Next-generation interaction proteomics for quantitative Jumbophage-bacteria interaction mapping

Andrea Fossati[1,2,3], Deepto Mozumdar[4], Claire Kokontis[4], Eliza Nieweglowska[4], Melissa Mendez[4], Adrian Pelin[1,2,3], Yuping Li[4], Baron Guo[4], Nevan J. Krogan[1,2,3], David A. Agard[4], Joseph Bondy-Denomy[4*] and Danielle L. Swaney[1,2,3*]

[1]J. David Gladstone Institutes, San Francisco, 94158, California, USA.
[2]Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, 94158, California, USA.
[3]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, 94158, California, USA.
[4]Department of Immunology and Microbiology, University of California San Francisco, San Francisco, 94158, California, USA.

*Corresponding author(s). E-mail(s): Joseph.Bondy-Denomy@ucsf.edu; danielle.swaney@ucsf.edu;
Contributing authors: andrea.fossati@gladstone.ucsf.edu; deepto.mozumdar@ucsf.edu; claire.kokontis@ucsf.edu; eliza@msg.ucsf.edu; melissa.mendez@ucsf.edu; adrian.pelin@ucsf.edu; yuping.li@ucsf.edu; baron.guo@ucsf.edu; nevan.krogan@ucsf.edu; agard@msg.ucsf.edu;

Abstract

Host-pathogen interactions (HPIs) are pivotal in regulating establishment, progression, and outcome of an infection. Affinity-purification mass spectrometry has become instrumental for the characterization of HPIs, however the targeted nature of exogenously expressing individual viral proteins has limited its utility to the analysis of relatively small pathogens. Here we present the use of co-fractionation mass spectrometry (SEC-MS) for the high-throughput analysis of HPIs from native viral infections of two jumbophages (ϕKZ and ϕPA3) in *Pseudomonas aeruginosa*. This enabled the detection >6000 unique host-pathogen and >200 pathogen-pathogen interactions for each phage, encompassing >50% of the phage proteome. Interactome-wide comparison across phages showed similar perturbed protein interactions suggesting fundamentally conserved mechanisms of phage predation within the KZ-like phage family. Prediction of novel ORFs revealed a ϕPA3 complex showing strong structural and sequence similarity to ϕKZ nvRNAp, suggesting ϕPA3 also possesses two RNA polymerases acting at different stages of the infection cycle. We further expanded our understanding on the molecular organization of the injected phage proteome by providing 23 novel virion components and 5 novel injected proteins, as well as providing the first evidence for phage manipulation of the host ribosome. To enable accessibility to this data, we developed PhageMAP, an online resource for network query, visualization, and interaction prediction http://phagemap.ucsf.edu/. We anticipate this study will lay the foundation for the application of co-fractionation mass spectrometry for the scalable profiling of host-pathogen interactomes and protein complex dynamics upon infection.

Introduction

Protein-protein interactions (PPIs) are the fundamental building blocks of cellular complexity and their perturbation and rewiring has profound effects on the proteome and cell fate. During an infection, the interactions between host and pathogen proteome are pivotal in regulating pathogen tropism, infection progression and, ultimately, infection outcome. Host-pathogen interaction (HPI) mapping using affinity-purification mass spectrometry (AP-MS) has been instrumental in identifying host targeted processes[1–5] and, recently, to predict potential therapeutics targets during the SARS-CoV2 pandemic[6–8].

Despite the successes of AP-MS for mapping HPIs, the exogenous expression and purification of individual pathogen proteins limits our our ability to characterize HPIs under native expression levels, and quantify how these interactions are regulated in the context of the full pathogen protein repertoire during an infection. The targeted nature of AP-MS also precludes the detection of downstream rearrangements in protein complexes beyond the viral protein of interest. Lastly, AP-MS is a labor-intensive process that requires the generation of numerous plasmids and hundreds or thousands of individual purifications to comprehensively probe protein-protein interactions for an entire viral proteome. This limits the scalability of AP-MS for the characterization HPIs for larger viruses or bacteria which express hundreds or thousands of proteins.

As a result, small eukaryotic viruses have been prioritized in host-pathogen interaction studies, thus extensive knowledge on interactions between larger prokaryotic viruses (bacteriophages) and their host is currently missing. This class of bacterial viruses hold great potential for treatment of multi-drug resistant bacteria which have increasingly been reported in the last two decades[9]. However, without a thorough understanding on putative interactions and functions of the phage gene products, it will be challenging to inform the rational design of the next generation of phage therapeutics.

To bridge this gap, here we have applied co-fractionation mass spectrometry using size-exclusion chromatography, coupled with fast data-independent acquisition MS (SEC DIA- MS),[10] to generate two phage-bacteria interactomes and to measure host PPI rewiring upon phage infection in *Pseudomonas aeruginosa*. Specifically, we provide the first interactome of two KZ-like phages

(ϕKZ[11] and ϕPA3[12]) which are archetype Jumbophages that possess gigantic genomes (>200 genes), with no known organization of genes by function, hence lacking synteny. Unique to this family of phages is the presence of a large proteinaceous shell acting analogous to the eukaryotic nucleus; thus decoupling transcription from translation. This structure confers resistance to several bacterial antiphage systems such as CRISPR[13, 14] and has a fundamental role in infection establishment[15] and virion egress[16]. Through the prediction of PPIs using deep learning and structural modeling, we derived system-level maps of Jumbophage infection encompassing a large fraction of the phage and bacterial host proteome. These host-pathogen interaction maps substantially extends previous knowledge on Jumbophage predation and, provide the first application of co-fractionation mass spectrometry for host-pathogen interaction profiling.

Results

A cross-phage study of viral infection cycle

To understands HPIs that mediate phage infection, we infected *Pseudomonas Aeruginosa* (strain PAO1) with either the ϕKZ or ϕPA3 bacteriophage for 60 minutes in biological duplicate. To control for virion protein complexes (i.e complexes present within the phage itself), parallel experiments were also performed using a naturally emerging PAO1 mutant (Δ*fliC*)[17] that cannot be infected (Fig.1A).

Infected cell lysates were then fractionated by size-exclusion chromatography, and each fraction (n=72) was analyzed using data independent acquisition MS (DIA-MS) coupled to high-throughput liquid chromatography[18]. To predict host-pathogen interactions, we used a modified version of the PCprophet

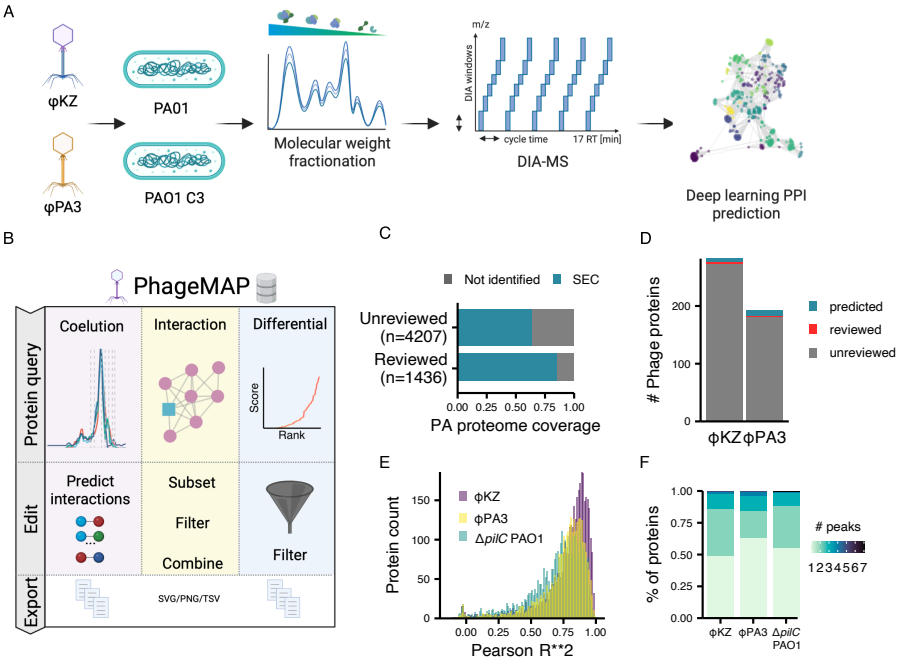


Fig. 1. High-throughput interaction proteomics for deep host-pathogen interaction mapping A. SEC-MS workflow and experimental design. **B**. Overview of PhageMAP analysis and workflows. **C**. Recovery of *Pseudomonas* proteome by SEC-MS. **D**. Barplot representing the number of phage proteins identified. **E**. Correlation between replicates for all proteins identified in the experiment (n=4132). **F**. Fractional distribution of the number of SEC peaks across the various phages and host.

toolkit[10], where the random forest classifier was replaced with a convolutional neural network that was trained for PPI prediction using > 10 million interactions from various co-fractionation experiments[19].

Derived host-pathogen interaction networks have been organized into a user-friendly website, PhageMAP, where users can query proteins of interest to visualize coelution patterns, interactomes, investigate different assembly states of the PAO1 proteome upon phage infection, and export their findings as publication-quality networks or coelution plots (Fig.1B).

This experimental workflow resulted in the high-throughput and comprehensive coverage of both the bacterial and the phage proteomes. Specifically, we detected 3782 PAO1 proteins, covering 83% of the validated SwissProt entries for the *Pseudomonas* pan-proteome, and 67% of the unreviewed entries (Fig.1B). Likewise, we detected 280 proteins for ϕKZ and 198 proteins for ϕPA3, covering 75% and 53% of their proteomes, respectively (Fig.1C).

To test the achievable robustness and resolution of our workflow we utilized two benchmarks. First, the robustness of fractionation was assessed by the Pearson R^2 between the two replicates of a given condition. Each condition showed an average correlation of >0.8 (Fig.1E), indicating high reproducibilty in both phage infection and SEC fractionation, with most of the SEC-profile peaks overlapping within 1-2 fractions (<0.250 μL). To test the resolution achievable with our chromatographic separations, we calculated the number of SEC peaks per proteins, which is a direct proxy for how many different complex assemblies a protein participates in. Approximately 45% of the identified proteins were detected in a single SEC-peak in each condition employed (Fig1F). While the presence of a single peak can represent detection of only a monomeric protein, we found the majority of these single-peaks proteins are not at their predicted monomeric molecular weight (Sup. Fig.S1). This suggests that the protein complex assembly state of the PAO1 proteome was preserved during sample preparation and SEC fractionation.

An high-quality interaction dataset for bacterial protein complexes

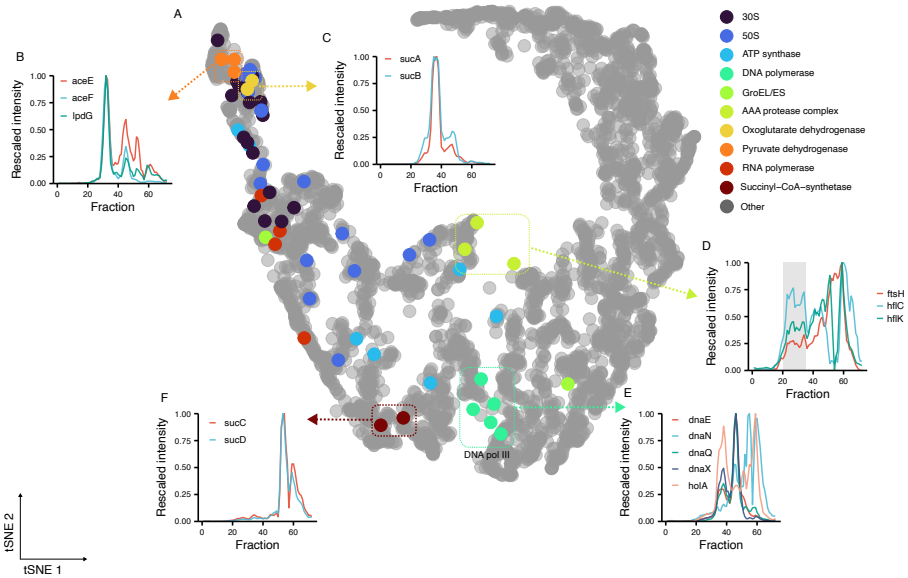


Fig. 2. *Pseudomonas* protein complexes identified in the SEC-MS data **A**. t-SNE plot for all the *P. aeruginosa* proteins detected in the Δ*fliC* experiment. Color code represents membership in reported protein complexes. Representative coelutions are showed for the pyruvate dehydrogenase complex (**B**), Oxoglutarate dehydrogenase complex (**C**), AAA protease complex (**D**), DNA polymerase III (**E**) and succinyl-coA synthetase (**F**). X axis shows the fraction number, while Y axis indicates the unit-rescaled intensity. The molecular weight (Kda) of the protein standard mix is represented as additional X axis (top). Line color shows the various subunits.

Next, we sought to investigate the recovery of known protein complexes by leveraging the partial conservation of core molecular assemblies between *P. aeruginosa* and other bacteria such as *E.coli*, for which protein complexes are more extensively annotated[20]. Smaller enzymes such as metabolic enzymes are usually co-expressed within the same operon[21] and have been reported to dimerize or multimerize. In line with this, we observed enzymes such as the

pyruvate dehydrogenase complex (Fig.2B) and the oxoglutarate dehydrogenase complex (Fig.2C), which migrated at an estimated MW of $\approx 3.5 * 10^6$ Da (expected MW $\approx 3.75 * 10^6$ Da) and $\approx 2.4 * 10^6$ Da, respectively. It is important to point that out that the molecular weight estimation for these large assemblies is subject to error due to these peaks being outside the external calibration curve. To achieve MW estimation we included in the calibration curve a pure SEC-separated 70S ribosome (Supplementary Fig S2).

Our sample preparation also preserved membrane-bound complexes. As example, the AAA protease complex, formed by four heaxamers of the AAA protease (ftsH) and 12 copies of each single-pass membrane proteins (HflK and HflC)[22], was recovered at high molecular weight in a broad peak as shown in Fig.2D. The large molecular weight range and sensitivity covered by our separation approach was also demonstrated in the recovery of more transient complexes such as the DNA polymerase III (dnaA, dnaE, and dnaQ) loaded with the γ complex (holA and dnaX) which plays a key role at the replication fork[23] (Fig.2E). Finally, heterodimeric complexes such as the succinyl-coA synthetase were also recovered as demonstrated by the coelution plot in (Fig.2F). Our manual inspections further confirms that prior knowledge can be easily incorporated into SEC-MS data analysis and allows for straight-forward identification of protein complexes.

Comparison of host-targeted processes reveals conserved and divergent predation mechanisms

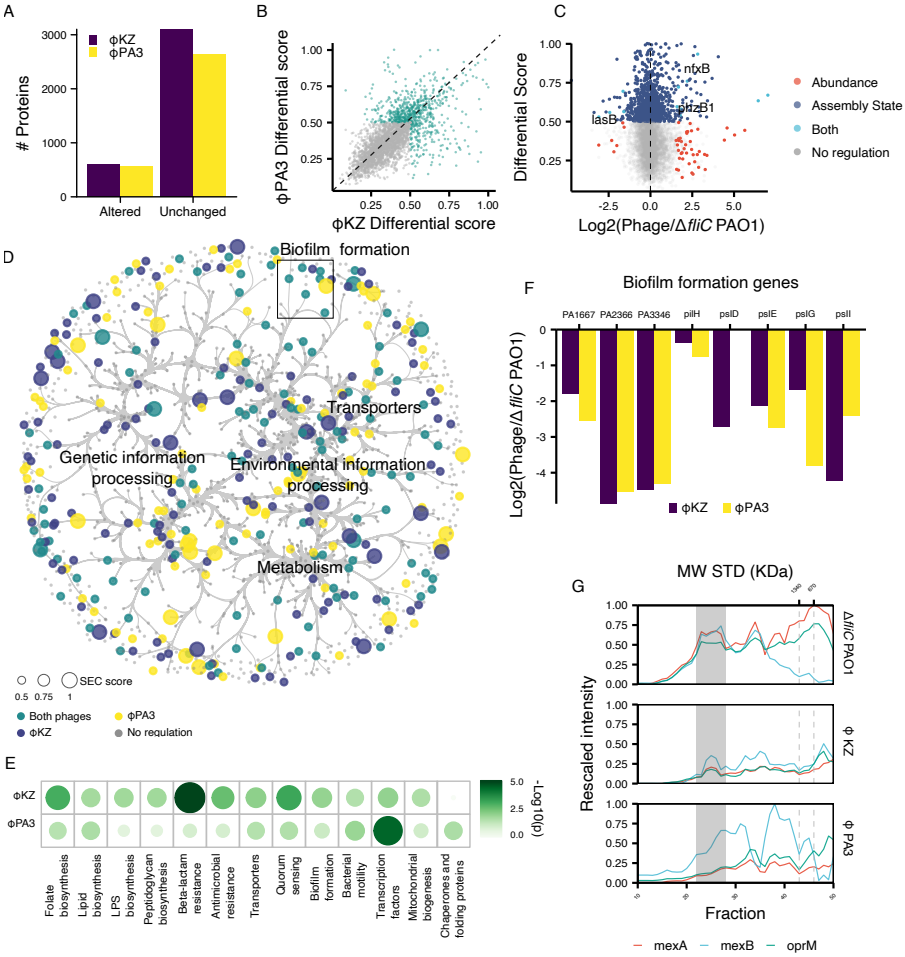


Fig. 3. Differential analysis of SEC-MS data A. Altered host proteins for each experiment. **B**. Scatterplot of differentially regulated proteins. Axis represents the differential SEC score while each dot represents a PAO1 protein. Color indicates significant regulation in either phage. **C**. 2D plot of differential SEC score (Y axis) vs LogFC from global proteome abundance (X axis). Color represents the different regulation level. Proteins highlighted in green are significantly regulated at the abundance level (Log2FC ≥ 2 and $q \leq 1\%$) and assembly state level (SEC score $\geq$ 0.5). Both phages are shown. **D**. SEC derived molecular network for PAO1 proteins. Node color represents the regulation status, while node size shows the SEC score (i.e differential score). Edges are bundled using KDEEB. **E**. Enriched KEGG terms for altered proteins. Node size and color represents the significance on a -log10 scale. **F**. Barplot representing the log2FC biofilm formation genes upon phage infection as compared to control (Δ*fliC*) experiments. **G**. Coelution plot for the efflux pump MexA/B-oprM. Different experiments are represented by the various subpanels.

After having demonstrated the proteome depth achieved in our SEC-MS dataset and the recovery of known complexes, we turned our attention to how Jumbophages re-wire *Pseudomonas* protein complexes by evaluating differences in SEC profiles upon phage infection. Variation in SEC profiles between conditions can arise from differential assembly state (i.e. a protein profile shifting to higher or lower molecular weight), different stoichiometry within a complex, or global alterations in protein abundance. To quantify these different cases, we employed a previously described Bayesian analysis module from the PCprophet package[10] to derive marginal likelihoods (SEC differential score) of protein-level SEC changes between ϕKZ and ϕPA3 versus the receptorless infected samples (i.e. Δ*fliC*). Comparing the SEC-profile differences between phage-infected PAO1 and Δ*fliC* revealed approximately 600 proteins showing SEC variation upon infection by either phage (Fig.3A). Notably, there is substantial consistency in which *Pseudomonas* proteins are altered, and the degree of change in their individual SEC profiles (Fig.3B, cor = 0.677), potentially pointing towards common pathways and complexes hijacked by ϕPA3 and ϕKZ for successful predation. Most of the changes at the assembly state level do not have a corresponding variation in protein abundance at the global proteome level, suggesting that SEC-MS offers an orthogonal view on effect of perturbations, such as infection, on the proteome (Fig.3C). Proteins changing both in assembly state and abundance could be dependent on deep interactome rewiring rooted in a strong transcriptional response to phage infection. For example, we identified several quorum sensing proteins such as lasB, phzB1, and nfxB in ϕKZ and ϕPA3 as proteins changing in both dimensions (SEC and abundance). To identify conserved KZ-like jumbophage manipulation of the host interactome, we mapped the SEC-derived PAO1 interaction network (Fig.3D) with the correspondent protein-level differential

data derived from the comparison between phage and uninfected samples. Albeit a large portion of the nodes do not have a functional annotation, we identified several classes where their component were significantly altered upon Jumbophage phage infection. Several pathways related to drug resistance (folate biosynthesis KEGG ID pae00790, beta lactam resistance KEGG ID pae01501, antimicrobial resistance KEGG ID pae01504) were enriched upon Jumbophage infection (Fig.3E). Moreover, quorum sensing and biofilm formation pathways were enriched in both phage infected samples ($q \leq 0.01$) and several prior studies have highlighted the role of phages in regulating formation of biofilms[24, 25]. We identified multiple proteins in this category having significantly decreased abundance in the high molecular weight region compared to their uninfected counterpart (Fig.3F) which adds further evidence to alteration of quorum sensing as the phenazine system directly triggers biofilm formation in *Pseudomonas*[26, 27]. Membrane proteins were particularly affected by Jumbophage infection with porins and multi-drug efflux proteins (KEGG pae02010: ABC transporters) displaying significant reduction in interactions. As example, the MexAB-OprM complex, a key efflux pump[28], shows almost complete reduction of the fully assembled complex (Fig.3G). It is important to point out that changes we observed could either be beneficial for the phage to overcome its host, or an attempt from the host to eliminate the phage.

Organization of the ϕKZ like Jumbophages viral interactome

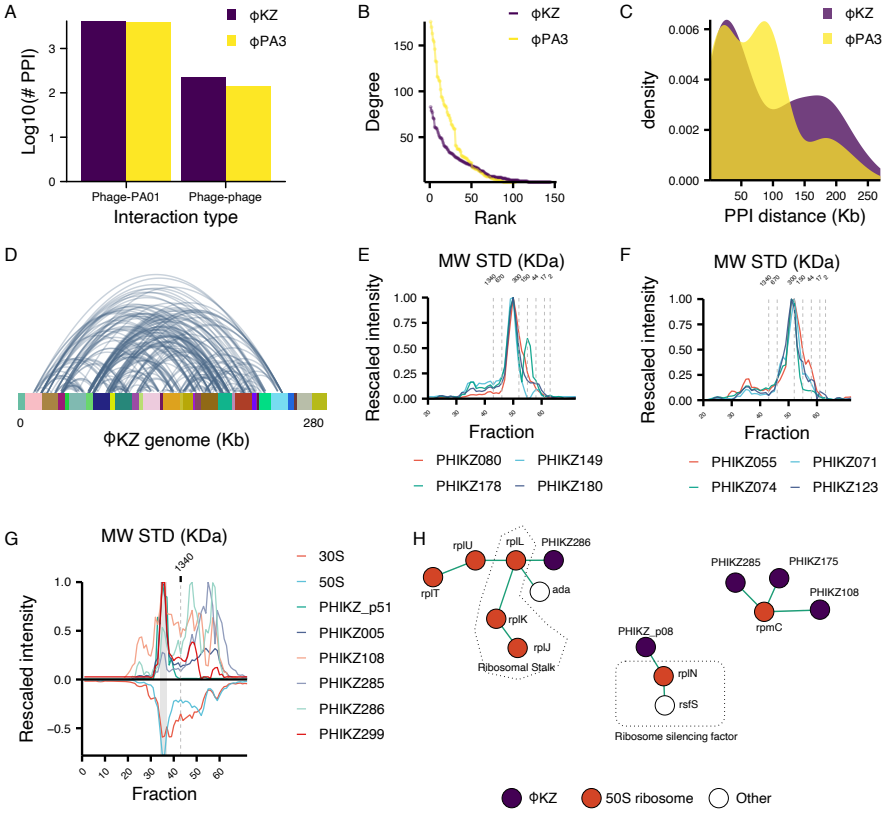


Fig. 4. Comparative analysis of ϕKZ and ϕPA3 interaction networks A. Number of PPIs identified for each phage (Y axis), separated by interaction type. **B**. Degree distribution (Y axis) versus the rank for a particular node (X axis). Line color represents the various phages. **C**. Density plot illustrating the genomic distance in Kb (X axis) for the phage-phage interactions. Color codes represent the two phages. **D**. Arcplot of ϕKZ interactions across the various operons. X axis shows the KZ genome (Kb) and the various colored boxes shows the reported operons. **E, F**. Coelution plot for the ϕKZ non-virion associated RNA-polymerase and the virion associated RNA-polymerase. **G**. Mirrorplot illustrating the coelution of ϕKZ proteins (upper panel) with the 70S ribosome (lower panel). **H**. Interaction network for ϕKZ proteins and ribosomal subunits from the SEC XL-MS experiment. Edges represent an identified crosslink. Dashed line represents a known ribosomal structural component, node color represents whether a protein is a phage protein (purple), ribosomal component (red) or another protein class (white)

The remodeling of host protein complexes can be the result of indirect rewiring of host cellular processes, or direct interactions with phage proteins. Thus, we next investigated interactions directly involving phage proteins, including complexes containing both phage-host and phage-phage interactions. The two Jumbophages used in this study, ϕKZ and ϕPA3 both have high sequence similarity[29] and shows high degree of immune evasion from bacterial antiphage systems[13, 30]. Following SEC-MS and PPI prediction, we defined high-confidence interactions as those with a probability score of $\geq$ 0.75. In total, we identified 292 interactions between pairs of ϕKZ viral proteins, and 6550 host-pathogen interactions between ϕKZ and PA01 proteins. ϕPA3 showed a similar trend with 145 viral-viral and 3979 host-pathogen protein interactions (Fig.4A). Topological analysis of these networks revealed a scale-free architecture (Fig.4B), in line with previous reports that SEC-MS derived networks presents the same architectural features as networks derived from literature curated studies and large PPI databases[10, 31, 32]. Next, we evaluated the distribution of predicted PPI (by SEC-MS) in phage infected PAO1 cells as a function of the genomic separation of their corresponding genes (Fig.4C,D). Indeed as observed with PAO1 cells infected with ϕKZ, the predicted PPI are not typically localized to proteins whose genes lie in a single/adjacent operon(s) but rather distributed all across the genome (Fig.4D). When looking at the overall density of PPI between a set of two phage proteins with respect to the distance of separation (in Kb) of their corresponding genes (Fig.4C), we observed that for both ϕKZ and ϕPA3, interactions are predicted between translation products of phage genes separated by distances as large as 250Kb. For ϕKZ the density displays a bimodal distribution with a sharp peak at $\approx$ 22 Kb and a broader peak at $\approx$ 185 Kb (Fig.4C, purple). Interestingly, for ϕPA3 this distribution is altered, as it shows a greater distribution

of PPI predicted at lower gene separations (two peaks at ≈ 19 Kb and ≈ 97 Kb), that tapers off at larger gene separations (with a third peak at ≈ 182 Kb) (Fig.4C, yellow). This discrepancy between the two phages is explained in part by the fact that as compared to ϕKZ, in ϕPA3 partial duplication or fragmentation of genes into multiple genes within a single/adjacent operon(s) is frequently observed (see Supplementary Table S1). Overall, the PPI distribution confirms the general lack of synteny within the genomes of ϕKZ-like jumbophages and shows the SEC-MS approach as a particularly advantageous technique to query phage encoded complexes at the protein, level agnostic to the overall genome organization (i.e. a guilt-by-association approach at the protein level).

The identified interactions allowed us to recapitulate several known complexes in the Jumbophage proteome, despite a limited number being described at present. As example, we recovered the non-virion associated RNA-polymerase[33] migrating at its expected molecular weight (apparent MW 271 KDa or fraction 53, correct MW ≈ 265 kDa) (Fig.4E) as well as the virion associated RNA polymerase[34] (apparent MW 300 KDa or fraction 52, correct MW ≈ 297 kDa) (Fig.4F). Overall, we observed the presence of several phage peaks groups at high molecular weight in both ϕPA3 and ϕKZ, suggesting the presence of yet undiscovered phage protein complexes or phage-host interactions (Supplementary Fig.S3). For example, we observed a clear peak group composed of PHIKZ005, PHIKZ108, PHIKZ285, PHIKZ286, PHIKZ299, and PHIKZ_p51 at approximately 4 MDa. Intriguingly, these proteins were predicted to be in complex with the fully assembled *P. aeruginosa* 70S ribosome as shown in Fig.4G. To validate these ϕKZ proteins as ribosomal interactors, we performed cross-linking mass spectrometry (XL-MS)[35] on a pooled

sample from the SEC fractions correspondening to the 70S ribosome (Supplementary Fig.2). We identified 975 crosslinks in total (202 hetero-links and 871 homo-links), covering several previously reported protein complexes such as the ATP synthetase, AAA protease complex, SEC complex D/F and succinyl-coA synthetase (Supplementary Fig.S4). The XL-MS data encompassed 24 *P. aeruginosa* ribosomal proteins (separated in 30S and 50S) of which 3 showed interaction with 6 ϕKZ proteins. Amongst the identified ϕKZ proteins, we recovered PHIKZ285, PHIKZ286, and PHIKZ108, which were predicted from the SEC-MS data to be in complex with the 70S ribosome. Moreover, we identified PHIKZ_p08 and PHIKZ175 as additional ribosomal interactors. All of these proteins interacted with distinct structural features of the ribosome as depicted in the interaction network in Fig.4H. PHIKZ286 bound the L1 ribosomal stalk (*rplL*, *rplK*, and *rplJ*) which has an important role in tRNA translocation[36] and is the contact site for several translation factors[37]. PHIKZ_p08 interacted with *rplN* bound to its ribosome silencing factor *rsfS* which slows down or represses translation[38]. Finally, PHIKZ285, PHIKZ175, and PHIKZ108 were bound to *rpmC* which is an accessory proteins positioned near the exit site and is required for triggering nascent polypeptide folding[39]. It is important to note that, these interactions could either be functional within the ribosome or represent active translation of the phage proteins. Further mechanistic characterization would be needed to determine if such phage proteins hijack ribosomal function.

Identification of novel protein-coding sequences by SEC-MS

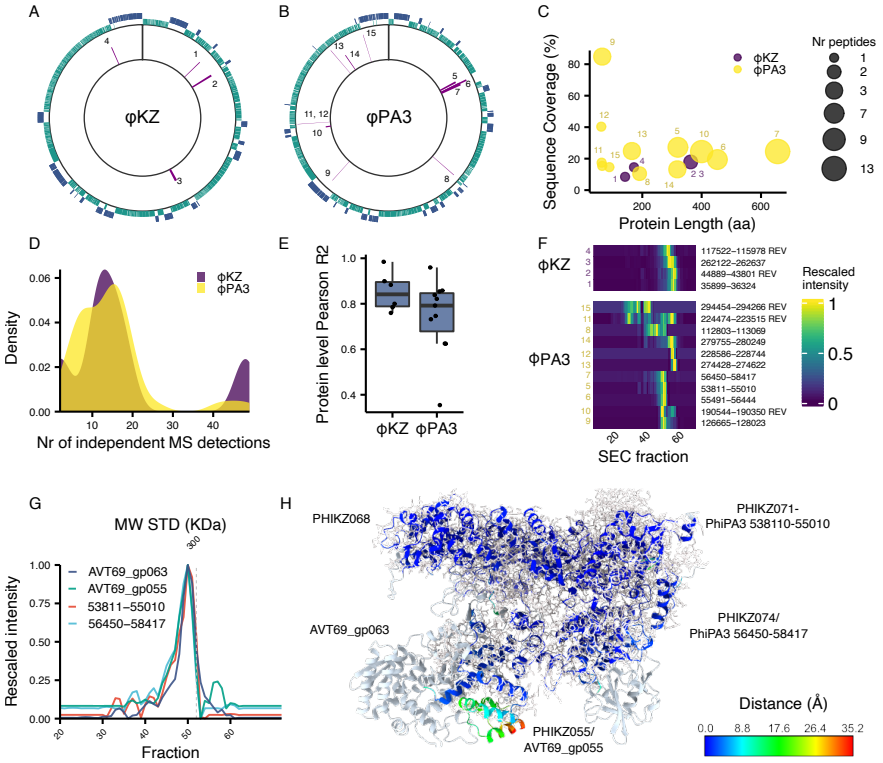


Fig. 5. Identification of novel phage proteins. A-B. CDS plot for ϕKZ and ϕPA3. Forward CDS are colored in green while reverse CDS are represented in purple. Identified novel proteins are highlighted in the histogram (inner circle). **C**. Scatterplot of protein length vs percentage of sequence coverage in the SEC-MS experiments. Dot size represents the number of proteotypic peptides identified. **D**. Distribution of the identifications (defined as number of independent MS detections using a 1% peptide spectrum matching FDR) for the novel ϕKZ and ϕPA3 proteins. **E**. Boxplot showing the Pearson correlation between the two replicates (n=72). Each novel protein is represented as a dot. The box boundaries show the interquantile range (IQR) and its whiskers 1.5×IQR. **F**. Heatmap representing the elution profile for all the novel ORFs. X axis represents the fraction number while the cell color shows the unit-rescaled intensity. **G**. Coelution profile for predicted nvRNAp in ϕPA3. **H**. Superimposition of reported structure for the ϕKZ nvRNAp (stick, grey) and predicted structure for the ϕPA3 nvRNAp (ribbons). Chains are colored by their distance to the ϕKZ nvRNAp structure after superimposition.

The multiplexed nature of DIA allows un-biased sampling of the full precursor space[40], hence we queried our data for the presence of peptides from novel phage proteins. To achieve this, we employed EMBOSS to build a custom protein FASTA comprised of predicted novel ORFs, and then extracted expected peptide fragment ion chromatograms for these novel ORFs from the DIA data. This resulted in the detection of 4 previously undescribed proteins for ϕKZ (2 forward and 2 reverse ORFs) and 11 for ϕPA3 (8 forward and 3 reverse) (Fig.5A and B). The authenticity of these novel proteins is supported by the detection of two or more proteotypic peptides for nearly all novel proteins (Fig.5C), as well as the reproducible detection of the same peptides in 15 or more consecutive fractions across independent experiments (Fig.5D). All novel proteins showed reproducible quantitation between biological duplicate experiments (n=72 per replicate), with an average peptide-level correlation of 0.75 for ϕPA3 proteins and 0.82 for ϕKZ proteins (Fig.5E). Most of these proteins did not migrate at their predicted molecular weight, suggesting their potential involvement in high-order assembly, with two ϕPA3 proteins eluting at $\approx$ 4 MDa (Fig.5F), pointing towards their potential association with large macromolecular complexes.

Some of the novel ORFs are further supported by a great degree of sequence overlap with other reported proteins. The most staggering example is the identification in ϕPA3 of the reverse sense ORF 56450-58417 which shows >70% sequence similarity with previously reported proteins from various *Pseudomonas spp* phages (ϕKZ, Psa21, Phabio, 201ϕ2-1, and PA1C)(Supplementary Fig. S5A). Interestingly, all proteins showing $\geq$ 50% homology to 56450-58417 are previously reported or proposed phage RNA polymerase components, such as ϕKZ gp74 (non-virion associated RNAp, UniprotID Q8SD88)[34, 41, 42]. To date, there is no experimental evidence

of a nvRNAP in ϕPA3. To derive other putative members of this complex, we extracted the predicted interactors of ORF 56450-58417 and performed BLASTp analysis to identify proteins showing homology to other Jumbophage RNA polymerase components. From this analysis, we selected 3 interactors (gp55, gp63, and the novel ORF 53811-55010) showing >50% conservation with multiple Jumbophage proteins annotated as RNAP components (Supplementary Fig.S5B-D). Specifically, we identified homologs of both the β^1 polymerase subunit (gp55 and ORF 56450-58417), as well as homologs of the β subunit (ORF 53811-55010). The ϕPA3 protein gp63 displays 57% homology to ϕKZ gp68, an essential nvRNAp component which lacks structural similarity to know components of previously reported RNA polymerases[41]. By employing the molecular weight derived from the SEC peak position, we estimated the nvRNAp MW in ϕPA3 being $\approx$ 321 KDa (Fig.5G). Assuming the lack of homodimers in the structure, the predicted MW for these four proteins was $\approx$ 253 KDa, suggesting a putative missing subunit. Of note, we did not identify 53811-55010 interactors corresponding to ϕKZ g123, another β subunit component, which could explain this observation. Nonetheless, to explore the possibility of these proteins (gp55, gp63, ORF 53811-55010, and ORF 53811-55010) folding into an RNA polymerase-like assembly we performed structural prediction of this peak group using AlphaFold2 multimer[43]. We achieved high prediction confidence, with inter-chain predicted confidence (ipTM + pTM) of 0.82 and low predicted aligned error (PAE) across the entire oligomer (Supplementary Fig.S6). We aligned the best scoring model to the reported structures for the ϕKZ nvRNAP (PDB 7OGP and 7OGR)[42] as depicted in Fig.5H. We reached a template modeling (TM) score of 0.503 using US-Align[44] and an average RMSD of 1.016 Å using MatchMaker[45] between our proposed ϕPA3

vRNAp and the ϕKZ RNAp (70GR), strongly pointing towards a global tertiary structure similarity between these two assemblies. As we obtained low distances for the β and β^1 subunits, we set to investigate the misaligned region at the C-term of the polymerase clamp (gp63 in ϕPA3 and gp68 in ϕKZ). Despite showing high sequence homology (68%), these two proteins share a large intrinsically disordered region (IDR) in the middle of the sequence (275-293 aa for gp68 and 277-301 aa gp6) as exemplified in Supplementary Fig.S7) which gives the central region extreme flexibility hence resulting in a random orientation for the folded C-term in ϕPA3 gp63 following AlphaFold prediction.

Discovery and validation of novel injected phage proteins

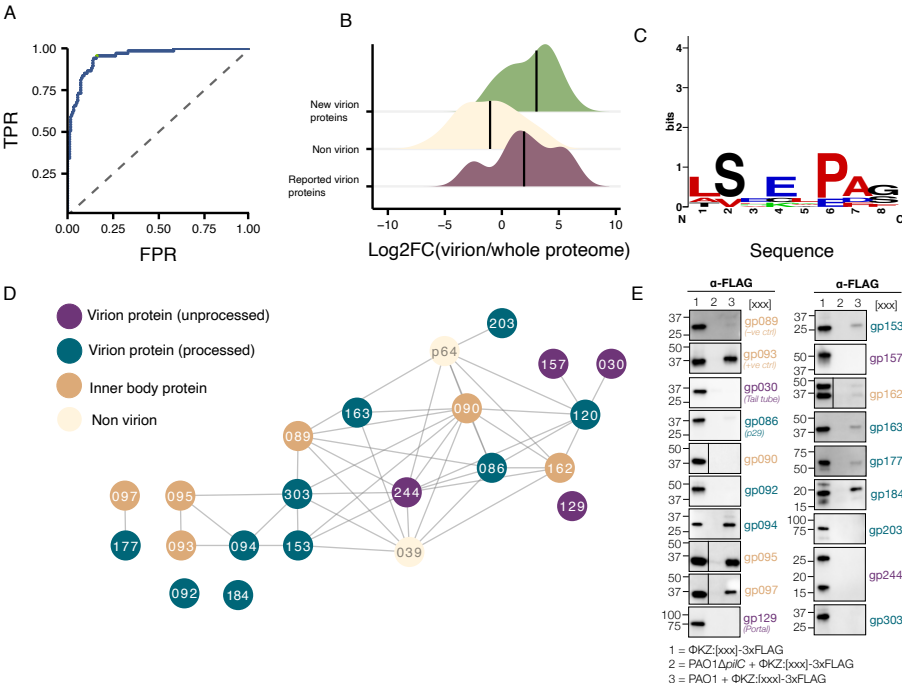


Fig. 6. Data-driven analysis of injected inner body proteins A. ROC curve for virion MS (AUC ≈ 0.94) using as ground truth the prior reported virion proteins. Green highlight selected threshold for maximimum sensitivity at the lowest FPR. **B**. Density plot representing the enrichment of virion proteins over a whole proteome infection experiment expressed as log2 fold change (X axis). Different colors represent whether a protein was previously reported as virion (bordeaux), novel from our virion dataset (green) or non-virion (cream). **C**. Sequence logo for the proteins in the IB interaction network. X axis shows the position from N to C term while Y axis represents conservation in bits. **D**. SEC-MS derived interaction network for the reported IB proteins (gp93/95/97). Color code represents the query protein (aquamarine), abundant virion protein (dark purple) defined as top 20% most abundant proteins, non abundant virion proteins (grey) and proteins not identified in the virion MS experiment (green). **E**. Injection of phage proteins evaluated by WB of 3x FLAG phage tagged proteins.

Although it is well known that ϕKZ phages guard their genome from nucleolytic host-immune systems by building a proteinaceous shell[13], this structure is only visible after 20 minutes of infection. Little is known about how the phage genome is protected or packaged prior to shell assembly. In

order to identify phage proteins proximal to the genome (with possible protective functions), we first set on obtaining an in-depth virion proteome as this will allow distinction of virion proteins (injected) from newly synthesized proteins. We performed cesium-gradient purification of ϕKZ coupled with deep peptide fractionation and long chromatographic acquisition (see Supplementary Methods for details). The 245 ϕKZ proteins identified in this dataset encompassed $\geq 90\%$ of previously reported head proteins (Supplementary Fig S.8A). This drastic increase in protein number is dependent on the increase sensitivity and sequencing speed of the MS utilized for acquisition, the extensive sample fractionation prior to MS acquisition (see Supplementary Fig S.8B), as well as possible contamination from the cesium-fractionation. To account for the latter, we compared our enriched virion sample with the previously reported virion proteins to derive an ROC curve, which we used to select an intensity threshold maximizing recall of known virion proteins and minimizing false positive rate (Fig 6A). Based on this, we selected 81 proteins in total (58/61 of previously reported), hence adding 23 proteins to the virion composition. This stringent filtering resulted in the selection of proteins which are strongly enriched over their corresponding protein abundance in a non-enriched samples (Fig 6B). As prior work reported extensive gp175-driven proteolysis of the head and inner body (IB) proteins[46], we performed a semi-tryptic search on the purified virion data to confirm prior reported cleavages and potentially identify novel ones. We recovered 63 semi-tryptic peptides, of which 15 could be mapped to prior data[46] ($\approx 40\%$ overlap). Within our semi-triptic peptides, we identified 20 cleavages corresponding to the reported IB proteins (gp93/95/97) and 12 mapping to 9 unreported proteins. Of note, 8 cleavages could be mapped to gp94, gp177, and gp303 (see Supplementary data and Supplementary Fig. S8C). To identify consensus

sequences within our set of IB interactors we performed motif enrichment analysis using STREME[47]. In this analysis, we employed as input the reported IB proteins (gp93/gp95/gp97) and the SEC-MS derived interactors after filtering using the virion proteome (gp080, gp094, gp177, gp235, gp237, and gp303). We identified the consensus motif LSxE as enriched (BH-adjusted $p = 6e^{-5}$) which confirms the previously reported one S/A/G-X-E motif, while providing additional specificity in the P2 position (Fig6C).

We then queried the ϕKZ interaction network using as input the virion proteome, to identify putative injected proteins (Fig. 6D). Building on this data, we selected the interactors of the previously reported proteins (gp94, gp153, gp162, gp163, and gp177) for further validation using our previously reported assay for evaluating injection[17]. By further lowering the interaction thresholds to all positive predicted interactions (i.e PPI probability ≥ 0.5 instead of 0.75 utilized to select high-confidence interactors), we further identified gp184 as IB interactor and validated it as injected proteins. These experiments, confirmed injection of the previously reported IB proteins (gp93, gp95, gp97) and furthermore validated the injection of all their interactors as showcased in Fig. 6E. Prior assignment of IB proteins was based on an arbitrary cutoff[46]. Here by using more sensitive MS of the virion combined with SEC-MS, we identify a full new set of proteins that are highly abundant, found in the virion, interact with the previously reported IB proteins and herein we validate injection of 8 proteins (3 previously reported). Overall, these proteins give us a starting point to unravel the interactome of the ejected phage genome and identify proteins that protect the genome from host nucleases.

Discussion

Understanding host and pathogen interactions and their dynamics upon infection is a crucial component to deepening our knowledge on the mechanisms regulating infection progression and outcome. To date most proteomics studies of infectious diseases focused on the analysis a few pathogen proteins by tag or antibody-based purification or the measurement of protein abundance variation in infected samples. Yet, it is widely known that the pathogen proteome works in ensemble through protein-protein interactions to hijack the host cell which in turn regulates both expression and interaction between host proteins. Hence, a system-wide view on the intrinsic modularity of the pathogen proteome and how it quantitatively regulates host complexes is key to understand pathogenic mechanisms at the molecular level.

In this study we demonstrate the first application of SEC-MS to systematically investigate pathogen proteome organization and host interactome plasticity upon Jumbophages infection in *P. aeruginosa*. KZ-like phages (specifically ϕKZ and ϕPA3) are potent killers of *P. aeruginosa* (with a broad host range), making them timely alternatives to antibiotics with many KZ-like phages already in clinical trials to treat bacterial infections. By obtaining an atlas of these phages interactomes, we can start to shed light into the mechanism of action of these phages in diverse aspects of their infection cycle (e.g. takeover of host translational apparatus, phage replication, host genome degradation or host lysis).

Our KZ-like phages interactomes offer evidence for the subdivision of Jumbophage proteomes into distinct assemblies such as virion and non-virion associated RNA polymerases as well as interactions of pathogen proteins with key host complexes like the ribosomal stem and ribosomal silencing factors.

Moreover, while the lack of immediate genome organization hinders the prediction of functions for phage proteins, the deep coverage and unbiased nature of SEC-MS data offers a straightforward approach to identify novel complexes and propose putative functions. As example, by using SEC-derived interactors of a de-novo predicted ϕPA3 protein (ORF 56450-58417), we identified an heterotetrameric assembly which we demonstrated bearing strong structural similarity to the reported nvRNAP in ϕKZ. This suggests that the unbiased nature of SEC-MS data allows to not only to identify uncharacterized, novel proteins, but also to probe their putative function through their interactions in a structurally defined complex that could be further investigated with structural biology approaches. Beside identification of interactions, these maps offer the opportunity to further quantify host interactome remodelling and disentangle variation in expression from assembly state. By comparing the *P. aeruginosa* interactome between infected and uninfected, we observed a large degree of changes during infection, with perturbation of similar complexes between the two jumbophages suggesting conserved mechanisms of phage predation. While we provided a first draft of the KZ-like jumbophage interactome it is important to point out the trade-off between specificity and throughput in interaction identification, which we tried to mitigate by utilizing only high-confidence interactions for analysis. Advances in deep learning models for prediction of interactions from co-fractionation mass spectrometry data and integration of orthogonal features (beside the coelution itself) such as predicted structure or function is expected to improve prediction accuracy and reduce false discovery rate for uncharacterized proteomes. Overall, the characterization of host-pathogen molecular networks remains challenging, but we provided the first interactome-wide study of infection progression using two models ϕKZ-like phages in *Pseudomonas aeruginosa*.

Wider application of SEC-MS is expected to significantly accelerate the characterization of pathogenic mechanisms by providing proteome-wide insights into the physical association between host and pathogen complexes, thus enabling identification of novel druggable targets, host vulnerabilities, or guidance in the development of novel biologicals.

Data availability

The supporting MS data is available via ProteomeXchange with the identifier PXDXXXX. Novel ϕPA3 and ϕKZ proteins have been submitted to UniProt. All the code to reproduce the plots as well as the intermediate data and Alphafold2 predicted structures are available on GitHub at https://github.com/anfoss/Phage_data.

Acknowledgments

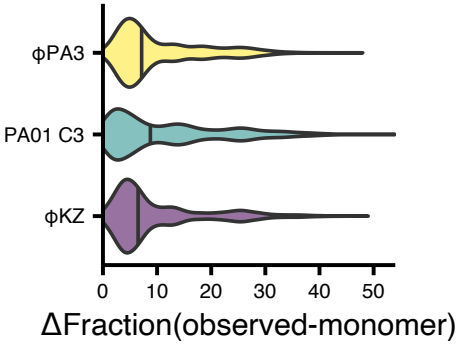
This work was supported by an NIH grant 1R01AI167412 to JBD and DLS. We thank Dr. James Wells at UCSF for the usage of the HPLC used to perform the size-exclusion experiments. Molecular graphics were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. Figure 1A-B was prepared with Biorender.

Author contributions

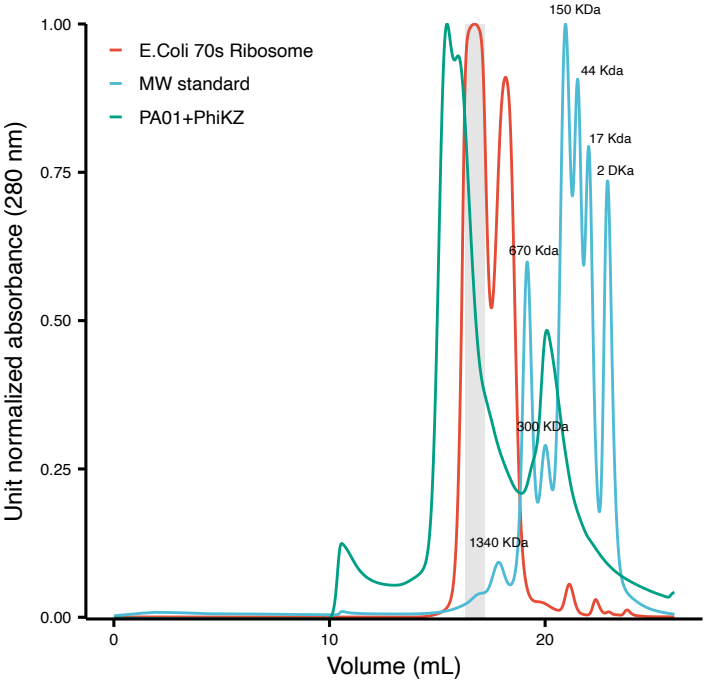
AF: Performed proteomics sample preparation and analysis all MS data, developed PhageMAP and wrote the manuscript. DLS, JBD, DA, NK: Conceptualization, supervision, writing, and funding acquisition. AP: Novel ORF

prediction DM. CK. BG: Phage infection experiments, virion enrichment, microscopy and WB for injected proteins EM, MM: Performed shell enrichment YP: Critical input in revising the manuscript All co-authors contributed in reviewing and editing the manuscript.

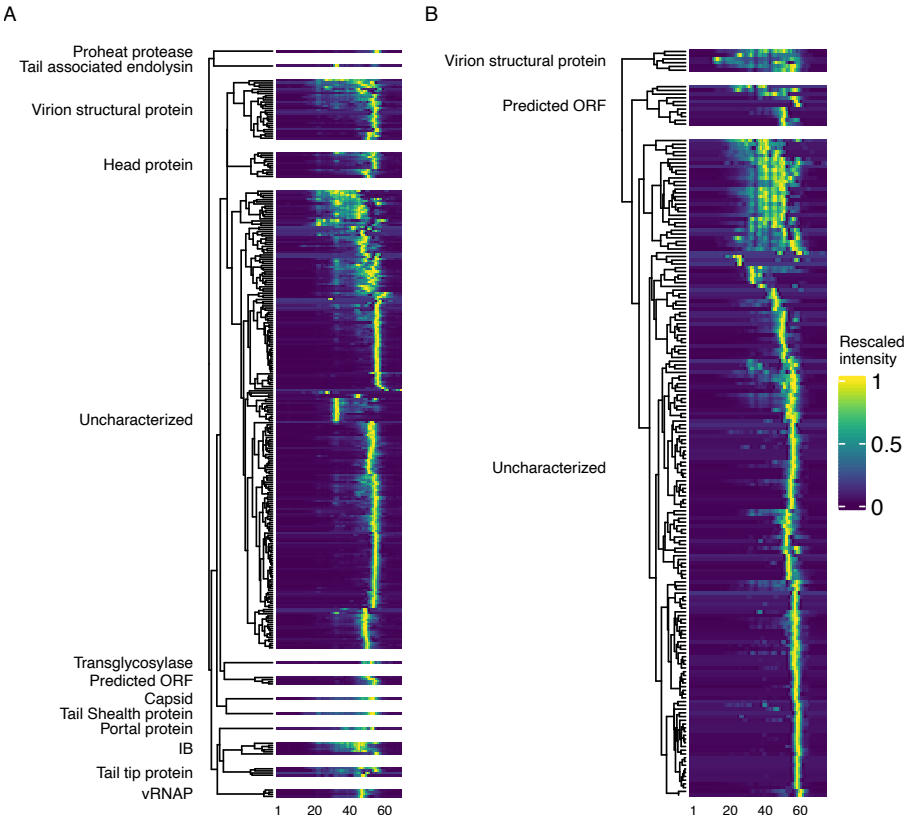
Supplementary figures



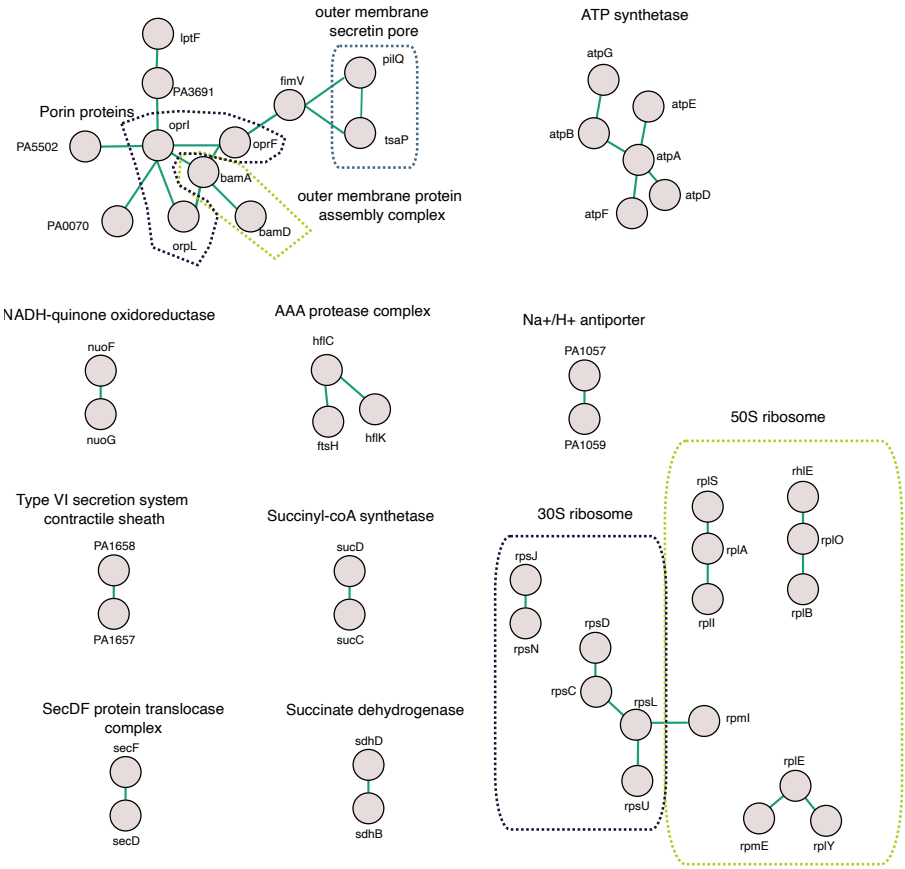
Supplementary Fig. S 1. Violin plot showing the distance between the observed protein SEC peak and their predicted molecular weight expressed as fraction number for the single-peak proteins. Black line represents the mean.



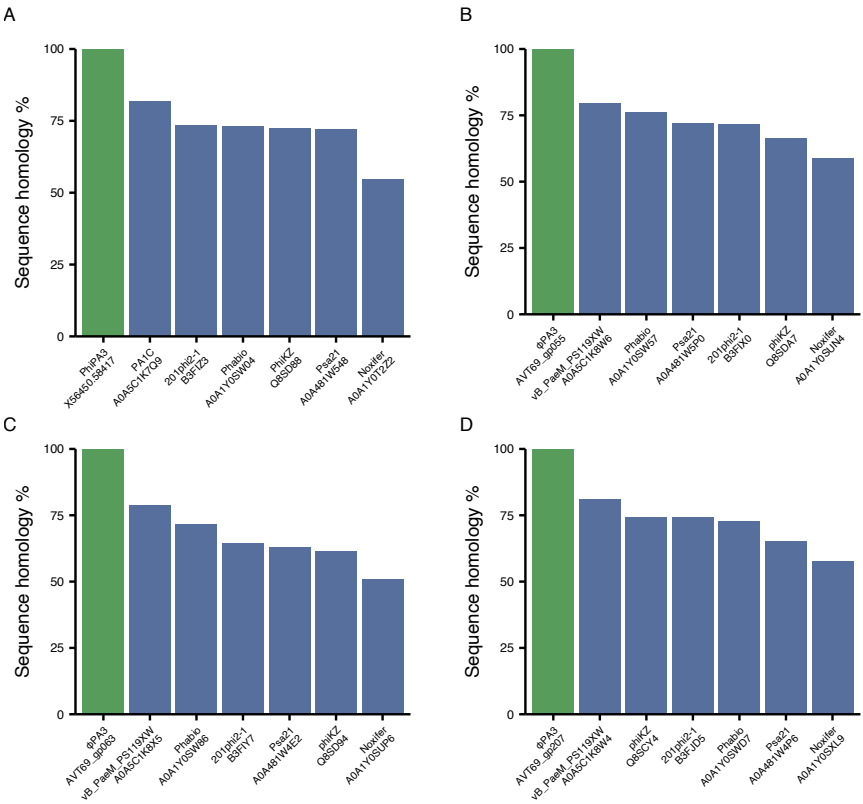
Supplementary Fig. S 2. Ribosome absorption profile and MW from SEC Chromatographic traces collected at 280 nm for the purified 70S ribosome (red), the protein mixture used as molecular weight standards (cyan) and the ϕKZ infected sample (green). The grey area shows the fractions used for the ribosomal crosslinking experiment.



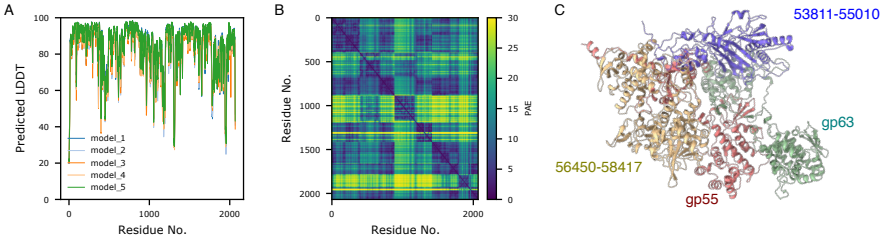
Supplementary Fig. S 3. Distribution of the KZ-like phages proteomes into discrete assemblies A-B. Coelution heatmap for all the phage proteins identified in ϕKZ (**A**) and ϕPA3 (**B**). The dendrogram branches are labelled based on manual literature curation for the corresponding proteins in the peak group. Color represents the unit-rescaled intensity. X axis represent the fraction number.



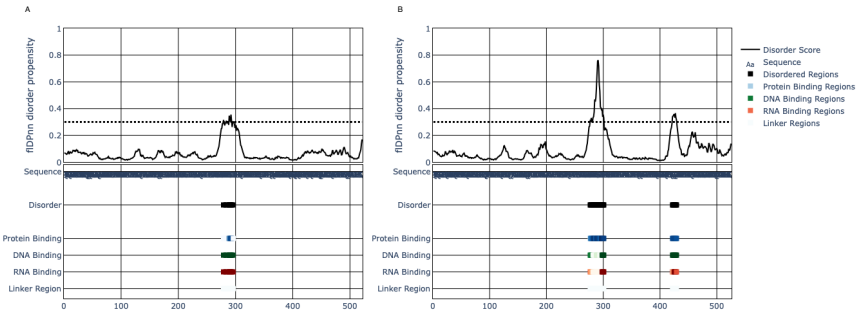
Supplementary Fig. S 4. Know protein complexes recovered by SEC-XL MS Interaction network derived from the heterolinks detected at 5% CSM-FDR for reported complexes in *Pseudomonas*. Text label shows the complex for the interacting proteins.



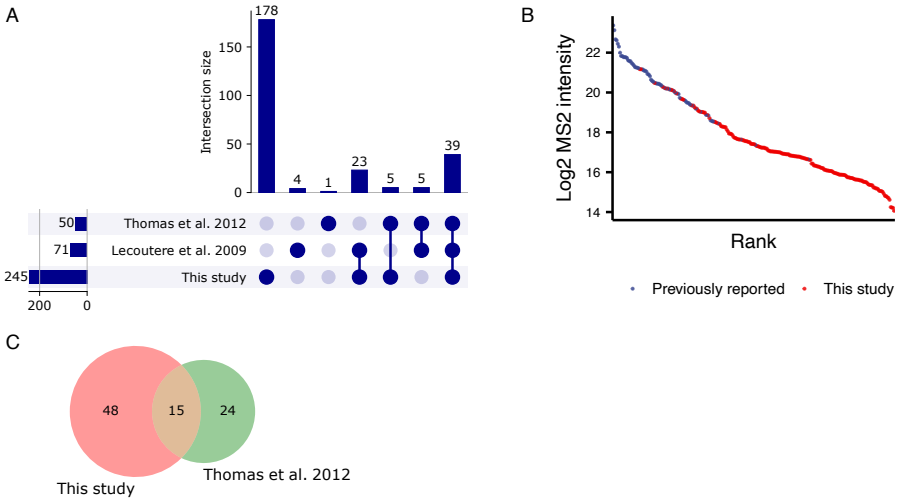
Supplementary Fig. S 5. Alignment of ϕPA3 56450-58417 interactors with other Jumbophage proteins A-D. Barplot showcasing the sequence homology between ORF 56450-58417 (**A**), gp55 (**B**), gp63 (**C**) and gp207 (**D**) to other *Pseudomonas* phages protein. ϕPA3 proteins are highlighted in green. Y axis shows the percentage of sequence homology.



Supplementary Fig. S 6. Prediction of ϕPA3 non virion associated RNA structure A. Per-residue local confidence (pLDDT) versus sequence (X axis). Different line colors represents the different AF2 model **B**. Predicted alignment error (PAE) heatmap **C**. Structure of best scoring model (iPTM + TM = 0.826)



Supplementary Fig. S 7. Intrinsically disordered region in ϕKZ gp68 and ϕPA3 gp63 Prediction of disordered regions using flDPnn[48]. X axis represent sequence, while different rows shows different local predicted properties between ϕKZ gp68 (**A**) and ϕPA3 gp63 (**B**)



Supplementary Fig. S 8. Benchmark of purified virion MS versus other reported studies of ϕKZ head proteins. **A**. Upset plot showing the overlap in protein IDs between this study, Thomas et al. (2012)[46] and Lecoutere et al. (2009) [49] **B**. Distribution of intensities for virion proteins identified in this study. Blue proteins were previously identified while red proteins are novel virion proteins from this study. **C**. Venn diagramm of semi-tryptic peptides detected in this study versus Thomas et al[46]

Supplementary tables

Table 1. Gene duplication in ϕKZ-like Jumbophages

ϕKZ gene	ϕPA3 gene	Virion protein
PHIKZ083	AVT69_gp236, AVT69_gp079, AVT69_gp081	YES
PHIKZ093	AVT69_gp097, AVT69_gp099	YES
PHIKZ094	AVT69_gp096, AVT69_gp097, AVT69_gp098	YES
PHIKZ118	AVT69_gp130, AVT69_gp132	NO
PHIKZ131	AVT69_148, AVT69_149	NO
PHIKZ133	AVT69_gp150, AVT69_gp152	NO
PHIKZ134	AVT69_gp150, AVT69_gp151, AVT69_gp152	NO
PHIKZ135	AVT69_gp150, AVT69_gp151, AVT69_gp152	NO
PHIKZ144	AVT69_gp161, AVT69_gp211	YES
PHIKZ145	AVT69_gp162, AVT69_gp163	YES
PHIKZ146	AVT69_gp164, AVT69_gp165, AVT69_gp167	YES
PHIKZ178	AVT69_gp207, AVT69_gp209	YES
PHIKZ179	AVT69_gp131, AVT69_gp208	NO
PHIKZ184	AVT69_gp215, AVT69_gp095	YES
PHIKZ209	AVT69_gp245, AVT69_gp251	NO

Material and methods

Cloning

All plasmids and primers used in this study are listed in the Supplementary Information. PAO1 cells were transformed with pHERD30T constructs using electroporation.

Bacterial culture and cloning

Pseudomonas aeruginosa strains PAO1 (WT or ΔfliC) were grown overnight in 3 mL LB at 37°C with aeration at 175 rpm. Cells were diluted 1:100 from a saturated overnight culture into 100 mL LB with 10mM $MgSO_4$ and grown for $\approx$ 2.5 hours at 37°C with aeration at 175 rpm. At OD600nm = 0.5-0.6 ($\approx 3e^8$ CFU/mL), the cell cultures were infected with bacteriophage (ϕKZ or

ϕPA3; MOI $\approx$ 1) on ice for 10 minutes (to allow complete adsorption of virions onto cells) and then incubated at 30°C for 50 minutes (total time of infection 60 minutes). Thereafter, the cell cultures were transferred to pre-chilled 50 mL falcon tubes, centrifuged at 6000x*g*, 0°C for 5 minutes. The supernatant was discarded and cell pellets were washed twice with 5 mL ice-cold LB and combined. After the final wash, the bacterial pellets were resuspended in 5 mL ice-cold LB. The concentrated cell culture was flash frozen in liquid nitrogen and subsequently mechanically lysed using a SPEX-freezer mill.

Cesium gradient purification of phage virions

Bacteriophages (ϕKZ or ϕPA3) were propagated in LB at 37°C with PAO1 as a host. Liquid growth curve experiments were used to ascertain the MOI of bacteriophage stock needed to ensure complete lysis of the bacteria following a substantial growth as ascertained by OD600 measurement. Growth curve experiments were carried out in a Synergy H1 micro-plate reader (BioTek, with Gen5 software). Cells were diluted 1:100 from a saturated overnight culture with 10 mM $MgSO_4$. Diluted culture (140 μl) was added together with 10 μl of 10X serial dilutions of bacteriophage stocks to wells in a 96-well plate. This plate was cultured with maximum double orbital rotation at 37°C for 24 h with OD600 nm measurements every 5 minutes. Thereafter, the bacteriophage stock was added at the appropriate MOI to a 1:100 back-dilution of a saturated PAO1 overnight culture in 100 mL LB with 10mM $MgSO_4$ and the bacterial culture incubated for 24 hours (37°C with aeration, 175 rpm). 5 mL of chloroform was added to the cultures in a fume-hood and the cultures were incubated to with chloroform for 15 minutes (37°C, 175 rpm) to ensure maximum lysis of bacterial cells. The cell cultures were transferred to 50 mL falcon tubes and centrifuged at 6000 x*g* for 15 min to pellet bacterial debris.

The supernatant (containing bacteriophages in high titer) was carefully transferred to a fresh set of 50mL falcon tubes and centrifuged and 6000xg for 15 min to pellet any residual bacterial debris. The supernatant was transferred to fresh 50 mL falcon tubes with 2 mL chloroform. To obtain high purity virion particles, a previously described protocol was followed[50]. The virions from the bacterial cell lysate were concentrated by slow stirring overnight at 4°C in 1 M NaCl and 10% PEG (final concentration) and then pelleted (11'300xg, 4°C, 30 min). Pellets were resuspended in 20 ml of SM buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM $MgSO_4$, 0.002% gelatin) containing Complete Protease Inhibitor (Roche). The phage suspension (5.8 mL/tube) were layered onto CsCl step gradients composed of the following concentrations of CsCl: 1.59 g/ml (0.75 ml), 1.52 g/ml (0.75 ml), 1.41 g/ml (1.2 ml), 1.30 g/ml (1.5 ml) and 1.21 g/ml (1.8 ml). The buffer used throughout the gradient was 10 mM Tris-HCl (pH 7.5) and 1 mM MgCl2. Tubes were spun at 31,000 rpm for 3h at 10°C in an SW41 rotor (Beckman Coulter ultracentrifuge) and the resulting phage band had a buoyant density of 1.36 g/ml. This fraction was collected and dialyzed against three changes of 50 mM Tris-HCl and 10 mM $MgCl_2$ at 4°C. This ultra-purified phage stock was diluted in SM buffer and its titer assessed using plaque assays. Finally, the phage virion stock was acetone precipitated using 8 volumes of ice-cold acetone.

Bacterial infection and SEC sample preparation

Cryomilled samples were resuspended in ≈ 4 ml of SEC running buffer (50 mM ammonium bicarbonate and 150 mM NaCl pH 7.4) supplemented with protease inhibitors (Roche) and ultracentrifuged at 60'000 g for x minutes at 4°C. The supernatant was concentrated to 100 μL using a 100 KDa molecular weight cutoff filters to simultaneously enrich for high-molecular weight assemblies and

deplete monomeric proteins. The concentrated sample was centrifuged once more at 10'000 g at 4°C to remove particles.

Size-exclusion chromatography

Approx 1000 μg per sample ($\approx 80 - 90\mu$L as estimated by Bradford's assay) were separated on a Agilent Infinity 1260 HPLC operating at 0.5 mL/minute in SEC running buffer with a Phenomenenex SRT-C1000 column connected and cooled at 4 °C. 72 fractions of 125 ul were collected after 3.75 ml until 13 ml and the column was then washed with 2 column volumes (18 mL) of SEC buffer. The MW was estimated using a protein mixture (Phenomenex AL0-3042), while a *E.Coli* 70s ribosome (NEB, cat nr P0763S) was used to estimate which fractions to use for ribosome XL-MS.

SEC-MS proteomics sample preparation

The SEC samples were prepared as we previously reported[51] using a 96 well filter-aided sample preparation (FASP). The FASP-filters were conditioned by washing twice with 100 μL of dd$H_2$0. SEC buffer was removed by centrifugation (1800 g 1 h) and proteins were resuspended in 50 μL of TUA buffer (TCEP 5 mM, Urea 8M, 20 mM ammonium bicarbonate) and incubated on a thermos shaker (37°C, 400 rpm) for 30 minutes. Cysteine residues were then alkylated by addition of 20 μL CAA buffer (Chloroacetamide 35 mM, 20 mM ammonium bicarbonate) for 1 h at 25°C in the dark. TCEP and IAA were removed by centrifugation (1800 g, 30 min) and filters were washed 3 times with 100 μL of 20 mM ammonium bicarbonate. Proteins were digested in 50 μL of 20 mM ammonium bicarbonate with 1 μg of tryspin per fraction. A 96 well receiver plate (Nucleon, Thermo-Fisher) was used to collect the peptides by centrifugation for 30 minutes at 1800g. The filter plates were washed once with 100 μL

of ddH_2O and centrifuged to dryness (1800g, 60 minutes). The peptides from the receiver plate were transferred to protein LoBind tubes (Eppendorf) and the corresponding well was washed with 50 μL of 50% acetonitrile (ACN) in ddH_20 to increase the recovery of hydrophobic peptides. The combined resulting peptides per each fraction were vacuum dried and stored at -80 C until MS-acquisition. For each phage, 5 μL from each fraction were pooled together to generate a phage-specific library. Each sample specific library was prepared on a C18 spin column (Nest). Following activation of the column with 1 column volume (CV) 100% ACN and wash with 2 CV of 0.1% formic acid the peptides were bound to the column and eluted using a step-wise gradient of ACN from 5 to 25 (5% increases) in 0.1% triethylamine to account for the increased hydrophobicity of the XL peptides compared to not modified ones. A final fraction at 80% ACN was added to recover hydrophobic peptides.

Proteomics sample preparation for virion enriched protein pellets

Dried proteins were resuspended in 100 μL of 8M urea, 100 mM ammonium bicarbonate (ABC) pH 8.1. TCEP (Thermo Fisher) was added to 5 mM final concentration and the samples were incubated at room temperature for 30 minutes. Reduced cysteines were alkylated with 10mM chloroacetamide (CAA) for 30 minutes in the dark. Following alkylation, the urea was diluted to 1 M with 100 mM ABC and the proteins were digested with 2 μg of trypsin per sample for 14 hrs at 37°C in a thermo-shaker (600 rpm). Digestion was stopped by acidification using 10% formic acid (FA) and the samples were desalted using a C18 spin column (Nest group). Briefly, columns were activated using 1 column volume (CV) of ACN and then equilibrated with 2 CV of 0.1% FA. Peptides were loaded twice and then washed with 3 CV of 0.1% FA. Elution

was done using 0.5 CV of 50% ACN 0.1% FA and repeated twice. Samples were dried under vacuum and stored at -80°C until acquisition.

Crosslinking MS sample preparation

ϕKZ infection and SEC-separation were performed as described above. Following separation, the SEC-fractions corresponding to the 70S ribosome peak (F33-F38) were pooled. The was crosslinked for 1 hr at RT using 5 mM DSSO from a freshly prepared 30 mM stock in water-free DMF. The reaction was quenched by addition of ABC to 50 mM for 30 minutes at RT and the proteins were precipitated using 8 volumes of ice-cold acetone. Following overnight incubation, pellets were washed 5 times with 8x volumes of ice-cold acetone and briefly dried under vacuum. The pools were reconstituted in 8M urea, 100 mM ABC and 5 mM TCEP and incubated for 30 minutes at RT. CAA was added to 10 mM final concentration and the samples were incubated in the dark for 1 hr. Urea was diluted to 1 M by addition of 100 mM ABC and the proteins were digested overnight with 2 ug of trypsin in a thermo shaker at 30 °C. Samples were acidified with 10% TFA and high-ph tip fractionation was performed as we previously described[51]. Briefly, following activation, equilibration and washing of the C18 resin, the elution was done using a step-wise gradient of ACN from 10 to 40 (5% increases) in 0.1% triethylamine to account for the increased hydrophobicity of the XL peptides compared to not modified ones. Resulting fractions were dried under vacuum.

SEC-MS and spectral library acquisition

Samples were resuspended in buffer A (0.1% FA) and approximately 200 ng were analyzed by DIA-PASEF on a Bruker TimsTOFpro interfaced with a

Ultimate3000 UHPLC. For the SEC-MS experiment, the peptides were separated on a PepSep column (15 cm, 150 um IID) using a 38-minute gradient at 0.6 μl/min. Following loading, the peptides were eluted in 20 minutes with a 5% to 30% B (0.1% FA in ACN) in 20 minutes. The column was then washed for 5 minutes at 90% and high flow (1 μl/min) and re-equilibrated at 5% ACN for the next run. The peptides were sprayed through a 20 mm ZDV emitter kept at 1700 V and 200 °C. The mass spectrometer was operated in positive mode using DIA-PASEF acquisition[52]. Briefly, 4 PASEF scans (0.85 $1/K0$ to 1.30 $1/K0$) were acquired and divided each precursor range into 24 windows of 32 Da (500.7502 – 966.67502 m/z) overlapping 1 Da. Each of the fractionated samples (phage-specific libraries) was acquired in DDA-PASEF using a similar gradient composition except for the elution which was performed in 90 minutes leading to a 120 minute gradient. For DDA-PASEF the ion mobility window and precursor range were matched to the DIA boundaries to allow for seamless library building and search.

XL-MS data acquisition

The XL-MS samples were acquired on a Bruker TimsTOFpro interfaced with a Ultimate3000 UHPLC. The peptides were separated using a 118 minutes linear gradient. Following loading, the percentage of B (80% ACN in 0.1% FA) was increased from 2% to 8% in 5 minutes and then to 43% in 90 minutes. Residual peptides were eluted at 50%B for 10 minutes and then the column was washed at 88% B for the remaining 13 minutes. The peptides were separated on a PepSep column (15 cm, 150 mm iid, 1.9 μm beads size). The mass spectrometer was operated in positive mode and data-dependent acquisition with the same source parameters as the SEC fractionated samples. To enrich for crosslinked peptides a custom IM polygon was employed[53] and charge

inclusion was enabled ($3+to8+$ precursors). Precursors having nominal intensity above 20'000 were selected for fragmentation using an inverted collision energy of 23 eV at 0.73 $1/k0$ and 95 eV at 1.6 $1/k0$.

SEC-MS data analysis

The DDA files were searched within the Fragpipe framework using MSfragger[54] and the 'DIA-speclib-quant' workflow using the *Pseudomonas aeruginosa* pan proteome FASTA (5564 entries, proteome ID UP000002438, downloaded on the 05/22). For each phage, the correspondent FASTA nucleotide file was downloaded from GenBank (NC_004629.1 for ϕKZ and NC_028999.1 for ϕPA3) and EMBOSS was used for novel ORFs prediction (see 'Prediction of novel ORFs' section for details). The GenBank files were translated to protein level using BioPython and supplemented to the *Pseudomonas* FASTA. Carbamylation of cysteines was set as fixed modification while oxidation of methionine, N-term acetylation (peptide level) and pyro-glu formation were set as variable modifications. EasyPQP (https://github.com/grosenberger/easypqp) was used to generate a spectral library. Following phage-specific library generation, PAO1 precursors from all libraries were transferred to ensure the presence of the same PAO1 proteins with the same peptides across all DIA experiments using lowess for RT realignment. The DIA-PASEF data was searched with DIA-NN[55] v.1.7.1 using a library-centric approach. Identified spectrum with MS1 precursors within 10 ppm and MS2 precursor within 15 ppm were selected and a second library was generated (double-pass mode). Quantification was set to robust (high-accuracy) and cross-run normalization was disabled.

XL-MS data analysis

XL-MS timsTOF files were converted to mgf using MSconvert. MS1 peak picking was enabled and the spectrum were denoised (top30 peaks in 100 m/z bins). Ion mobility scans were combined. Following the conversion, the peak files were searched in XiSearch[56] using a fraction-specific FASTA containing only the protein ids identified by SEC-MS in the corresponding MW range. MS1 and MS2 tolerances were fixed to 10 and 15 ppm with 10 ppm of peptide tolerance. DSSO was selected as crosslinker (158.0037648 Da) and the correspondent oxidized and amidated crosslinker were added as modifications. Link-FDR was fixed at 5% (boosted) and the resulting file were imported into XiView (https://xiview.org) for manual inspection of crosslinked spectrums.

Data analysis for DDA purified virion samples

TimsTOF DDA files were searched in MSfragger using the LFQ-MBR workflow. Cysteine carbamylation was selected as fixed modification while N-term acetylation and deamidation were enabled as variable modification with a max of 3 variable modifications per peptide. Peptides of length 7 to 50 were searched again a database of phage, *Pseudomonas aeruginosa* plus contaminants. Decoys were generated by pseudo-inversion. Percolator was used for FDR-control at 1% PSM.

Protein-protein interaction prediction from SEC-MS data

DIA-NN report were filtered at 1% library Q-value and, to infer protein quantities, the top2 peptides yielding the highest intra-protein correlation were averaged (sibling peptide correlation strategy). This step was performed across all samples to ensure the same peptides were used for every replicate and condition. The raw MS2 profiles were smoothed using a Savitzky-Golay

filter and rescaled in a 0-1 range. A dot product matrix between all proteins was calculated and protein showing $r^2 \geq 0.3$ were selected as putative interactors for prediction. For every pair we calculated 5 features: (i) sliding window (q=6) correlation, (ii) fraction-wide intensity difference, (iii) peak shift, (iv) Euclidean distance and (v) contrast angle dot-product.

For prediction, we utilized a fully-connected neural network implemented in Tensorflow (https://www.tensorflow.org). Briefly, we set the input layer as number of features (147) followed by a fully connected layer with 100 neurons and a dropout layer (0.2 %) and a fully connected layer with 72 neurons. A final output layer using sigmoid as activation function was used for classifying co-eluting and not-coeluting proteins. For training, a previously reported dataset was used[31]. To select for positive we utilized protein pairs in STRING using a combined score of 0.9 and experimental evidence, while negative were randomly selected. The DNN model was trained for 100 epochs using ADAM (learning rate = 0.001) and binary cross-entropy as loss function. Early stopping (patience = 20) to avoid overfitting. To further removed spuriously co-eluting PPIs after the prediction step, we calculated an equal number of decoy PPIs by randomly sampling the remaining proteins and utilized the DNN model to predict their coelution probability. We then utilized these two distributions to perform target-decoy competition (TDC) using posterior probabilities.

ORFs prediction from nucleotide FASTA

EMBOSS v6.6.0.0 subroutine getorf was used to predict open reading frames (ORFs) with a minimum size of 50 AA. Existing annotated genes were removed from the predicted ORFs using bedtools subroutine subtract, allowing us to differentiate between existing and novel ORFs.

Structural prediction and alignment for ϕPA3 vRNAp

Protein complex prediction was performed using AlphaFold 2 (https://github.com/deepmind/alphafold). AF2 was run with full database size and the multimer preset. OpenMM energy minimization was performed to generate relaxed models and 5 models per complex were generated. Models were ranked by ipTM + TM and the PAE and LDDT were extracted for visualization. Each complex was submitted as a FASTA file, with proteins ordered from the longest to the shortest sequence. The alignment was performed using US-Align[44] (https://zhanggroup.org/US-align/) and the oligomer option was selected. Alignments of predicted complex structures (ϕKZ vRNAp and 4 proteins ϕPA3 vRNAp) were performed by multiple structure alignment (MSTA) using US-align with default parameters and a TM-cutoff of 0.45 was used to estimate topological similarities between the two structures. For visualization purposes, the structure of vRNAp (70GR) without PHIKZ123, which lacked homologs identification in ϕPA3, was used as template in MatchMaker.

Generation of ϕKZ particles packaged with 3xFLAG fusions of ϕKZ virion proteins.

ϕKZ particles packaged with virion proteins bearing a C-terminal 3xFLAG-tag were generated by adapting a protocol used to generate ϕKZ particles packaged with mNeonGreen-tagged inner body proteins[17, 57] . PAO1 cells transformed with the appropriate pHERD30T−(PHIKZxxx)−3xFLAG construct were grown overnight in 3 mL LB supplemented with gentamicin (50 μg/ml) at 37°C with aeration at 175 rpm. Cells were diluted 1:100 from a saturated overnight culture into 5mL LB supplemented with $MgSO_4$ (10mM) and Gentamicin (50 μg/ml) and grown for $\approx$ 2.5 hours at 37°C with aeration at 175 rpm. At OD600nm = 0.5-0.6 (3E8 CFU/mL), the bacterial cultures

were infected with ϕKZ (WT, MOI $\approx$ 1) for 2.5 hours. Thereafter 1 mL of chloroform was added to the cultures in a fume-hood and the cultures were incubated to with chloroform for 15 minutes (37°C, 175 rpm) The cell cultures were transferred to 15 mL falcon tubes and centrifuged at 6000xg for 15 min to pellet bacterial debris. The supernatant (containing bacteriophages in high titer) was carefully transferred to a fresh set of 50mL falcon tubes and centrifuged and 6000xg for 15 min to pellet any residual bacterial debris. Thereafter, 4 mL the supernatant was filtered and concentrated ($\approx$10x) using Amicon-100 centrifugal filters to remove excess 3xFLAG-tagged proteins. The concentrated supernatant was used for western blot experiments.

Western blot and blot analysis

PA01 cells were grown as previously described and upon reaching 0.5 OD (600 nm), gentamicin was added (50 μg/ml) and the celles were chilled on ice for 5 minutes to stall translation. Thereafter PAO1 cells ($\approx$1 OD equivalent) were infected with ϕKZ particles packaged with virion proteins bearing a C-terminal 3xFLAG-tag (MOI $\approx$ 1) on ice for 10 minutes (to allow complete adsorption of virions onto cells) and then incubated at 30℃ for 15 minutes. Thereafter, the cell cultures were transferred to pre-chilled 15mL falcon tubes, centrifuged at 6000xg, 0°C for 5 minutes. The supernatant was discarded and the cell pellet was washed twice with 2 mL of pre-chilled (0°C) LB to remove excess unbound virions. The cell pellet was lysed in 100 μL of lysis buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 2% glycerol, 1% TTX-100, CompleteMini EDTA-free protease inhibitor cocktail). The lysed suspension was further sonicated on ice using a Q125 sonicator (10 pulses, 1s ON, 1s OFF, 30% amplitude). The cell lysate was centrifuged at 15000xg (15min, 0°C) to remove cellular debris. The clarified cellular lysate (100 μL) was boiled with 33 μLL of 4X Laemmli

Buffer (with Beta-mercaptoethanol) for 10 min. 14 μLL of lysate samples was loaded. For virion control samples, 10 μLL of purified virions were boiled with 3.3 μLL of 4X Laemmli Buffer (with Beta-mercaptoethanol) for 10 min and 2 μLL of samples were loaded. SDS-PAGE gels were run with running buffer (100 mL 10X Tris-Glycine SDS Buffer, 900 mL Milli-Q water) at 130V for 1 hour (constant voltage setting). The SDS-PAGE gels were transferred onto 0.2 μM PVDF membranes using a wet transfer (Transfer Buffer: 100 mL 10X Tris-Glycine Buffer, 200 mL methanol, 700 mL Milli-Q water; 100V, 1 hour, 4°C). The membranes were incubated with blocking buffer (5% Omniblock milk, non fat-dry in 1X TBST (200 mL Tris Buffer Saline, 0.20 mL Tween-20)) for 1 hour at room temperature. Thereafter the blocking buffer was discarded and the membranes were incubated with 1:1000 dilutions of mouse anti-FLAG M2 antibody (Sigma-Aldrich) in 1X TBST (overnight, 4°C, with constant shaking). Thereafter the membranes were washed thrice for 10 min with TBST and incubated with 1:3000 dilution of Goat anti-mouse HRP (Thermo Fischer) in blocking buffer for 1 hour at room temperature with constant shaking. Finally, the membranes were washed thrice for 10 min with TBST and incubated with Clarity Western ECL substrate. Membranes were imaged on an Azure 500 imager.

References

[1] Shah, P. S. *et al.* Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175** (7), 1931–1945.e18 (2018). https://doi.org/10.1016/j.cell.2018.11.028 .

[2] Hiatt, J. *et al.* A functional map of HIV-host interactions in primary human T cells. *Nature communications* **13** (1), 1752 (2022). https://doi.

org/10.1038/s41467-022-29346-w .

[3] Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan, N. J. A systems approach to infectious disease. *Nature reviews. Genetics* **21** (6), 339–354 (2020). https://doi.org/10.1038/s41576-020-0212-5 .

[4] Batra, J. *et al.* Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* **175** (7), 1917–1930.e13 (2018). URL https://doi.org/10.1016/j.cell.2018.08.044. https://doi.org/10.1016/j.cell.2018.08.044 .

[5] Hashimoto, Y., Sheng, X., Murray-Nerger, L. A. & Cristea, I. M. Temporal dynamics of protein complex formation and dissociation during human cytomegalovirus infection. *Nature Communications* **11** (1), 806 (2020). URL https://doi.org/10.1038/s41467-020-14586-5. https://doi.org/10.1038/s41467-020-14586-5 .

[6] Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020). https://doi.org/10.1038/s41586-020-2286-9 .

[7] Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594** (7862), 246–252 (2021). URL https://doi.org/10.1038/s41586-021-03493-4. https://doi.org/10.1038/s41586-021-03493-4 .

[8] Meyers, J. M. *et al.* The proximal proteome of 17 SARS-CoV-2 proteins links to disrupted antiviral signaling and host translation. *PLoS Pathogens* **17** (10), 1–30 (2021). URL http://dx.doi.org/10.1371/journal.ppat.1009412. https://doi.org/10.1371/journal.ppat.1009412 .

[9] Dadgostar, P. Antimicrobial resistance: implications and costs. *Infection and Drug Resistance* **12**, 3903–3910 (2019). https://doi.org/10.2147/IDR.S234610 .

[10] Fossati, A. *et al.* PCprophet: a framework for protein complex prediction and differential analysis using proteomic data. *Nature methods* (2021). URL http://www.ncbi.nlm.nih.gov/pubmed/33859439. https://doi.org/10.1038/s41592-021-01107-5 .

[11] Krylov, V. *et al.* Phage phikz—the first of giants. *Viruses* **13** (2), 1–18 (2021). https://doi.org/10.3390/v13020149 .

[12] Monson, R., Foulds, I., Foweraker, J., Welch, M. & Salmond, G. P. The Pseudomonas aeruginosa generalized transducing phage ϕPA3 is a new member of the ϕKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology* **157** (3), 859–867 (2011). https://doi.org/10.1099/mic.0.044701-0 .

[13] Mendoza, S. D. *et al.* A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577** (7789), 244–248 (2020). URL http://dx.doi.org/10.1038/s41586-019-1786-y. https://doi.org/10.1038/s41586-019-1786-y .

[14] Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature microbiology* **5** (1), 48–55 (2020). https://doi.org/10.1038/s41564-019-0612-5 .

[15] Chaikeeratisak, V., Birkholz, E. A. & Pogliano, J. The Phage Nucleus and PhuZ Spindle: Defining Features of the Subcellular Organization and Speciation of Nucleus-Forming Jumbo Phages. *Frontiers in Microbiology* **12** (July), 1–8 (2021). https://doi.org/10.3389/fmicb.2021.641317 .

[16] Chaikeeratisak, V. *et al.* Subcellular organization of viral particles during maturation of nucleus-forming jumbo phage. *Science Advances* **8** (18), 8–9 (2022). https://doi.org/10.1126/sciadv.abj9670 .

[17] Li, Y. *et al.* A family of novel immune systems targets early infection of nucleus-forming jumbo phages. *bioRxiv* 2022.09.17.508391 (2022). URL http://biorxiv.org/content/early/2022/09/18/2022.09.17.508391.abstract. https://doi.org/10.1101/2022.09.17.508391 .

[18] Fossati, A. *et al.* System-Wide Profiling of Protein Complexes Via Size Exclusion Chromatography-Mass Spectrometry (SEC-MS). *Methods in molecular biology (Clifton, N.J.)* **2259**, 269–294 (2021). https://doi.org/10.1007/978-1-0716-1178-4_18 .

[19] Skinnider, M. A. & Foster, L. J. Meta-analysis defines principles for the design and analysis of co-fractionation mass spectrometry experiments. *Nature Methods* **18** (7), 806–815 (2021). URL https://doi.org/10.1038/s41592-021-01194-4. https://doi.org/10.1038/s41592-021-01194-4 .

[20] Caufield, J. H., Abreu, M., Wimble, C. & Uetz, P. Protein Complexes in Bacteria. *PLOS Computational Biology* **11** (2), 1–23 (2015). URL https://doi.org/10.1371/journal.pcbi.1004107. https://doi.org/10.1371/journal.pcbi.1004107 .

[21] Lawrence, J. G. Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* **110** (4), 407–413 (2002). https://doi.org/10.1016/s0092-8674(02)00900-5 .

[22] Qiao, Z. *et al.* Cryo-EM structure of the entire FtsH-HflKC AAA protease complex. *Cell reports* **39** (9), 110890 (2022). https://doi.org/10.1016/j.celrep.2022.110890 .

[23] Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. Crystal structure of the processivity clamp loader gamma (γ) complex of E. coli DNA polymerase III. *Cell* **106** (4), 429–441 (2001). https://doi.org/10.1016/S0092-8674(01)00463-9 .

[24] Sutherland, I. W., Hughes, K. A., Skillman, L. C. & Tait, K. The interaction of phage and biofilms. *FEMS Microbiology Letters* **232** (1), 1–6 (2004). URL https://www.sciencedirect.com/science/article/pii/S0378109704000412. https://doi.org/https://doi.org/10.1016/S0378-1097(04)00041-2 .

[25] Silpe, J. E. & Bassler, B. L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* **176** (1-2), 268–280.e13 (2019). https://doi.org/10.1016/j.cell.2018.10.059 .

[26] Schiessl, K. T. *et al.* Phenazine production promotes antibiotic tolerance and metabolic heterogeneity in Pseudomonas aeruginosa biofilms. *Nature Communications* **10** (1), 762 (2019). URL https://doi.org/10.1038/s41467-019-08733-w. https://doi.org/10.1038/s41467-019-08733-w .

[27] Das, T. *et al.* Phenazine virulence factor binding to extracellular DNA is important for Pseudomonas aeruginosa biofilm formation. *Scientific reports* **5**, 8398 (2015). https://doi.org/10.1038/srep08398 .

[28] Andrésen, C. *et al.* Critical biophysical properties in the Pseudomonas aeruginosa efflux gene regulator MexR are targeted by mutations conferring multidrug resistance. *Protein Science* **19** (4), 680–692 (2010). https://doi.org/10.1002/pro.343 .

[29] Nazir, A., Ali, A., Qing, H. & Tong, Y. Emerging Aspects of Jumbo Bacteriophages. *Infection and drug resistance* **14**, 5041–5055 (2021). https://doi.org/10.2147/IDR.S330560 .

[30] Chaikeeratisak, V. *et al.* The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages. *Cell Reports* **20** (7), 1563–1571 (2017). URL http://dx.doi.org/10.1016/j.celrep.2017.07.064. https://doi.org/10.1016/j.celrep.2017.07.064 .

[31] Salas, D., Stacey, R. G., Akinlaja, M. & Foster, L. J. Next-generation interactomics: Considerations for the use of co-elution to measure protein interaction networks. *Molecular and Cellular Proteomics* **19** (1), 1–10 (2020). https://doi.org/10.1074/mcp.R119.001803 .

[32] Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150** (5), 1068–1081 (2012). URL http://www.ncbi.nlm.nih.gov/pubmed/22939629{%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3477804. https://doi.org/10.1016/j.cell.2012.08.011 .

[33] Yakunina, M. *et al.* A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. *Nucleic Acids Research* **43** (21), 10411–10420 (2015). https://doi.org/10.1093/nar/gkv1095 .

[34] Ceyssens, P.-J. *et al.* Development of Giant Bacteriophage KZ Is Independent of the Host Transcription Apparatus. *Journal of Virology* **88** (18), 10501–10510 (2014). https://doi.org/10.1128/jvi.01347-14 .

[35] Lenz, S. *et al.* Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nature Communications* **12** (1), 1–11 (2021). URL http://dx.doi.org/10.1038/s41467-021-23666-z. https://doi.org/10.1038/s41467-021-23666-z .

[36] Réblová, K., Sponer, J. & Lankas, F. Structure and mechanical properties of the ribosomal L1 stalk three-way junction. *Nucleic acids research* **40** (13), 6290–6303 (2012). https://doi.org/10.1093/nar/gks258 .

[37] Maruyama, K. *et al.* Switch of the interactions between the ribosomal stalk and EF1A in the GTP- and GDP-bound conformations. *Scientific Reports* **9** (1), 14761 (2019). URL https://doi.org/10.1038/s41598-019-51266-x. https://doi.org/10.1038/s41598-019-51266-x .

[38] Häuser, R. *et al.* RsfA (YbeB) proteins are conserved ribosomal silencing factors. *PLoS genetics* **8** (7), e1002815 (2012). https://doi.org/10.1371/journal.pgen.1002815 .

[39] Kramer, G. *et al.* L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419** (6903), 171–174 (2002). URL https://doi.org/10.1038/nature01047. https://doi.org/10.1038/nature01047 .

[40] Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics* **11** (6), O111.016717 (2012). URL http://www.mcponline.org/lookup/doi/10.1074/mcp.O111.016717. https://doi.org/10.1074/mcp.O111.016717 .

[41] Orekhova, M., Koreshova, A., Artamonova, T., Khodorkovskii, M. & Yakunina, M. The study of the phiKZ phage non-canonical non-virion RNA polymerase. *Biochemical and Biophysical Research Communications* **511** (4), 759–764 (2019). URL https://www.sciencedirect.com/science/article/pii/S0006291X1930333X. https://doi.org/https://doi.org/10.1016/j.bbrc.2019.02.132 .

[42] de Martín Garrido, N. *et al.* Structure of the bacteriophage PhiKZ non-virion RNA polymerase. *Nucleic Acids Research* **49** (13), 7732–7739 (2021). URL https://doi.org/10.1093/nar/gkab539. https://doi.org/10.1093/nar/gkab539 .

[43] Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* (2021). URL https://www.biorxiv.org/content/early/2021/10/04/2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034 .

[44] Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods* (2022). URL https://doi.org/10.1038/s41592-022-01585-1. https://doi.org/10.1038/s41592-022-01585-1 .

[45] Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7** (1), 339 (2006). URL https://doi.org/10.1186/

1471-2105-7-339. https://doi.org/10.1186/1471-2105-7-339 .

[46] Thomas, J. A. *et al.* Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage ϕKZ. *Molecular Microbiology* **84** (2), 324–339 (2012). https://doi.org/10.1111/j.1365-2958.2012.08025.x .

[47] Bailey, T. L. STREME: Accurate and versatile sequence motif discovery. *Bioinformatics (Oxford, England)* **37** (18), 2834–2840 (2021). https://doi.org/10.1093/bioinformatics/btab203 .

[48] Hu, G. *et al.* flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications* **12** (1), 4438 (2021). URL https://doi.org/10.1038/s41467-021-24773-7. https://doi.org/10.1038/s41467-021-24773-7 .

[49] Lecoutere, E. *et al.* Identification and comparative analysis of the structural proteomes of phiKZ and EL, two giant Pseudomonas aeruginosa bacteriophages. *Proteomics* **9** (11), 3215–3219 (2009). https://doi.org/10.1002/pmic.200800727 .

[50] Wu, W., Thomas, J. A., Cheng, N., Black, L. W. & Steven, A. C. Bubblegrams reveal the inner body of bacteriophage ϕKZ. *Science (New York, N.Y.)* **335** (6065), 182 (2012). https://doi.org/10.1126/science.1214120 .

[51] Fossati, A. *et al.* Toward Comprehensive Plasma Proteomics by Orthogonal Protease Digestion. *Journal of proteome research* (20), 4031–4040 (2021). https://doi.org/10.1021/acs.jproteome.1c00357 .

[52] Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods* **17** (12),

1229–1236 (2020). URL https://doi.org/10.1038/s41592-020-00998-0. https://doi.org/10.1038/s41592-020-00998-0 .

[53] Steigenberger, B. *et al.* Benefits of Collisional Cross Section Assisted Precursor Selection (caps-PASEF) for Cross-linking Mass Spectrometry. *Molecular & cellular proteomics : MCP* **19** (10), 1677–1687 (2020). https://doi.org/10.1074/mcp.RA120.002094 .

[54] Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods* **14** (5), 513–520 (2017). https://doi.org/10.1038/nmeth.4256 .

[55] Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **17** (1), 41–44 (2020). URL http://dx.doi.org/10.1038/s41592-019-0638-x. https://doi.org/10.1038/s41592-019-0638-x .

[56] Mendes, M. L. *et al.* An integrated workflow for crosslinking mass spectrometry. *Molecular Systems Biology* **15** (9), e8994 (2019). URL https://www.embopress.org/doi/abs/10.15252/msb.20198994. https://doi.org/https://doi.org/10.15252/msb.20198994 .

[57] Guan, J. *et al.* Bacteriophage genome engineering with CRISPR-Cas13a. *Nature microbiology* **7** (12), 1956–1966 (2022). https://doi.org/10.1038/s41564-022-01243-4 .