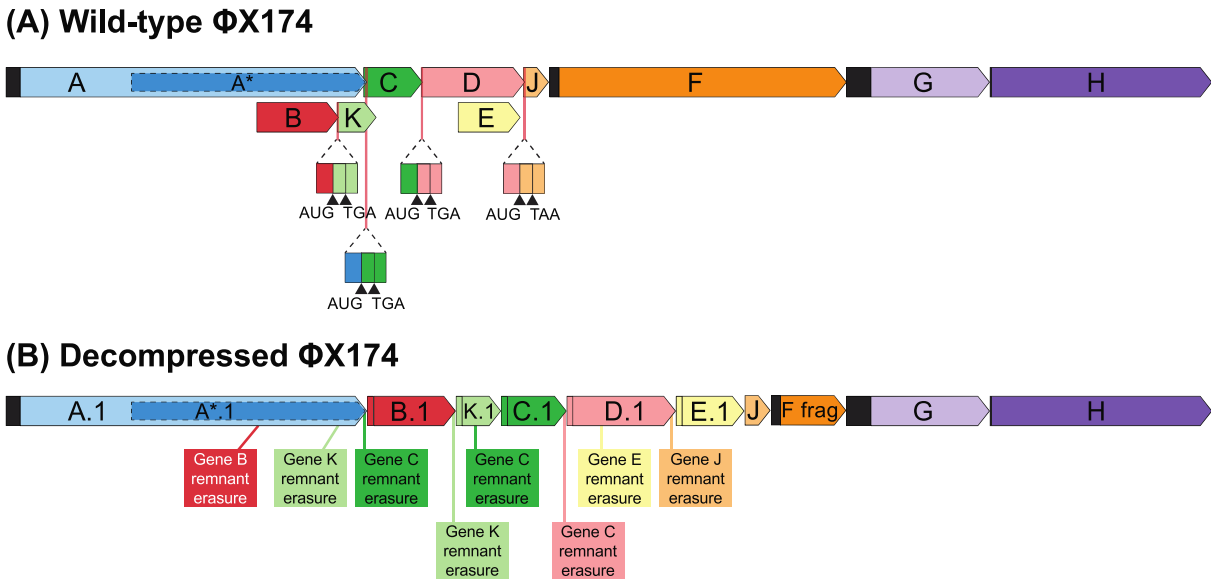


Figure 1. Wild-type and decompressed φX174 genome topologies. (A) In the wild-type φX174 genome, genes B and E are examples of overprinted gene overlap while tandem same-strand start/stop codon overlaps are shown for overlaps between genes B−K, A−C, C−D, and D−J. (B) In the decompressed φX174 genome all overprinted genes and start/stop codon overlaps are removed. Locations of synonymous codon modification to remove legacy start codons and ribosome binding sites are shown. Genes with the 0.1 nomenclature have had synonymous codon changes in at least one location within the gene. Black boxes are intergenic regions.

to date, no study has defined the molecular mechanisms underlying these fitness defects.

The functional effects of gene overlap disruption are extremely pertinent to genome engineering and refactoring projects, which frequently disrupt sequence overlaps during reconfiguration of native genes and pathways. A number of projects have cautiously sought to minimize the potential impacts of overlap disruption through explicit design choices such as overlap sequence duplication,[48−50] but these efforts to address the biological implications of gene overlap disruption have been *ad hoc*. Currently, there is no consensus on how to treat gene overlaps in genome engineering and refactoring projects.

In this study, we comprehensively define the range of phenotypes displayed by a whole genome modularized bacteriophage that was constructed previously.[44] Apart from the in-frame initiating A* gene, the aptly named "decompressed" φX174 had all instances of gene overlap modularized. The decompressed φX174 genome modifications removed gene overlaps of entirely overprinted genes (gene B within A, gene K within A and C, and gene E within D) (Figure 1). The genome modifications also removed the start/stop codon overlaps of tandem same-strand gene overlaps of (genes B−K, A−C, C−D, and D−J) (Figure 1). These changes were all made while maintaining absolute genome length, codon usage bias, and G+C content.[44]

Initial characterization of the resulting phage revealed comparable lysis efficiency with respect to the wild-type strain under the limited conditions tested.[44] To build upon that work, here we have deeply characterized decompressed φX174 using targeted proteomics to temporally resolve the complete proteome of decompressed and wild-type φX174 across an entire infection cycle. We find that decompressed φX174 has severe deficiencies in expression of several key replication proteins and displays a broad range of phenotypes associated with low fitness. As a result of this work we gain new insights into the effect of genome modularization and highlight the need to consider gene overlaps as a critical component of genome architecture, especially in the context of future genome engineering projects.

RESULTS AND DISCUSSION

Decompressed φX174 Displays Comparable Lysis Timing but Impaired Progeny Production. To understand more deeply the phenotypic effects of removing all overlaps from the genome of bacteriophage φX174, we first performed plaque assays and liquid growth assays of decompressed φX174. Wild-type and decompressed φX174 phage were used to infect *Escherichia coli* C122(pJ804(Gene F)) strain under inducing conditions. Plaque assays of the infections showed that the decompressed φX174 plaque size was significantly smaller, at 4.1 ± 0.9 mm^2, compared to 29.2 ± 8.0 mm^2 for wild-type φX174 (p-value < 0.00001) (Figure 2A).

During phage infection, there is an intricate balance between progeny production and lysis initiation, and there are a range of factors that can decrease plaque size, including decreased latent period or burst-size.[51−53] The φX174 protein E mediates host lysis, with an estimated 500 molecules sufficient to cause membrane disruption and cell death.[54] If decompressed φX174 was producing protein E sooner in the infection cycle than wild-type φX174 or was producing less E protein, that may explain the reason for smaller plaque size due to reduced progeny production from premature lysis or inefficient lysis.

To determine whether lysis-timing or lack of effective lysis was the cause of the decompressed φX174 small-plaque phenotype, we infected exponentially growing *E. coli* C122(pJ804(Gene F)) strain with either decompressed or wild-type φX174 with a multiplicity of infection (MOI) = 5 and measured culture absorbance for 60 min. We used a high phage load to ensure the majority (>99%) of host cells had one or more phage attached and we were only observing the effects of one infection cycle. The results of this time-course showed that decompressed and wild-type φX174 have comparable lysis timing under our

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

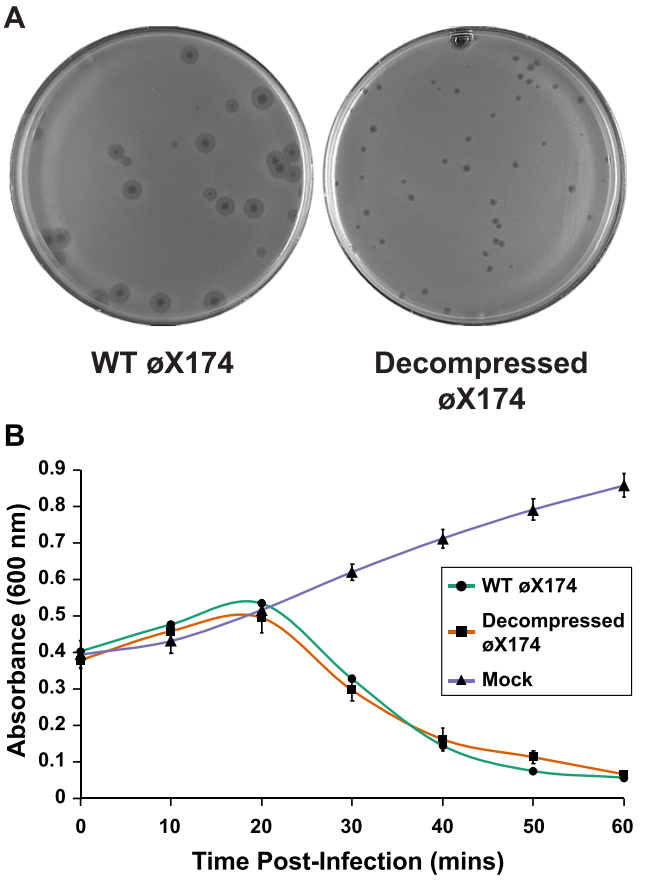


Figure 2. Decompressed φX174 lyses the host at similar efficiency to wild-type φX174 at high multiplicity of infections but has reduced fecundity overall. (A) Double-agar overlay plates of wild-type φX174 and decompressed φX174 infecting host *E. coli* C122(pJ804(Gene F)). Wild-type φX174 plates contained 2 mM glucose and decompressed φX174 plates contained 2 mM rhamnose. Plate diameter = 85 mm. (B) Lysis curve at multiplicity of infection (MOI) = 5. Error bars represent one standard deviation (n = 2−3).

conditions (Figure 2B) with no evidence for phage cross-contamination (Figure S1).

We next measured the burst-size of decompressed φX174 to determine if the small-plaque phenotype could be due to fewer viable progeny produced per infected cell. We determined the burst-size of wild-type φX174 to be 145 ± 63 phages produced per cell while decompressed φX174 produced only 16 ± 2 phages per cell. Previous measurements of wild-type φX174 burst-size have reported between 160 and 180 phages produced per cell.[55−57]

The significant burst-size deficiency in decompressed φX174 is the likely reason for the small plaque size in this engineered phage (Figure 2A). Poor lysis is not the reason for altered burst-size in decompressed φX174 as lysis timing is not altered in a single-step growth experiment (Figure 2B). Additionally, we removed all unattached phages before infection initiation in this assay, effectively excluding aberrant host attachment or recognition as the cause of this burst-size deficiency. We next assessed whether phage virion structure or genome ejection into cells could be the cause of the reduced burst-size. The decompressed φX174 genome encodes proteins with the same amino acid sequence as the wild-type φX174 genome, and on this account, we expect any structural changes to be either from disruption of the intimate genome-capsid interactions due to packaging a genome of different sequence or from inefficient assembly due to altered protein levels during capsid assembly.

Decompressed φX174 Virion Has Poor Heat Stability and Attachment Efficiency. To identify whether decompressed φX174 virions are less stable than wild-type due to the differences in the packaged genome, we measured the effect of heat stress on phage infectivity. Samples of purified wild-type and decompressed φX174 virions were subjected to 60 °C over a range of times, and the remaining infectivity was compared to untreated samples. Decompressed φX174 virions were rapidly inactivated by the heat treatment after only 5 min, while wild-type φX174 was slightly more infectious after this treatment (Figure 3A). A longer 15 min treatment also significantly decreased decompressed φX174 infectivity more than wild-type (Figure 3A).

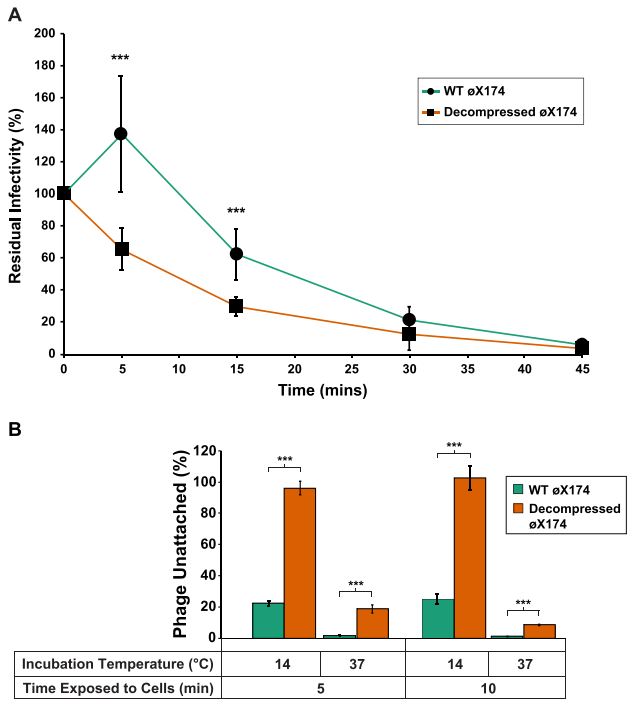


Figure 3. Decompressed φX174 is both less heat stable and poorer at attaching to host cells than wild-type φX174. (A) Heat stability of phage. Samples of decompressed φX174 and wild-type φX174 virus were exposed to heat stress (60 °C) for 5, 15, 30, and 45 min followed by measurement of residual infectivity by double-agar plate plaque enumeration. Error bars represent one standard deviation (n = 9). ***p-value $<1.0 \times 10^{-4}$. (B) Viral attachment assay. Lysis resistant *E. coli* C900 Δ*slyD* cells were incubated at the specified temperature with either wild-type φX174 or decompressed φX174 phage for either 5 or 10 min followed by enumerating unattached phage using the double-agar overlay method. Error bars represent one standard deviation (n = 4). ***p-value $<1.0 \times 10^{-4}$.

This result indicates that the modularized genome of decompressed φX174 is having a negative effect on the thermal stability of the virion. This is not totally unexpected, as studies addressing heat stability of viral capsids with genomic variation have revealed different thermal stability between counterparts.[58−60] The physical properties of constituents within capsid shells (for example, salt ions and nucleotides) have been shown to directly influence a capsid's internal osmotic pressure, and consequently, mechanical stability.[61−64] So, while the decom-

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

pressed φX174 genome encodes for the same genes and proteins, and is of the same absolute nucleotide length as the wild-type, it is reasonable to suggest that genome organization is critical to maintaining thermal stability by maintaining optimal internal pressure and internal capsid interactions. Genome modularization and the resulting alterations to the genome secondary structure (Figure S2) could alter the electrostatic interactions within the capsid,[65−67] therefore directly influencing capsid rigidity. These complex interactions and their impact on particle stability and genome packaging have been observed before in other viruses.[60,68−70]

Given the lower thermal stability of the decompressed φX174 virion we next measured the attachment efficiency to determine virion host-recognition and subsequent attachment. Attachment for wild-type and decompressed φX174 was measured at 14 and 37 °C. At 14 °C in starvation buffer φX174 adsorbs to the host cell without undergoing eclipse (irreversible structural change).[71,72] This characteristic of φX174 is particularly advantageous as it allows for the synchronization of infection. By contrast, at 37 °C in starvation buffer φX174 undergoes an irreversible structural change leading to the partial ejection of the genome.[71]

The results of the attachment assay at 14 °C showed that decompressed φX174 virions could not remain stably attached to host cells, as opposed to wild-type φX174 that had 77.8% (±1.6%) virions remaining attached after incubation with hosts for 5 min (Figure 3B). At 37 °C decompressed φX174 virions attach more stably to host cells, with 81.3% (±2.7%) left attached, but was still significantly less than wild-type under the same conditions (91.5% (±0.6%), *p*-value = 0.00004). Longer 10 min incubations did not substantially improve relative decompressed φX174 attachment (Figure 3B). These attachment assay results suggest that decompressed φX174 virions can recognize the host cell receptor, as they are able to remain stably attached at 37 °C and eclipse, but at 14 °C, pre-eclipsed decompressed φX174 does not remain stably adsorbed to the host.

Together, these results provide evidence that the packaged genome of decompressed φX174 is impacting the properties of the virion, which in turn impairs attachment. We did not measure the specific infectivity of decompressed φX174 as particles of the requisite purity were not obtainable, but other work has shown a critical effect on capsid and viral infectivity from genomic cargo. For instance, Zeng *et al.* (2017),[73] measured the mechanochemical properties of adeno-associated virus 2 and brome mosaic virus and found that adding exogenous genetic material to these viral genomes has an influence on virus-substrate adhesion interactions with differing deformation and strong orientation biases for adsorption depending upon the genomic content.[73] Further work by this group[74] has illuminated the capsid orientation bias and deformity arising from adsorption in the structurally similar brome mosaic virus. The orientation bias as a result of the deformity of the capsid to maximize contact area would involve compressing the capsid inner surface and extending the outer surface,[74] actions that will be impacted by properties of the encapsulated genome.[73] As reflected by the attachment assay results, these mechanochemical properties are likely perturbed in the decompressed φX174 virion. To further support the notion that the genome-capsid interactions have profound effects on the dynamic function of virions, previous work in φX174 has shown that altering the internal capsid environment through mutant J proteins also alters attachment efficiency.[75−77] Additionally, altering the packaged genome sequence was also found to have profound effects on capsid biophysical properties.[76] These studies, along with our experiments, suggest that the genome of the decompressed φX174 virion is influencing the structural integrity and surface properties of the capsid, leading to perturbation of pre-eclipse attachment and capsid heat stability. To our knowledge, this study demonstrates for the first time that native gene architecture is critical for optimal capsid function.

Decompressed φX174 Capsid Morphology Is Identical to Wild-Type. To identify any major structural aberrations in the decompressed φX174 capsid that may be causing the decreased attachment efficiency, we incubated purified virions with either DNase I or proteinase K and measured changes to viability. The results of this assay revealed that decompressed φX174 virions are slightly more stable under protease stress than wild-type (*p*-value 0.0173), but there was no significant difference under DNase I treatment, thereby making altered capsid permeability or partial genome ejection unlikely (Figure S3).

To characterize subtler variations in capsid structure between wild-type and decompressed φX174 we used cesium chloride (CsCl) gradient purification to isolate virions and examined them by transmission electron microscopy (TEM). Visualization and subsequent 2D-classification of wild-type and decompressed φX174 virions revealed no observable compositional differences in major spike protein (G) or major capsid protein (F) in the decompressed virion (Figure 4 and S4). Additionally, the diameter of both decompressed and wild-type virions was measured to be 33 nm, in close agreement with wild-type measurements in other studies,[78−80] which is suggestive of correct protein J and protein H composition. As no gross structural deformities were observed with the decompressed

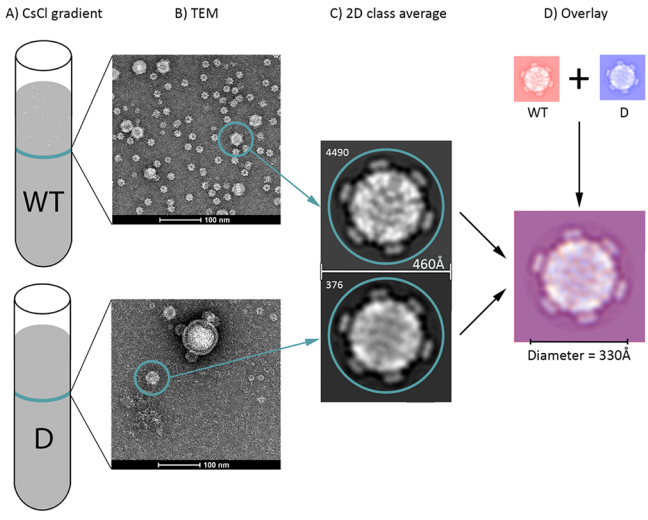


Figure 4. Decompressed φX174 and wild-type φX174 virion morphologies are indistinguishable. (A) Phage band isolated after CsCl step-gradient ultracentrifugation. WT = wild-type and D = decompressed φX174. (B) Uranyl-acetate negative stained particles visualized using transmission electron microscopy (TEM). Phage particles found in both preparations (teal circle). (C) 2D class averaging of phage capsids highlighted in (B). 4490 particles averaged for the wild-type capsid, and 376 particles averaged for the decompressed φX174. Additional classifications and particle information available as Figure S4 and S5. (D) 2D class averaged images of capsids overlaid reveal no obvious morphology differences between wild-type φX174 and decompressed φX174.

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

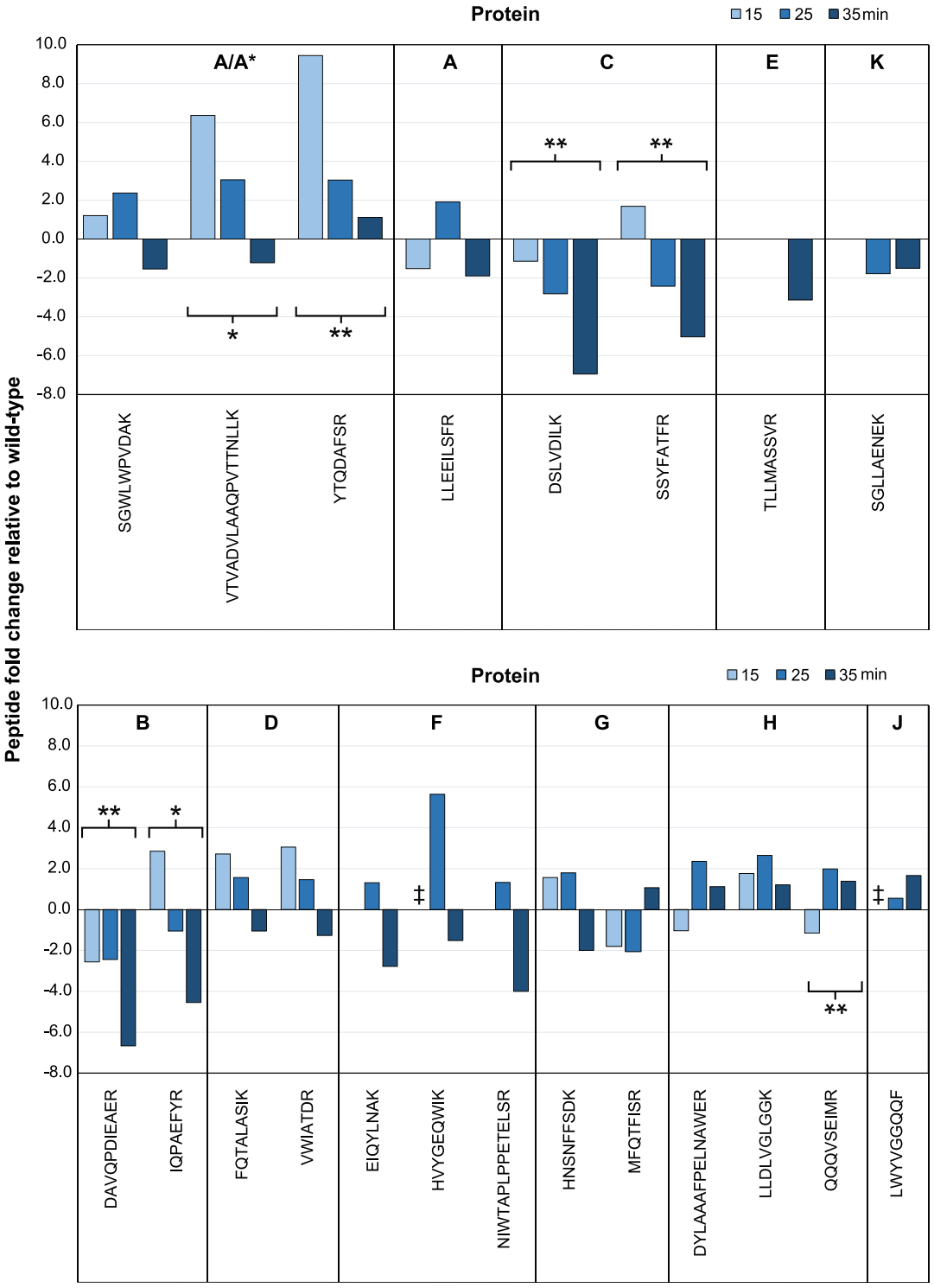


Figure 5. Mass spectrometric quantitation of φX174 proteins produced during infection shows altered expression levels of proteins involved in genome replication and capsid packaging and assembly in decompressed φX174. PRM quantitation of targeted viral peptides across four time-points comparing decompressed relative to wild-type. (Top) Fold-change of peptides from replication associated proteins. (Bottom) Fold-change of peptides from capsid associated proteins. ‡ = Peptide detected in decompressed φX174 but not in the wild-type φX174. Two-factor ANOVA significance *p-value < 0.05. **p-value < 0.005. Plotted summed intensities available as Figures S6–S9.

virion, we lean toward the hypothesis that the mechanical properties of the capsid have been altered due to the packaging of the decompressed genome, thereby producing altered thermal stability and attachment deficiencies (Figure 3).

Curiously, we also observed *E. coli* chaperone complex GroEL structures mixed with φX174-containing CsCl bands (Figure S5). As many viruses rely on host cellular chaperones to mediate their protein folding,[81−83] including GroEL,[84−86] the copurification of this large protein complex is not surprising, but has never been associated with φX174-infections before. Moreover, we note that in the absence of highly purified particles, as GroEL

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

co-isolation suggests, structural details of internal capsid proteins J and H cannot be determined.

Targeted Proteomics Enables Time-Resolved Measurement of φX174 Protein Production. To identify whether there are any defects in phage protein production that could explain the decompressed φX174 reduced burst-size, we quantified the level of individual φX174 proteins produced during infection of the *E. coli* host. We infected *E. coli* C122(pJ804(Gene F)) with either wild-type or decompressed φX174 phage at MOI = 1.2 and captured samples over a time-course. Proteins from four post-infection time-points (0, 15, 25, and 35 min) were extracted and analyzed by parallel reaction monitoring mass spectrometry (PRM).[87] PRM is a targeted mass spectrometry acquisition method that enables highly sensitive and reproducible measurements of predetermined peptide ions,[87] and was therefore a well-suited analytical method for quantitation of phage proteins *in vivo*. This approach successfully detected and quantified the relative abundance of phage proteins from within the dynamic background of the *E. coli* proteome, enabling comparative analyses of the wild-type and decompressed φX174 phage proteomes across the infection cycle. Stoichiometry between distinct phage proteins (*e.g.*, H and J) could not be determined due to the nature of our relative measurements; however, interprotein differences between decompressed and wild-type still enabled us to draw robust conclusions toward expression perturbations. For quantitation, each phage protein was measured using at least two peptides (except for proteins K, J, and E, where only one tryptic peptide was suitable (File S1)).

Using PRM we measured viral proteins at infection initiation (time-point 0 min) for all proteins apart from A, C, and K (File S2). There are two possibilities for this observation: (1) the decompressed φX174 phage possessed some altered infection dynamics under the cold starvation conditions (Methods), or (2) contaminating noninfectious phage particles and phage protein debris carried over from the inoculation stock into the 0 min measurements. The first explanation is unlikely as incubation in starvation buffer and low temperature conditions (14 °C) will mitigate host transcription and translation, and φX174 genome penetration.[71,72,88] Contaminating noninfectious phage particles and debris are a more likely explanation as lower specific infectivity of decompressed virions are a strong possibility given the altered capsid properties, thus requiring more phage particles to be added for the requisite level of MOI. Contaminating nonstructural proteins A* and E can be explained given their association with membrane bound DNA (A*) or host membrane material (E) which may have been pelleted during cell harvesting. We therefore normalized our peptide abundance measurements by subtracting 0 min abundance from later abundance values.

Proteomics Reveals Decompressed φX174 Infection Is Deficient in Proteins B and C, while Protein A* Is Overexpressed. Next we determined whether any φX174 proteins (Table S1) produced during the infection cycle were altered between wild-type and decompressed φX174. Two-factor ANOVA analysis was performed on each peptide with independent variables set to time and phage strain. The results of this analysis showed there were significant alterations in the expression of proteins A*, B, and C, and possible alterations in H during decompressed φX174 infection (Figures 5, S6–S9, and Files S1 and S2). φX174 capsid protein F was supplied *in trans* for decompressed φX174 due to genome size limitations. We found that there was no significant difference in protein F expression for decompressed φX174 in comparison to wild-type samples, alleviating any concern for protein F deficiency as a cause of the phenotypic differences. Furthermore, the lysis protein E was only measured late in infection (35 min) for decompressed and wild-type samples (Figures 5, S6, and S7). This corroborates our earlier findings (Figure 2B) in which decompressed φX174 has a comparable lysis timing to that of wild-type at MOI = 5, and supports the idea that the temporal regulation of the lysis protein was not disrupted during modularization.

Significantly, we found protein C expression was diminished in the decompressed φX174 infection (Figures 5, S6, and S7). In the wild-type genome, gene C has significant levels of upstream sequence overlap with gene K as well as an overlapping start–stop codon junction with genes A/A* (Figure 6). The modularization of these junctions, resulting in a modified 5′-UTR region and removal of the A/C overlap may be responsible for decreased protein C expression.

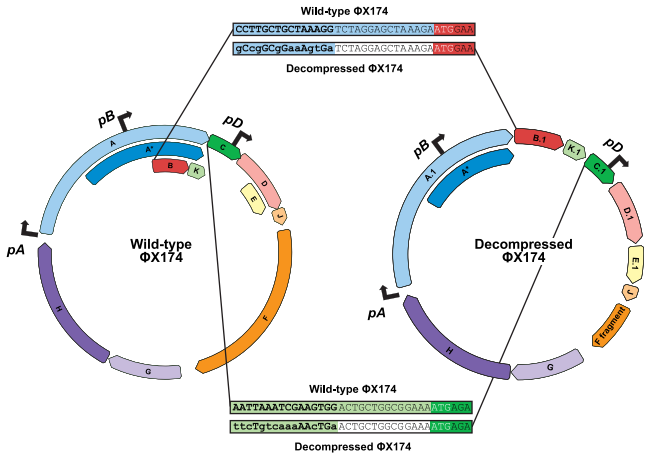


Figure 6. The heavily modified 5′-UTR of genes B and C in the decompressed φX174 genome may result in altered protein expression. Gene B and C sequences in the wild-type and decompressed φX174 genomes are highlighted (boxed) and show the region 5′ of the start codons (white bolded ATG text). Colored background to text indicates which gene the sequence belongs to: gene A (light blue), gene K (light green), gene C (dark green), and no gene (white). Wild-type gene C ATG start codon overlaps with gene A/A* TGA stop codon. Bolded text and lower case nucleotides show positions in the 5′-UTRs of decompressed φX174 genes that differ from wild-type sequence.

During φX174 infections, protein C is responsible for the shift from single-strand genome production seeding the generation of more dsDNA replicative forms (stage II) to the production of ssDNA genomes destined for packaging within capsids to assemble infectious virions (stage III).[89,90] More specifically, protein C functions by competitively binding to the φX174 origin of replication with an antagonistic partner, host protein single-strand binding protein (SSB).[90] If protein C binds, then a round of stage III DNA synthesis occurs in which the production of single-stranded genomic DNA occurs, followed by packaging into the viral procapsid.[89] If SSB protein binds, then stage II synthesis continues.[89] It is known that an increase in host SSB protein relative to protein C during stationary phase infection results in reduced plaque sizes.[90] The ∼6-fold decreased expression of protein C at 35 min postinfection likely results in a reduced ability to transition from stage II to stage III DNA synthesis and less efficient packaging of ssDNA genomes into

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079–3090

available procapsids. Our observation of reduced protein C may also explain the reduced burst and plaque sizes for decompressed φX174 (Figure 2A).

We also found that protein A* was significantly upregulated during decompressed φX174 infection. Although protein A and A* share 100% sequence identity, we were able to discriminate between protein A and A* production through the measurement of the unique A peptide, LLEEILSFR. By comparing LLEEILSFR levels to three other peptides shared by proteins A and A* we observed that protein A* is highly upregulated while protein A is not (Figures 5, S6, and S7).

Protein A* controls the delicate balance between packaging and replication during φX174 infections.[91] This function is accomplished through binding within the J−F intergenic region of the φX174 genome. This A* binding leads to inhibiting DNA unwinding in this region and disruption of protein A rolling circle genome amplification.[91−93] Therefore, stochiometry differences between proteins A and A* brought about by A* upregulation in decompressed φX174 are likely to cause issues with genome replication and the maintenance of genome packaging fidelity.

Protein B is an essential internal scaffolding protein that is needed for the assembly of the procapsid, an intermediary capsid particle devoid of the phage genome.[94−96] We also find significantly reduced internal scaffolding protein B production in the decompressed φX174 infection (Figures 5, S6, and S7), indicating another potential reason for reduced viral plaque and burst-size (Figure 2A). A lack of internal scaffolding protein B would prevent intermediary assembly particles, 9S* and 6S, from forming 12S* particles,[96] thereby stalling productive virion assembly.[96]

In the decompressed φX174 genome gene B has been moved from its nested position within genes A/A* to a fully separated and downstream position (Figure 6). Furthermore, the gene B 5′-UTR region is identical up to 15 nucleotides 5′ of the gene B start codon but further upstream is completely different (Figure 6) and may be a factor for its reduced protein output.

To understand possible reasons for the large differences in protein expression between decompressed and wild-type φX174, we analyzed mRNA folding across gene 5′ termini and entire coding sequences (Figure S10). Other work focused only on synonymous codon changes within φX174 has found correlations between gene mRNA folding energy and viral fitness.[97,98] By contrast, our analyses showed predictive mRNA structure does not explain the difference in protein production between the strains, further emphasizing our current inability to accurately predict the effects of large-scale genome architectural changes.

CONCLUSION

Re-engineering and recapitulating natural genes and gene pathways of interest is now enabling researchers in synthetic biology to create genomes with novel configurations never before seen in natural systems. However, our understanding of the intricacies of genome architecture, such as the functional importance of overlapping genes, is still lacking.

In this study, we have revealed the biological ramifications of whole genome modularization in the model virus φX174. For the first time, using the fully refactored φX174 and careful phenotypic and targeted proteomic measurements, we have conclusively demonstrated that removing coding overlap causes major defects in protein expression and reproduction efficiency. More specifically, we show that the decompressed φX174 infection cycle has significantly altered protein expression for 3 of the 11 (∼30%) phage proteins (A*, B, and C), as well as perturbed burst-size, capsid stability, and host-cell attachment. Further investigative work to determine decompressed protein stoichiometry, specific infectivity, and genome replication fidelity should lend deeper insights into decompressed φX174 phenotype deficiencies. This work provides key evidence that gene overlap is a crucial feature in genomes and points at a direct mechanism for their retention during evolutionary selection. Lastly, our work shows that in order to retain the highest fitness levels possible, conserving overlapped gene topologies should be strongly considered in genome engineering projects going forward.

METHODS

Strains. Phage strains φX174 and φX174.1f (termed decompressed φX174 throughout) were used throughout this work.[44,98] The host strain used for experiments and virus propagation (unless stated otherwise) was *E. coli* C122 (Public Health England NCTC122) which carried a plasmid pJ804-(Gene F) containing the φX174 gene F (Genbank No. NC_001422.1) under a rhamnose-inducible promoter.[44] Growth of the host strain was conducted at 37 °C, 250 rpm (with 25 mm orbital shaking) in LB Miller broth (10 g/L tryptone, 10 g/L NaCl, and 5 g/L yeast extract) containing 50 μg/mL carbenicillin and 2 mM $CaCl_2$ and either 2 mM rhamnose for decompressed φX174 or 2 mM glucose for wild-type φX174.

Phage Infection. Infections with either the wild-type or decompressed phage were conducted by first growing the host strain overnight (∼18 h) followed by subculturing the cells 1/100 in fresh LB broth (Miller) containing 50 μg/mL carbenicillin and 2 mM calcium chloride, and growing until an O.D. of 0.5−0.6. The cells were then pelleted (4000 RCF, 14 °C, 8 min) and subjected to two wash and pellet cycles with ice-cold HFB buffer (60 mM ammonium chloride, 90 mM sodium chloride, 100 mM potassium chloride, 1 mM magnesium sulfate, 1 mM calcium chloride, and 100 mM tris base, pH 7.4), after which the cell pellet was resuspended to 1/10 of the original culture volume using ice-cold HFB followed by addition of virus at required MOI. The cell−virus mixture was left at 14 °C for 30 min to allow viral attachment (infection synchronization).[99] Infection initiation was triggered by addition of prewarmed (37 °C) LB broth (Miller) containing 50 μg/mL carbenicillin and 2 mM $CaCl_2$ (and either 2 mM rhamnose or glucose) to a final volume constituting original culture volume.

Phage Isolation and Purification. Phage stocks were produced through two methods. Both involved growth of 800 mL of *E. coli* C122(pJ804(Gene F)) culture to mid log, as described. If propagation of decompressed φX174 was required, the media included 50 μg/mL carbenicillin and 2 mM rhamnose. Virus was spiked into the culture and allowed to propagate until culture clearance. The first method, involved pelleting of cleared cultured to remove bacterial debris (4000 RCF for 8 min). The supernatant was retained and filtered using Amicon Ultra-15 Centrifugal Filter Units (Merck, Germany) with a 100 kDa molecular weight cutoff. The second method was for ultrapure phage stocks. After culture clearance, NaCl was added to a final concentration of 1 M, cooled to 4 °C, followed by pelleting of bacterial debris (4000 RCF for 8 min) and retaining the supernatant and addition of poly(ethylene glycol) (PEG) 8000 to a final concentration of 15% (w/v). The supernatant was stored overnight (∼18 h) at 4 °C, followed by

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

pelleting the phage at 4000 RCF for 8 min. Phage pellets were resuspended in HFB buffer and PEG 8000 was removed using three wash cycles of equal volume chloroform. Finally, the phage was CsCl step-gradient purified using CsCl density steps of 1.30, 1.40, and 1.50 g/mL using the Beckman Coulter SW32Ti swinging bucket rotor at 32 000 rpm for 16 h at 10 °C. Phage bands were removed and buffer-exchanged using the Amicon Ultra-15 Centrifugal Filter Units (Merck, Germany) with a 100 kDa molecular weight cutoff in HFB buffer. The titer for viral stocks was determined using the double-agar overlay method.[100] Soft agar was supplemented with 50 μg/mL carbenicillin and 2 mM rhamnose for enumeration of the decompressed φX174 phage titer.

Genome Folding Prediction. Minimum free energy (MFE) secondary structure predictions for decompressed and wild-type φX174 were performed through the ViennaRNA websuite's RNAfold server.[101] Default settings were used, except for energy parameters which were set as DNA parameters.

mRNA Structure Predictions. RNA folding simulations were performed on an 83-nt window around the start codon of each gene using the Nucleic Acid Package (NUPACK) web server with default settings,[102] or using the entire coding sequence using mfold web server with default settings.[103] Lowest-energy structures were reported.

Bacteriophage Lysis Curve Measurements. Lysis curve for decompressed and wild-type φX174 was conducted in parallel 10 mL triplicate cultures, and grown and synchronized as previously described. 2 mM rhamnose was added to decompressed φX174 cultures to induce gene F on pJ804(Gene F), and 2 mM glucose to wild-type cultures to suppress the induction of the plasmid-borne gene F. Culture samples for O.D. measurements were taken every 10 min for 60 min. PCR was performed on cultures at the end of the experiment to confirm there was no cross-contamination of phage strains (Figure S1).

Bacteriophage Attachment. Cultures were grown as described, except, lysis resistant *E. coli* C900 (Δ*slyD*) which contains a mutation conferring resistance to φX174 E-mediated lysis[104] was used as the host cell for attachment (a gift from Bentley A. Fane). At mid log, cells were pelleted, washed and resuspended in microcentrifuge tubes to 1/10th of the original culture volume with HFB buffer that was either precooled for the 14 °C assay, or prewarmed for the 37 °C assay. Phage stock was then added to the cells, followed by incubation at the required temperature with continuous culture agitation at 900 rpm using a benchtop heating block with Eppendorf holding cassette. At 5 and 10 min post-incubation, cells were immediately pelleted at 4000 RCF for 5 min, followed by removal of the top 100 μL of the supernatant for assessment of viral titer using the double-agar overlay method.

Bacteriophage Progeny Production Measurements. Cultures and infection synchronization at 14 °C was performed as described, except the virus was added to a final MOI of 0.01. After synchronization, cells were pelleted and the supernatant was retained for titer. Cultures were then resuspended using prewarmed (37 °C) LB with 2 mM $CaCl_2$ and 50 μg/mL carbenicillin, and supplemented with either 2 mM glucose for the wild-type virus, or 2 mM rhamnose for the decompressed phage. Resuspended cultures were then diluted with media according to predetermined levels to allow for appropriate enumeration of plaques. Cultures were then incubated at 37 °C with 250 rpm agitation for 40 min with sampling of the culture at 5, 10, 15, 20, 25, 30, 35, and 40 min time-points. Sampling involved removing 200 μL of the culture and immediately mixing with 100 μL of chloroform, followed by storage on ice. Samples were then spun at 5000 RCF for 5 min, after which 100 μL of the top layer was removed and put into a fresh tube. Assessment of viral titer was then performed using the double-agar overlay method. To account for viral attachment differences, the PFU counts were normalized to the amount of phage attached by assaying the supernatant of the spin-down and subtracting the unattached phage titer from the phage titer at the start of the experiment. These values were then used to calculate burst size.

Virion Heat Stability. Phage stocks were diluted to approximately 10^2 PFU/μL using LB broth (Miller) supplemented with 10% (v/v) glycerol, and 2 mM calcium chloride. Eppendorf tubes containing 90 μL of the same LB broth diluent as well as 10 μL of the diluted phage stock. The tubes were individually labeled according to predetermined harvest points of 0, 5, 15, 30, and 45 min, with the 0 min tube being placed immediately on ice. The rest of the tubes were incubated at 60 °C, and then placed upon ice when reaching their time point. Titers of the samples was performed as described using the double-agar overlay method.

DNase I and Proteinase K Stability Assays. For both assays, individual plaques were stabbed from double-agar overlay plates and eluted into HFB buffer for the DNase I assay, or 10 mM Tris-HCl for the proteinase K assay. The DNase I assay consisted of aliquoting 7.5 μL of decompressed or wild-type plaque suspension into a 200 μL capacity microcentrifuge tubes containing 0.5 μL DNase I (2000 U/mL) (or water (control)), 1 μL DNase I buffer (10×) and incubated at 37 °C for 1-h. The proteinase K assay consisted of aliquoting 9 μL of plaque suspension into a 200 μL capacity microcentrifuge tube and either 1 μL of 1 μg/μL proteinase K, or water (control) and incubated at 37 °C of for 4 h. For both assays, plaques were enumerated through the double-agar overlay assay.[100]

Electron Microscopy. A 4 μL solution of decompressed or wild-type φX174 virions were applied to glow-discharged copper grids (carbon type B with Formvar, 200 mesh (Ted Pella Inc., United States)), then stained with 2% uranyl acetate and dried at room temperature. The screening of the capsids was carried on FEI Tecnai G2 20 electron microscope (ThermoFisher Scientific, United States) at 200 kV accelerating voltage. Data images for 2D classification of both capsids were automatically collected on a Talos Arctica transmission electron microscope operated at 200 kV acceleration voltage using EPU software (ThermoFisher Scientific, United States). All data sets were collected on a Falcon III detector in linear mode at 92 000× magnification with a pixel size of 1.6 Å per pixel.

Image Processing (2D Classification). MotionCor2[105] was used to correct local beam-induced motion and to align resulting frames. Micrographs were then sorted on the basis of image quality, and 1946 micrographs for wild-type and 1633 micrographs for decompressed φX174 were used for subsequent reconstruction using RELION 3.0 beta.[106−108] Particles (37 150 for wild-type and 2971 for decompressed φX174) were picked automatically from the full data set and extracted by using a box size of 300 × 300 pixels. The particles were then analyzed by 2D classification in RELION 3.0 beta.

Label-Free Quantitative Mass Spectrometry−Parallel Reaction Monitoring (PRM). *E. coli* C122(pJ804(Gene F)) was grown from overnight (~18 h) cultures to mid log phase in LB broth (Miller) supplemented with 2 mM $CaCl_2$ and 50 μg/mL carbenicillin. Cultures were then prepared for infection synchronization with either decompressed or wild-type φX174

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

strains at MOI = 1.2 (to the starved hosts, we added 230 μL of LB broth (Miller) with 2 mM $CaCl_2$ and 50 μg/mL carbenicillin containing phage at the appropriate PFU), as described earlier, after which, the infection was initiated by the addition of warm (37 °C) LB broth (Miller) supplemented with 2 mM $CaCl_2$, 50 μg/mL carbenicillin and either 2 mM rhamnose for the decompressed φX174 strain cultures, or glucose for the wild-type cell cultures. 0 min time-points were collected immediately (2.25 mL) and placed on ice and then the cultures were incubated at 37 °C, 250 rpm. At time-points 15, 25, and 35 min postinfection initiation, 2.25 mL of culture was harvested and stored on ice. Samples were pelleted, and washed twice with ice-cold HFB buffer, followed by storing the pellets at −30 °C overnight (∼18 h).

The next day, sample pellets were resuspended in 200 μL lysis buffer (100 mM triethylammonium bicarbonate buffer (TEAB) with 1% (v/v) sodium deoxycholate (SDC) and 1× protease inhibitor cocktail solution (Roche, Switzerland)), and subjected to probe sonication (30% amplitude, 5 bursts). Reduction and alkylation of cysteine residues (dithiothreitol (DTT) to 10 mM, incubated at 60 °C for 30 min followed by iodoacetamide (IAA) to 30 mM, incubated 1-h in the dark, and last, quenching with equal molarity of DTT), followed by chloroform−methanol protein precipitation to obtain pure protein pellets that were subsequently resuspended in digestion buffer (100 mM TEAB with 1% (v/v) SDC) and quantified by the Pierce BCA Protein Assay Kit (ThermoFisher Scientific, United States). From each sample, 30 μg of protein was taken for digestion with trypsin (2% (w/w)) (Promega, United States) and left overnight at 37 °C (∼18 h). After digestion, samples were acidified with 100% formic acid to 1% (v/v) and SDC was removed by spinning samples at 20 000 RCF and collecting the supernatant into a fresh tube. Samples were then vacuum centrifuged and dry peptides were stored at −30 °C until ready for mass spectrometry analysis.

Peptides were reconstituted in 2% (v/v) acetonitrile (ACN) and 0.1% (v/v) formic acid (buffer A) to a final concentration of 0.1 μg/μL. 1 μg (10 μL) of each sample was analyzed on a Q-Exactive mass spectrometer (ThermoFisher Scientific, United States) coupled to an EASY-nLC1000 system (ThermoFisher Scientific, United States). Peptide samples were injected onto the LC system using buffer A and were bound on a 75 μM × 100 mM C18 HALO column (2.7 μM bead size, 160 Å pore size). A flow rate of 300 nL/min using an increasing linear gradient of buffer B (99.9% (v/v) acetonitrile, 0.1% (v/v) formic acid) was run from 1% to 50% for 110 min followed by 85% buffer B for 10 min.

An inclusion list was used to target predefined precursor mass-to-charge (m/z) (File S3), and the mass spectrometer was operated to perform one full-ms scan (70 000 resolution (at 400 m/z) across the m/z range of 320−1800 m/z) with an automatic gains control (AGC) target of 1×10^6 (or a maximum fill time of 200 ms), followed by sequential PRM scans using the inclusion list (loop count 39, isolation window of 2 m/z). The selected precursor ions from PRM scans had an AGC target of 1×10^6 and a maximum fill time of 58 ms, after which, they were transferred from the C-trap to the higher energy collision dissociation (HCD) cell for fragmentation at a normalized collision energy of 27. MS/MS spectra were collected at a resolution of 17 500 (at 400 m/z).

Raw files were analyzed by MaxQuant version 1.6.10.43[109] against a φX174 protein sequence list generated from GenBank file NC_001422.1. Cysteine carbamidomethylation was selected as a fixed modification and protein and peptide sequence match identifications were set at a 1% false discovery rate cutoff. Result files were imported and analyzed in Skyline version 19.1.0.193,[110] and peptides with less than 3 matching fragment ions were filtered out. The summed area values of all individual transitions for each precursor (total area) were extracted after applying Savitzky−Golay smoothing and these values were used to quantify peptide relative abundance. The median total ion current (TIC) area across all samples was used for normalization by correcting for injection amount. The median TIC was divided by the sample's TIC and the ratio of this was used to normalize the total transition area. Average and standard deviation of the total area were calculated across biological triplicates, constituting relative peptide (and by extension, protein) abundance. Statistical analysis was performed individually on each peptide and significance level was established using single or two-factor ANOVA where appropriate (significance = p-value < 0.05). Fold-change was expressed using decompressed φX174 average intensity over wild-type φX174 average intensity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00323.

Figures S1−S10, Table S1, Supporting Methods: Representative sample purity test using PCR; computational prediction of 2D folding structure of wild-type and decompressed genomes; phage stability assays; cryo-EM; mass spectrometry spectra; N-terminal RNA folding energies; list of genes of φX174 (PDF)

Supporting Files S1−S3: File of detected MS peptides; PRM results for viral peptides; PRM inclusion list (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Mark P. Molloy − *Kolling Institute, Northern Clinical School, The University of Sydney, Sydney, NSW 2006, Australia;* orcid.org/0000-0003-4679-5868; Email: m.molloy@sydney.edu.au

Paul R. Jaschke − *Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia;* orcid.org/0000-0003-2359-1301; Email: paul.jaschke@mq.edu.au

Authors

Bradley W. Wright − *Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia;* orcid.org/0000-0002-6233-831X

Juanfang Ruan − *Electron Microscope Unit, Mark Wainwright Analytical Centre and School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW 2052, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00323

Author Contributions

BWW contributed conceptualization, formal analysis, investigation, methodology, visualization, writing (original draft), and writing (review and editing). JR contributed formal analysis, investigation, methodology, resources, validation, visualization, writing (original draft), and writing (review and editing). PRJ contributed conceptualization, funding acquisition, project administration, resources, validation, visualization, writing

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

(review and editing), and supervision. MPM contributed conceptualization, funding acquisition, resources, writing (review and editing), supervision.

Notes

The authors declare no competing financial interest.

Raw and processed proteomics data generated in this study are available in the Panorama Repository under https://panoramaweb.org/phix174.url and ProteomeXchange under PXD019681.

■ ACKNOWLEDGMENTS

We recognize that this research was conducted on the traditional lands of the Wattamattagal clan of the Darug nation. BWW was supported by a Macquarie Research Excellence Ph.D. Scholarship, PRJ was supported by the Molecular Sciences Department, Faculty of Science & Engineering, and the Deputy Vice-Chancellor (Research) of Macquarie University. JFR thanks the Cryo Electron Microscopy Facility through the Victor Chang Innovation Centre, funded by the NSW government, and the Electron Microscope Unit at UNSW, Sydney. We thank Dr. Pascal Steffen (The University of Sydney), Dr. Roy Walker (Macquarie University), and Prof. Bentley Fane (University of Arizona) for helpful discussions and feedback, Dr. Chao Shen (Macquarie University) for assistance with microscopy. Aspects of this research were conducted at the Australian Proteome Analysis Facility.

■ ABBREVIATIONS

dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; MOI, multiplicity of infection; PRM, parallel reaction monitoring; TEM, transmission electron microscopy; UTR, untranslated region.

■ REFERENCES

(1) Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., and Lenhard, B. (2017) Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun. 8*, 441.

(2) Tsochatzidou, M., Malliarou, M., Papanikolaou, N., Roca, J., and Nikolaou, C. (2017) Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of Saccharomyces cerevisiae. *Nucleic Acids Res. 45*, 5818−5828.

(3) Chen, N., and Stein, L. D. (2006) Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans. *Genome Res. 16*, 606−617.

(4) Lawrence, J. G. (2002) Shared Strategies in Gene Organization among Prokaryotes and Eukaryotes. *Cell 110*, 407−413.

(5) Lim, H. N., Lee, Y., and Hussein, R. (2011) Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci. U. S. A. 108*, 10626.

(6) Cavalli, G., and Misteli, T. (2013) Functional implications of genome topology. *Nat. Struct. Mol. Biol. 20*, 290−299.

(7) Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet. 8*, 104−115.

(8) Tamames, J. (2001) Evolution of gene order conservation in prokaryotes. *Genome Biol. 2*, research0020.1.

(9) Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci. 23*, 324−328.

(10) Davila Lopez, M., Martinez Guerra, J. J., and Samuelsson, T. (2010) Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. *PLoS One 5*, No. e10654.

(11) Poyatos, J. F., and Hurst, L. D. (2007) The determinants of gene order conservation in yeasts. *Genome Biol. 8*, R233.

(12) Singer, G. A. C., Lloyd, A. T., Huminiecki, L. B., and Wolfe, K. H. (2005) Clusters of Co-expressed Genes in Mammalian Genomes Are Conserved by Natural Selection. *Mol. Biol. Evol. 22*, 767−775.

(13) Minot, S., Wu, G. D., Lewis, J. D., and Bushman, F. D. (2012) Conservation of gene cassettes among diverse viruses of the human gut. *PLoS One 7*, No. e42342.

(14) Herniou, E. A., Olszewski, J. A., Cory, J. S., and O'Reilly, D. R. (2003) The Genome Sequence and Evolution of Baculoviruses. *Annu. Rev. Entomol. 48*, 211−234.

(15) Wakimoto, B. T., and Hearn, M. G. (1990) The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. *Genetics 125*, 141−154.

(16) Endy, D., You, L., Yin, J., and Molineux, I. J. (2000) Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. U. S. A. 97*, 5375.

(17) Flanagan, E. B., Schoeb, T. R., and Wertz, G. W. (2003) Vesicular stomatitis viruses with rearranged genomes have altered invasiveness and neuropathogenesis in mice. *J. Virol. 77*, 5740−5748.

(18) Pesko, K., Voigt, E. A., Swick, A., Morley, V. J., Timm, C., Yin, J., and Turner, P. E. (2015) Genome rearrangement affects RNA virus adaptability on prostate cancer cells. *Front. Genet.*, DOI: 10.3389/fgene.2015.00121.

(19) Springman, R., Badgett, M. R., Molineux, I. J., and Bull, J. J. (2005) Gene order constrains adaptation in bacteriophage T7. *Virology 341*, 141−152.

(20) Cecchini, N., Schmerer, M., Molineux, I. J., Springman, R., and Bull, J. J. (2013) Evolutionarily stable attenuation by genome rearrangement in a virus. *G3: Genes, Genomes, Genet. 3*, 1389−1397.

(21) Belshaw, R., Pybus, O. G., and Rambaut, A. (2007) The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res. 17*, 1496.

(22) Brandes, N., and Linial, M. (2016) Gene overlapping and size constraints in the viral world. *Biol. Direct 11*, 26.

(23) Johnson, Z. I., and Chisholm, S. W. (2004) Properties of overlapping genes are conserved across microbial genomes. *Genome Res. 14*, 2268−2272.

(24) Huvet, M., and Stumpf, M. P. H. (2014) Overlapping genes: a window on gene evolvability. *BMC Genomics 15*, 721−721.

(25) Chung, W.-Y., Wadhawan, S., Szklarczyk, R., Pond, S. K., and Nekrutenko, A. (2007) A First Look at ARFome: Dual-Coding Genes in Mammalian Genomes. *PLoS Comput. Biol. 3*, No. e91.

(26) Neme, R., and Tautz, D. (2013) Phylogenetic patterns of emergence of new genes support a model of frequent de novoevolution. *BMC Genomics 14*, 117.

(27) Mouilleron, H., Delcourt, V., and Roucou, X. (2016) Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Res. 44*, 14−23.

(28) Schlub, T. E., and Holmes, E. C. (2020) Properties and abundance of overlapping genes in viruses. *Virus Evol.*, DOI: 10.1093/ve/veaa009.

(29) Pavesi, A., Vianelli, A., Chirico, N., Bao, Y., Blinkova, O., Belshaw, R., Firth, A., and Karlin, D. (2018) Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS One 13*, No. e0202513.

(30) Veeramachaneni, V., Makałowski, W., Galdzicki, M., Sood, R., and Makałowska, I. (2004) Mammalian overlapping genes: the comparative perspective. *Genome Res. 14*, 280−286.

(31) Sanna, C. R., Li, W.-H., and Zhang, L. (2008) Overlapping genes in the human and mouse genomes. *BMC Genomics 9*, 169.

(32) Sun, H., Yang, S., Tun, L., and Li, Y. (2015) IAOseq: inferring abundance of overlapping genes using RNA-seq data. *BMC Bioinf. 16* (Suppl 1), S3−S3.

(33) Chirico, N., Vianelli, A., and Belshaw, R. (2010) Why genes overlap in viruses. *Proc. R. Soc. London, Ser. B 277*, 3809−3817.

(34) Jin, H., Vacic, V., Girke, T., Lonardi, S., and Zhu, J.-K. (2008) Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis. *BMC Mol. Biol. 9*, 6.
(35) Lapidot, M., and Pilpel, Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep. 7*, 1216−1222.
(36) Krakauer, D. C. (2000) Stability and evolution of overlapping genes. *Evolution 54*, 731−739.
(37) Delaye, L., DeLuna, A., Lazcano, A., and Becerra, A. (2008) The origin of a novel gene through overprinting in Escherichia coli. *BMC Evol. Biol. 8*, 31.
(38) Van Oss, S. B., and Carvunis, A.-R. (2019) De novo gene birth. *PLoS Genet. 15*, No. e1008160.
(39) Keese, P. K., and Gibbs, A. (1992) Origins of genes: "big bang" or continuous creation? *Proc. Natl. Acad. Sci. U. S. A. 89*, 9489.
(40) Logel, D. Y., and Jaschke, P. R. (2020) A high-resolution map of bacteriophage ϕX174 transcription. *Virology 547*, 47−56.
(41) Brantl, S. (2002) Antisense-RNA regulation and RNA interference. *Biochim. Biophys. Acta, Gene Struct. Expression 1575*, 15−25.
(42) Temme, K., Zhao, D., and Voigt, C. A. (2012) Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc. Natl. Acad. Sci. U. S. A. 109*, 7085−7090.
(43) Li, G.-q., Ma, T., Li, S.-s., Li, H., Liang, F.-l., and Liu, R.-l. (2007) Improvement of Dibenzothiophene Desulfurization Activity by Removing the Gene Overlap in the dsz Operon. *Biosci., Biotechnol., Biochem. 71*, 849−854.
(44) Jaschke, P. R., Lieberman, E. K., Rodriguez, J., Sierra, A., and Endy, D. (2012) A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast. *Virology 434*, 278−284.
(45) Chan, L. Y., Kosuri, S., and Endy, D. (2005) Refactoring bacteriophage T7. *Mol. Syst. Biol. 1*, 2005.0018.
(46) Ghosh, D., Kohli, A. G., Moser, F., Endy, D., and Belcher, A. M. (2012) Refactored M13 bacteriophage as a platform for tumor cell imaging and drug delivery. *ACS Synth. Biol. 1*, 576−582.
(47) Fernandes, J. D., Faust, T. B., Strauli, N. B., Smith, C., Crosby, D. C., Nakamura, R. L., Hernandez, R. D., and Frankel, A. D. (2016) Functional Segregation of Overlapping Genes in HIV. *Cell 167*, 1762−1773.
(48) Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D., and Church, G. M. (2013) Probing the limits of genetic recoding in essential genes. *Science 342*, 361−363.
(49) Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G., and Venter, J. C. (2016) Design and synthesis of a minimal bacterial genome. *Science 351*, No. aad6253.
(50) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006.0008.
(51) Gallet, R., Kannoly, S., and Wang, I.-N. (2011) Effects of bacteriophage traits on plaque formation. *BMC Microbiol. 11*, 181.
(52) Abedon, S. T., and Culler, R. R. (2007) Optimizing bacteriophage plaque fecundity. *J. Theor. Biol. 249*, 582−592.
(53) Abedon, S. T., and Culler, R. R. (2007) Bacteriophage evolution given spatial constraint. *J. Theor. Biol. 248*, 111−119.
(54) Zheng, Y., Struck, D. K., and Young, R. (2009) Purification and Functional Characterization of ϕX174 Lysis Protein E. *Biochemistry 48*, 4999−5006.
(55) Hutchison, C. A., and Sinsheimer, R. L. (1966) The process of infection with bacteriophage ΦX174: X. Mutations in a ΦX lysis gene. *J. Mol. Biol. 18*, 429−IN422.
(56) Gillam, S., Atkinson, T., Markham, A., and Smith, M. (1985) Gene K of bacteriophage phi X174 codes for a protein which affects the burst size of phage production. *J. Virol. 53*, 708−709.
(57) Brown, C. J., Stancik, A. D., Roychoudhury, P., and Krone, S. M. (2013) Adaptive regulatory substitutions affect multiple stages in the life cycle of the bacteriophage ϕX174. *BMC Evol. Biol. 13*, 66.
(58) Saha, B., Wong, C. M., and Parks, R. J. (2014) The adenovirus genome contributes to the structural stability of the virion. *Viruses 6*, 3563−3583.
(59) Emerson, S. U., Arankalle, V. A., and Purcell, R. H. (2005) Thermal Stability of Hepatitis E Virus. *J. Infect. Dis. 192*, 930−933.
(60) Ojosnegros, S., García-Arriaza, J., Escarmís, C., Manrubia, S. C., Perales, C., Arias, A., Mateu, M. G., and Domingo, E. (2011) Viral Genome Segmentation Can Result from a Trade-Off between Genetic Content and Particle Stability. *PLoS Genet. 7*, No. e1001344.
(61) Ivanovska, I., Wuite, G., Jönsson, B., and Evilevitch, A. (2007) Internal DNA pressure modifies stability of WT phage. *Proc. Natl. Acad. Sci. U. S. A. 104*, 9603−9608.
(62) Roos, W. H., Bruinsma, R., and Wuite, G. J. L. (2010) Physical virology. *Nat. Phys. 6*, 733−743.
(63) Evilevitch, A., Lavelle, L., Knobler, C. M., Raspaud, E., and Gelbart, W. M. (2003) Osmotic pressure inhibition of DNA ejection from phage. *Proc. Natl. Acad. Sci. U. S. A. 100*, 9292.
(64) Evilevitch, A., Fang, L. T., Yoffe, A. M., Castelnovo, M., Rau, D. C., Parsegian, V. A., Gelbart, W. M., and Knobler, C. M. (2008) Effects of Salt Concentrations and Bending Energy on the Extent of Ejection of Phage Genomes. *Biophys. J. 94*, 1110−1120.
(65) Belyi, V. A., and Muthukumar, M. (2006) Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. U. S. A. 103*, 17174−17178.
(66) Šiber, A., Božič, A. L., and Podgornik, R. (2012) Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Phys. Chem. Chem. Phys. 14*, 3746−3765.
(67) Erdemci-Tandogan, G., Wagner, J., van der Schoot, P., Podgornik, R., and Zandi, R. (2014) RNA topology remolds electrostatic stabilization of viruses. *Phys, Rev. E 89*, 032707.
(68) Snijder, J., Uetrecht, C., Rose, R. J., Sanchez-Eugenia, R., Marti, G. A., Agirre, J., Guérin, D. M. A., Wuite, G. J. L., Heck, A. J. R., and Roos, W. H. (2013) Probing the biophysical interplay between a viral genome and its capsid. *Nat. Chem. 5*, 502−509.
(69) Mateo, R., Luna, E., Rincón, V., and Mateu, M. G. (2008) Engineering Viable Foot-and-Mouth Disease Viruses with Increased Thermostability as a Step in the Development of Improved Vaccines. *J. Virol. 82*, 12232−12240.
(70) Bauer, D. W., Li, D., Huffman, J., Homa, F. L., Wilson, K., Leavitt, J. C., Casjens, S. R., Baines, J., and Evilevitch, A. (2015) Exploring the Balance between DNA Pressure and Capsid Stability in Herpesviruses and Phages. *J. Virol. 89*, 9288−9298.
(71) Newbold, J. E., and Sinsheimer, R. L. (1970) The process of infection with bacteriophage φX174: XXXII. Early steps in the infection process: Attachment, eclipse and DNA penetration. *J. Mol. Biol. 49*, 49−66.
(72) Denhardt, D. T., and Sinsheimer, R. L. (1965) The process of infection with bacteriophage φX174: III. Phage maturation and lysis after synchronized infection. *J. Mol. Biol. 12*, 641−646.
(73) Zeng, C., Moller-Tank, S., Asokan, A., and Dragnea, B. (2017) Probing the Link among Genomic Cargo, Contact Mechanics, and Nanoindentation in Recombinant Adeno-Associated Virus 2. *J. Phys. Chem. B 121*, 1843−1853.
(74) Zeng, C., Hernando-Perez, M., Dragnea, B., Ma, X., van der Schoot, P., and Zandi, R. (2017) Contact Mechanics of a Small Icosahedral Virus. *Phys. Rev. Lett. 119*, 038102.
(75) Bernal, R. A., Hafenstein, S., Esmeralda, R., Fane, B. A., and Rossmann, M. G. (2004) The phiX174 protein J mediates DNA packaging and viral attachment to host cells. *J. Mol. Biol. 337*, 1109−1122.
(76) Hafenstein, S., and Fane, B. A. (2002) phi X174 genome-capsid interactions influence the biophysical properties of the virion: evidence for a scaffolding-like function for the genome during the final stages of morphogenesis. *J. Virol. 76*, 5350−5356.
(77) Hafenstein, S. L., Chen, M., and Fane, B. A. (2004) Genetic and functional analyses of the oX174 DNA binding protein: the effects of

https://dx.doi.org/10.1021/acssynbio.0c00323
ACS Synth. Biol. 2020, 9, 3079−3090

substitutions for amino acid residues that spatially organize the two DNA binding domains. *Virology 318*, 204−213.
(78) Kazumori, Y. (1981) Electron microscopic studies of bacteriophage phi X174 intact and "eclipsing' particles, and the genome by the staining, and shadowing method. *J. Virol. Methods 2*, 159−167.
(79) McKenna, R., Xia, D., Willingmann, P., Ilag, L. L., Krishnaswamy, S., Rossmann, M. G., Olson, N. H., Baker, T. S., and Incardona, N. L. (1992) Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications. *Nature 355*, 137−143.
(80) Sun, Y., Roznowski, A. P., Tokuda, J. M., Klose, T., Mauney, A., Pollack, L., Fane, B. A., and Rossmann, M. G. (2017) Structural changes of tailless bacteriophage ΦX174 during penetration of bacterial cell walls. *Proc. Natl. Acad. Sci. U. S. A. 114*, 13708.
(81) Mayer, M. (2005) Recruitment of Hsp70 chaperones: a crucial part of viral survival strategies. *Rev. Physiol. Biochem. Pharmacol. 153*, 1−46.
(82) Geller, R., Taguwa, S., and Frydman, J. (2012) Broad action of Hsp90 as a host chaperone required for viral replication. *Biochim. Biophys. Acta, Mol. Cell Res. 1823*, 698−706.
(83) Xiao, A., Wong, J., and Luo, H. (2010) Viral interaction with molecular chaperones: role in regulating viral infection. *Arch. Virol. 155*, 1021−1031.
(84) Hanninen, A. L., Bamford, D. H., and Bamford, J. K. (1997) Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. *Virology 227*, 207−210.
(85) Bouvaine, S., Boonham, N., and Douglas, A. E. (2011) Interactions between a luteovirus and the GroEL chaperonin protein of the symbiotic bacterium Buchnera aphidicola of aphids. *J. Gen. Virol. 92*, 1467−1474.
(86) Ding, Y.-h., Duda, R. L., Hendrix, R. W., and Rosenberg, J. M. (1995) Complexes between chaperonin GroEL and the capsid protein of bacteriophage HK97. *Biochemistry 34*, 14918−14931.
(87) Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics 11*, 1475−1488.
(88) Newbold, J. E., and Sinsheimer, R. L. (1970) The process of infection with bacteriophage φX174: XXXI. Abortive infection at low temperatures. *J. Mol. Biol. 49*, 23−47.
(89) Aoyama, A., and Hayashi, M. (1986) Synthesis of bacteriophage ϕX174 in vitro: Mechanism of switch from DNA replication to DNA packaging. *Cell 47*, 99−106.
(90) Doore, S. M., Baird, C. D., Roznowski, A. P., and Fane, B. A. (2014) The Evolution of Genes within Genes and the Control of DNA Replication in Microviruses. *Mol. Biol. Evol. 31*, 1421−1431.
(91) Roznowski, A. P., Doore, S. M., Kemp, S. Z., and Fane, B. A. (2020) Finally, a role befitting Astar: the strongly conserved, unessential microvirus A* proteins ensure the product fidelity of packaging reactions. *J. Virol.*, JVI.01593-19 DOI: 10.1128/JVI.01593-19.
(92) van Mansfeld, A. D., van Teeffelen, H. A., Fluit, A. C., Baas, P. D., and Jansz, H. S. (1986) Effect of SSB protein on cleavage of single-stranded DNA by phi X gene A protein and A* protein. *Nucleic Acids Res. 14*, 1845−1861.
(93) Eisenberg, S., and Ascarelli, R. (1981) The A* protein of phi X174 is an inhibitor of DNA replication. *Nucleic Acids Res. 9*, 1991−2002.
(94) Siden, E. J., and Hayashi, M. (1974) Role of the gene B product in bacteriophage φX174 development. *J. Mol. Biol. 89*, 1−16.
(95) Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A., and Rossmann, M. G. (1997) Structure of a viral procapsid with molecular scaffolding. *Nature 389*, 308−313.
(96) Novak, C. R., and Fane, B. A. (2004) The functions of the N terminus of the phiX174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system. *J. Mol. Biol. 335*, 383−390.
(97) Van Leuven, J. T., Ederer, M. M., Burleigh, K., Scott, L., Hughes, R. A., Codrea, V., Ellington, A. D., Wichman, H., and Miller, C. (2020) ΦX174 Attenuation by Whole Genome Codon Deoptimization. *bioRxiv*, Feb 11, 2020, DOI: 10.1101/2020.02.10.942847 accessed 2020-02-18.
(98) Jaschke, P. R., Dotson, G. A., Hung, K. S., Liu, D., and Endy, D. (2019) Definitive demonstration by synthesis of genome annotation completeness. *Proc. Natl. Acad. Sci. U. S. A. 116*, 24206.
(99) Newbold, J. E., and Sinsheimer, R. L. (1970) Process of infection with bacteriophage phi-X174. XXXIV. Kinetic of the attachment and eclipse steps of the infection. *J. Virol. 5*, 427−431.
(100) Kropinski, A. M., Mazzocco, A., Waddell, T. E., Lingohr, E., and Johnson, R. P. (2009) Enumeration of bacteriophages by double agar overlay plaque assay. *Methods Mol. Biol. 501*, 69−76.
(101) Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol. 6*, 26.
(102) Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011) NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem. 32*, 170−173.
(103) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406−3415.
(104) Roof, W. D., Horne, S. M., Young, K. D., and Young, R. (1994) slyD, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. *J. Biol. Chem. 269*, 2902−2910.
(105) Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods 14*, 331−332.
(106) Scheres, S. H. W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol. 180*, 519−530.
(107) Fernandez-Leiro, R., and Scheres, S. H. W. (2017) A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D Struct Biol. 73*, 496−502.
(108) Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., and Scheres, S. H. W. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife 7*, No. e42166.
(109) Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol. 26*, 1367−1372.
(110) MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics 26*, 966−968.