

001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046

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

Andrea Fossati^{1,2,3}, Deepto Mozumdar⁴, Claire Kokontis⁴, Eliza Nieweglowska⁴, Melissa Mendez⁴, Adrian Pelin^{1,2,3}, Yuping Li⁴, Baron Guo⁴, Nevan J. Krogan^{1,2,3}, David A. Agard⁴, Joseph Bondy-Denomy^{4*} and Danielle L. Swaney^{1,2,3*}

¹J. David Gladstone Institutes, San Francisco, 94158, California,
USA.

²Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, 94158, California, USA.

³Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, 94158, California, USA.

⁴Department 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 (ϕ KZ and ϕ PA3) 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 ϕ PA3 complex showing strong structural and sequence similarity to ϕ KZ nvRNAP, suggesting ϕ PA3 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 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

093
094
095
096
097
098
099
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138

(ϕ KZ[11] and ϕ PA3[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.

160

161

162

163

164 Results

165

166

167 A cross-phage study of viral infection cycle

168

169 To understand HPIs that mediate phage infection, we infected *Pseudomonas*
170 *Aeruginosa* (strain PAO1) with either the ϕ KZ or ϕ PA3 bacteriophage for
171 60 minutes in biological duplicate. To control for virion protein complexes
172 (i.e complexes present within the phage itself), parallel experiments were also
173 performed using a naturally emerging PAO1 mutant ($\Delta fliC$)[17] that cannot
174 be infected (Fig.1A).

175

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

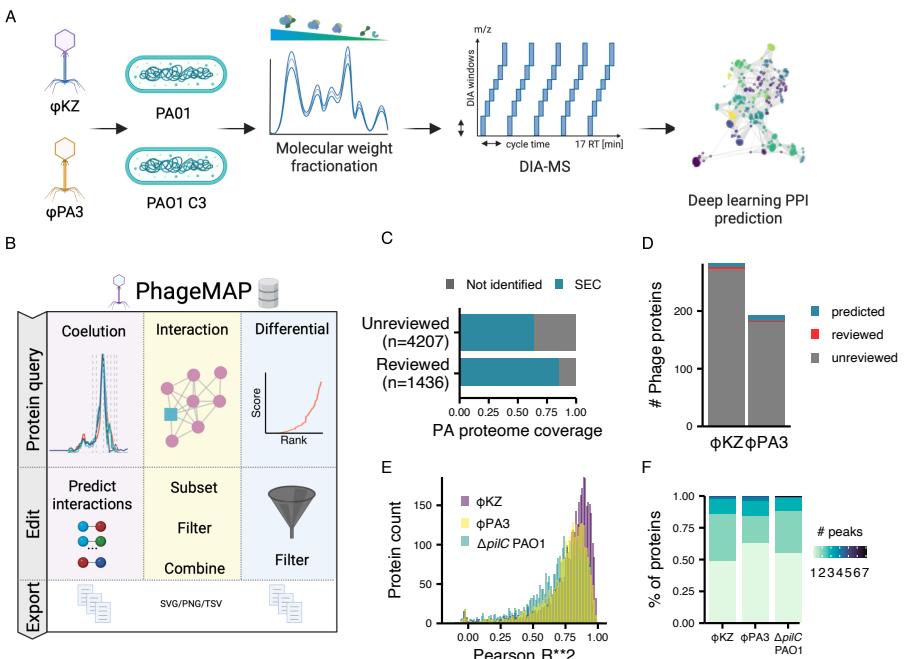


Fig. 1. High-throughput interaction proteomics for deep host-pathogen interaction mapping **A.** SEC-MS workflow and experimental design. **B.** Overview of PhageMAP analysis and workflows. **C.** Recovery of *Pseudomonas* proteome by SEC-MS. **D.** Barplot representing the number of phage proteins identified. **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).

231 This experimental workflow resulted in the high-throughput and compre-
232 hensive coverage of both the bacterial and the phage proteomes. Specifically,
233 we detected 3782 PAO1 proteins, covering 83% of the validated SwissProt
234 entries for the *Pseudomonas* pan-proteome, and 67% of the unreviewed entries
235 (Fig.1B). Likewise, we detected 280 proteins for ϕ KZ and 198 proteins for
236 ϕ PA3, covering 75% and 53% of their proteomes, respectively (Fig.1C).

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

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 reproducibility in both phage infection and SEC fractionation, with most of the SEC-profile peaks overlapping within 1-2 fractions ($<0.250 \mu\text{L}$). 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 (Fig.1F). 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

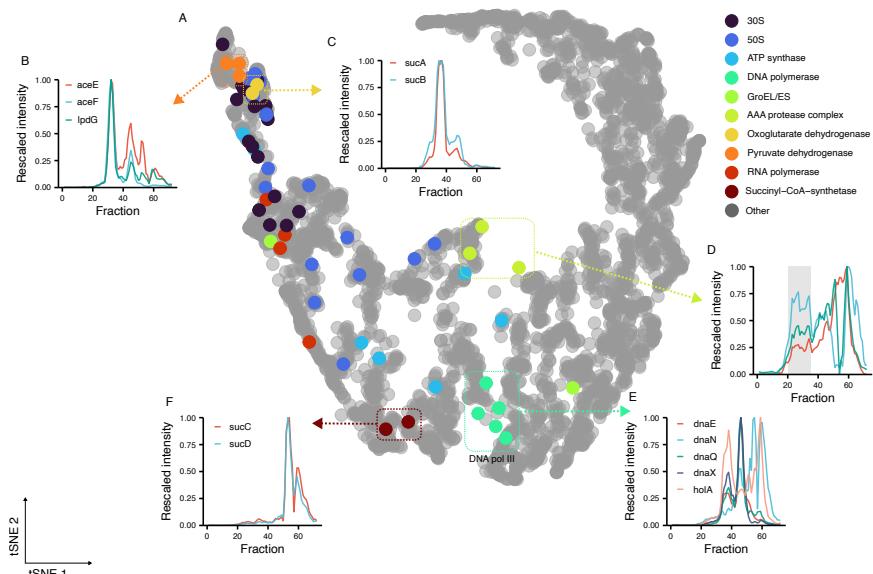


Fig. 2. *Pseudomonas* protein complexes identified in the SEC-MS data

A. t-SNE plot for all the *P. aeruginosa* proteins detected in the $\Delta fliC$ 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

323 pyruvate dehydrogenase complex (Fig.2B) and the oxoglutarate dehydroge-
324 nase complex (Fig.2C), which migrated at an estimated MW of $\approx 3.5 * 10^6$ Da
325 (expected MW $\approx 3.75 * 10^6$ Da) and $\approx 2.4 * 10^6$ Da, respectively. It is impor-
326 tant to point that out that the molecular weight estimation for these large
327 assemblies is subject to error due to these peaks being outside the external
328 calibration curve. To achieve MW estimation we included in the calibration
329 curve a pure SEC-separated 70S ribosome (Supplementary Fig S2).
330
331 Our sample preparation also preserved membrane-bound complexes. As exam-
332 ple, the AAA protease complex, formed by four hexamers of the AAA
333 protease (ftsH) and 12 copies of each single-pass membrane proteins (HflK
334 and HflC)[22], was recovered at high molecular weight in a broad peak as
335 shown in Fig.2D. The large molecular weight range and sensitivity covered by
336 our separation approach was also demonstrated in the recovery of more tran-
337 sient complexes such as the DNA polymerase III (dnaA, dnaE, and dnaQ)
338 loaded with the γ complex (holA and dnaX) which plays a key role at the
339 replication fork[23] (Fig.2E). Finally, heterodimeric complexes such as the
340 succinyl-coA synthetase were also recovered as demonstrated by the coelution
341 plot in (Fig.2F). Our manual inspections further confirms that prior knowl-
342 edge can be easily incorporated into SEC-MS data analysis and allows for
343 straight-forward identification of protein complexes.
344
345
346
347
348
349
350
351
352
353
354
355

356
357
358
359
360
361
362
363
364
365
366
367
368

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

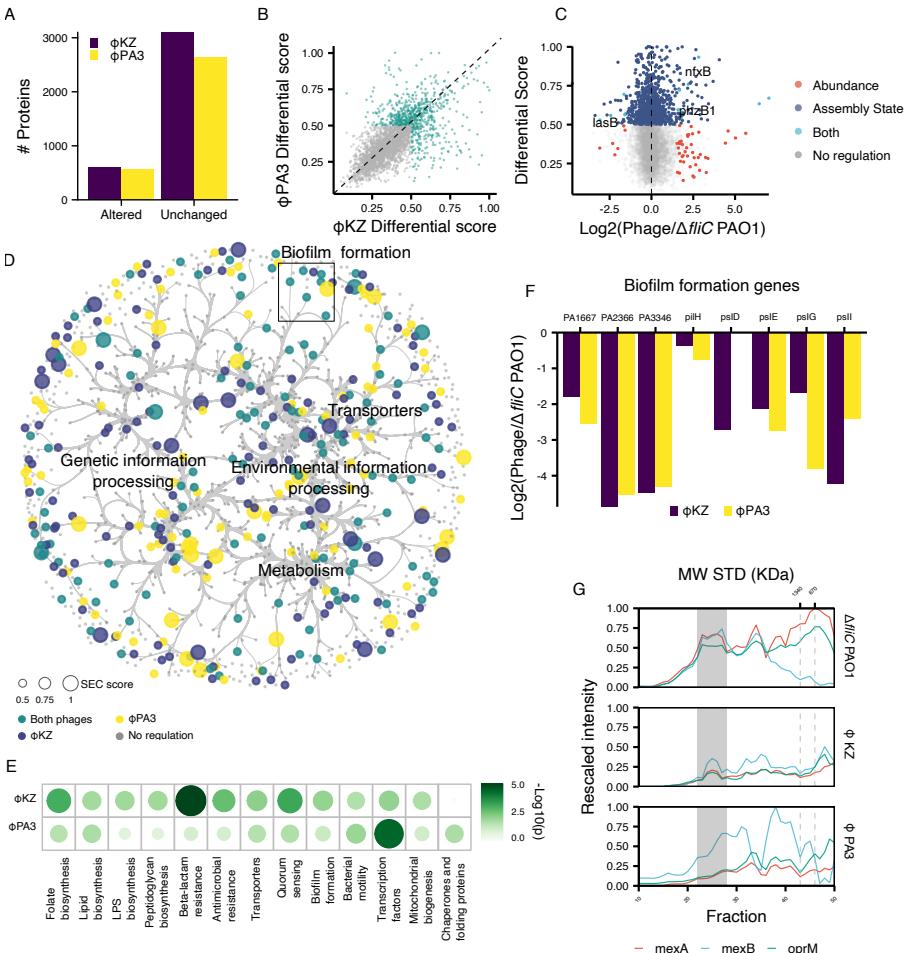


Fig. 3. Differential analysis of SEC-MS data A. Altered host proteins for each experiment. B. Scatterplot of differentially regulated proteins. Axis represents the differential SEC score while each dot represents a PAO1 protein. Color indicates significant regulation in either phage. C. 2D plot of differential 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 ($\text{Log}2\text{FC} \geq 2$ and $q \leq 1\%$) and assembly state level ($\text{SEC score} \geq 0.5$). 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 ($\Delta fliC$) experiments. G. Coelution plot for the efflux pump MexA/B-oprM. Different experiments are represented by the various subpanels.

415 After having demonstrated the proteome depth achieved in our SEC-MS
416 dataset and the recovery of known complexes, we turned our attention to how
417 Jumbophages re-wire *Pseudomonas* protein complexes by evaluating differ-
418 ences in SEC profiles upon phage infection. Variation in SEC profiles between
419 conditions can arise from differential assembly state (i.e. a protein profile
420 shifting to higher or lower molecular weight), different stoichiometry within a
421 complex, or global alterations in protein abundance. To quantify these differ-
422 ent cases, we employed a previously described Bayesian analysis module from
423 the PCprophet package[10] to derive marginal likelihoods (SEC differential
424 score) of protein-level SEC changes between ϕ KZ and ϕ PA3 versus the recep-
425 torless infected samples (i.e. $\Delta fliC$). Comparing the SEC-profile differences
426 between phage-infected PAO1 and $\Delta fliC$ revealed approximately 600 proteins
427 showing SEC variation upon infection by either phage (Fig.3A). Notably, there
428 is substantial consistency in which *Pseudomonas* proteins are altered, and
429 the degree of change in their individual SEC profiles (Fig.3B, cor = 0.677),
430 potentially pointing towards common pathways and complexes hijacked by
431 ϕ PA3 and ϕ KZ for successful predation. Most of the changes at the assem-
432 bly state level do not have a corresponding variation in protein abundance
433 at the global proteome level, suggesting that SEC-MS offers an orthogonal
434 view on effect of perturbations, such as infection, on the proteome (Fig.3C).
435 Proteins changing both in assembly state and abundance could be dependent
436 on deep interactome rewiring rooted in a strong transcriptional response to
437 phage infection. For example, we identified several quorum sensing proteins
438 such as lasB, phzB1, and nfxB in ϕ KZ and ϕ PA3 as proteins changing in both
439 dimensions (SEC and abundance). To identify conserved KZ-like jumbophage
440 manipulation of the host interactome, we mapped the SEC-derived PAO1
441 interaction network (Fig.3D) with the correspondent protein-level differential
442

data derived from the comparison between phage and uninfected samples. 461
Albeit a large portion of the nodes do not have a functional annotation, 462
we identified several classes where their component were significantly altered 463
upon Jumbophage phage infection. Several pathways related to drug resistance 464
(folate biosynthesis KEGG ID pae00790, beta lactam resistance KEGG ID 465
pae01501, antimicrobial resistance KEGG ID pae01504) were enriched upon 466
Jumbophage infection (Fig.3E). Moreover, quorum sensing and biofilm for- 467
mation pathways were enriched in both phage infected samples ($q \leq 0.01$) 468
and several prior studies have highlighted the role of phages in regulating 469
formation of biofilms[24, 25]. We identified multiple proteins in this category 470
having significantly decreased abundance in the high molecular weight region 471
compared to their uninfected counterpart (Fig.3F) which adds further evi- 472
dence to alteration of quorum sensing as the phenazine system directly triggers 473
biofilm formation in *Pseudomonas*[26, 27]. Membrane proteins were partic- 474
ularly affected by Jumbophage infection with porins and multi-drug efflux 475
proteins (KEGG pae02010: ABC transporters) displaying significant reduction 476
in interactions. As example, the MexAB-OprM complex, a key efflux pump[28], 477
shows almost complete reduction of the fully assembled complex (Fig.3G). It 478
is important to point out that changes we observed could either be beneficial 479
for the phage to overcome its host, or an attempt from the host to eliminate 480
the phage. 481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506

507 Organization of the ϕ KZ like Jumbophages viral
 508 interactome
 509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

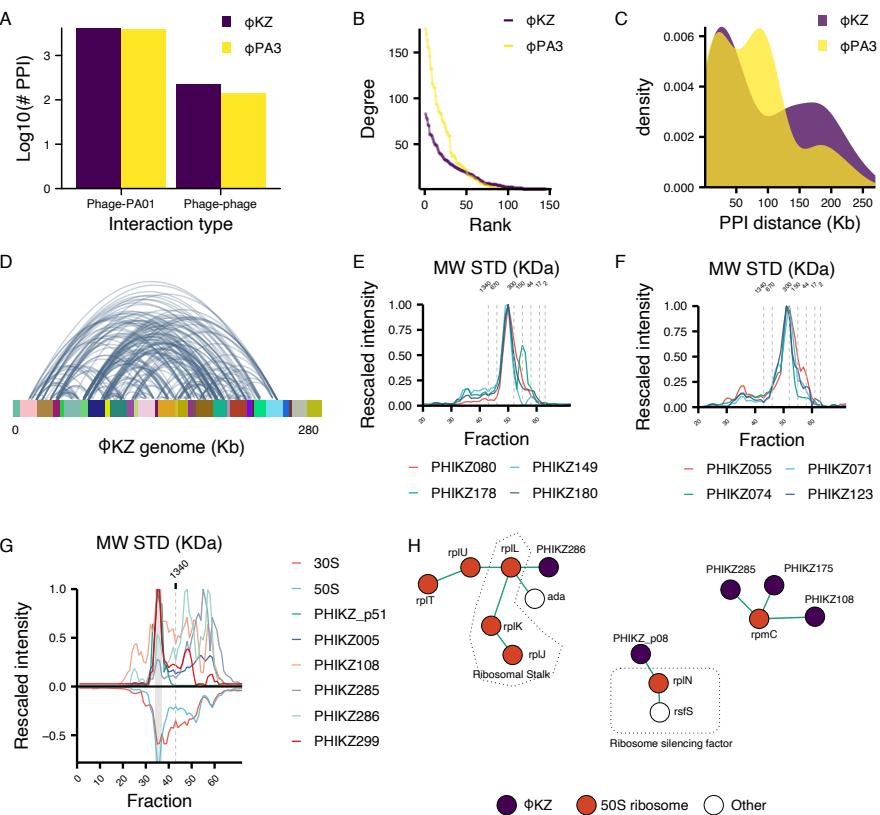


Fig. 4. Comparative analysis of ϕ KZ and ϕ PA3 interaction networks **A.** Number of PPIs identified for each phage (Y axis), separated by interaction type. **B.** Degree distribution (Y axis) versus the rank for a particular node (X axis). Line color represents the various phages. **C.** Density plot illustrating the genomic distance in Kb (X axis) for the phage-phage interactions. Color codes represent the two phages. **D.** Arcplot of ϕ KZ 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 ϕ KZ non-virion associated RNA-polymerase and the virion associated RNA-polymerase. **G.** Mirrorplot illustrating the coelution of ϕ KZ proteins (upper panel) with the 70S ribosome (lower panel). **H.** Interaction network for ϕ KZ 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, ϕ KZ and ϕ PA3 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 ≥ 0.75 . In total, we identified 292 interactions between pairs of ϕ KZ viral proteins, and 6550 host-pathogen interactions between ϕ KZ and PA01 proteins. ϕ PA3 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 ϕ KZ, 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 ϕ KZ and ϕ PA3, interactions are predicted between translation products of phage genes separated by distances as large as 250Kb. For ϕ KZ the density displays a bimodal distribution with a sharp peak at ≈ 22 Kb and a broader peak at ≈ 185 Kb (Fig.4C, purple). Interestingly, for ϕ PA3 this distribution is altered, as it shows a greater distribution

553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598

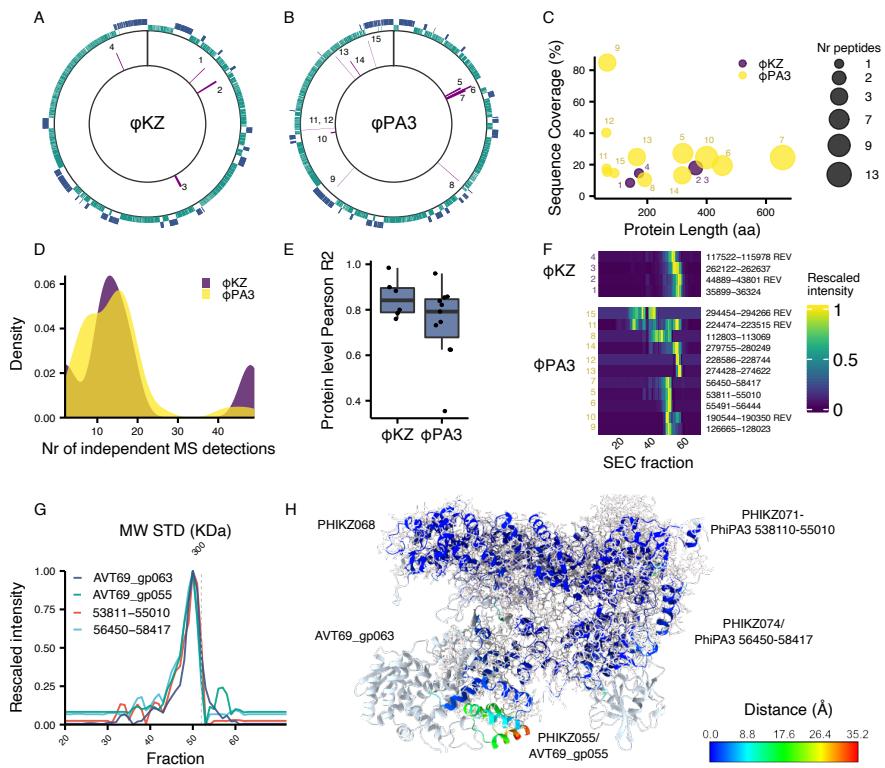
599 of PPI predicted at lower gene separations (two peaks at \approx 19 Kb and \approx 97
600 Kb), that tapers off at larger gene separations (with a third peak at \approx 182
601 Kb) (Fig.4C, yellow). This discrepancy between the two phages is explained
602 in part by the fact that as compared to ϕ KZ, in ϕ PA3 partial duplication or
603 fragmentation of genes into multiple genes within a single/adjacent operon(s)
604 is frequently observed (see Supplementary Table S1). Overall, the PPI distri-
605 bution confirms the general lack of synteny within the genomes of ϕ KZ-like
606 jumbophages and shows the SEC-MS approach as a particularly advantageous
607 technique to query phage encoded complexes at the protein, level agnostic to
608 the overall genome organization (i.e. a guilt-by-association approach at the
609 protein level).

610 The identified interactions allowed us to recapitulate several known
611 complexes in the Jumbophage proteome, despite a limited number being
612 described at present. As example, we recovered the non-virion associated RNA-
613 polymerase[33] migrating at its expected molecular weight (apparent MW 271
614 KDa or fraction 53, correct MW \approx 265 kDa) (Fig.4E) as well as the virion
615 associated RNA polymerase[34] (apparent MW 300 KDa or fraction 52, correct
616 MW \approx 297 kDa) (Fig.4F). Overall, we observed the presence of several phage
617 peaks groups at high molecular weight in both ϕ PA3 and ϕ KZ, suggesting the
618 presence of yet undiscovered phage protein complexes or phage-host interac-
619 tions (Supplementary Fig.S3). For example, we observed a clear peak group
620 composed of PHIKZ005, PHIKZ108, PHIKZ285, PHIKZ286, PHIKZ299, and
621 PHIKZ_p51 at approximately 4 MDa. Intriguingly, these proteins were pre-
622 dicted to be in complex with the fully assembled *P. aeruginosa* 70S ribosome
623 as shown in Fig.4G. To validate these ϕ KZ proteins as ribosomal interac-
624 tors, we performed cross-linking mass spectrometry (XL-MS)[35] on a pooled
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644

sample from the SEC fractions corresponding 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 ϕ KZ proteins. Amongst the identified ϕ KZ 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.

645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690

691 Identification of novel protein-coding sequences by
 692 SEC-MS
 693



722 **Fig. 5. Identification of novel phage proteins.** **A-B.** CDS plot for ϕKZ and
 723 ϕPA3 . Forward CDS are colored in green while reverse CDS are represented in
 724 purple. Identified novel proteins are highlighted in the histogram (inner circle). **C.**
 725 Scatterplot of protein length vs percentage of sequence coverage in the SEC-MS
 726 experiments. Dot size represents the number of proteotypic peptides identified. **D.**
 727 Distribution of the identifications (defined as number of independent MS detections
 728 using a 1% peptide spectrum matching FDR) for the novel ϕKZ and ϕPA3
 729 proteins. **E.** Boxplot showing the Pearson correlation between the two replicates
 730 ($n=72$). Each novel protein is represented as a dot. The box boundaries show the
 731 interquartile range (IQR) and its whiskers $1.5 \times \text{IQR}$. **F.** Heatmap representing the
 732 elution profile for all the novel ORFs. X axis represents the fraction number while
 733 the cell color shows the unit-rescaled intensity. **G.** Coelution profile for predicted
 734 nvRNAP in ϕPA3 . **H.** Superimposition of reported structure for the ϕKZ nvRNAP
 735 (stick, grey) and predicted structure for the ϕPA3 nvRNAP (ribbons). Chains are
 736 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 ϕ KZ (2 forward and 2 reverse ORFs) and 11 for ϕ PA3 (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 ϕ PA3 proteins and 0.82 for ϕ KZ 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 ϕ PA3 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 ϕ PA3 of the reverse sense ORF 56450-58417 which shows >70% sequence similarity with previously reported proteins from various *Pseudomonas spp* phages (ϕ KZ, Psa21, Phabio, 201 ϕ 2-1, 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 ϕ KZ gp74 (non-virion associated RNAP, UniprotID Q8SD88)[34, 41, 42]. To date, there is no experimental evidence

783 of a nvRNAP in ϕ PA3. To derive other putative members of this complex,
784 we extracted the predicted interactors of ORF 56450-58417 and performed
785 BLASTp analysis to identify proteins showing homology to other Jumbophage
786 RNA polymerase components. From this analysis, we selected 3 interactors
787 (gp55, gp63, and the novel ORF 53811-55010) showing >50% conservation
788 with multiple Jumbophage proteins annotated as RNAP components (Sup-
789 plementary Fig.S5B-D). Specifically, we identified homologs of both the β^1
790 polymerase subunit (gp55 and ORF 56450-58417), as well as homologs of the
791 β subunit (ORF 53811-55010). The ϕ PA3 protein gp63 displays 57% homology
792 to ϕ KZ gp68, an essential nvRNAP component which lacks structural simi-
793 larity to known components of previously reported RNA polymerases[41]. By
794 employing the molecular weight derived from the SEC peak position, we esti-
795 mated the nvRNAP MW in ϕ PA3 being \approx 321 KDa (Fig.5G). Assuming the
796 lack of homodimers in the structure, the predicted MW for these four proteins
797 was \approx 253 KDa, suggesting a putative missing subunit. Of note, we did not
798 identify 53811-55010 interactors corresponding to ϕ KZ g123, another β subunit
799 component, which could explain this observation. Nonetheless, to explore the
800 possibility of these proteins (gp55, gp63, ORF 53811-55010, and ORF 53811-
801 55010) folding into an RNA polymerase-like assembly we performed structural
802 prediction of this peak group using AlphaFold2 multimer[43]. We achieved high
803 prediction confidence, with inter-chain predicted confidence (ipTM + pTM) of
804 0.82 and low predicted aligned error (PAE) across the entire oligomer (Supple-
805 mentary Fig.S6). We aligned the best scoring model to the reported structures
806 for the ϕ KZ nvRNAP (PDB 70GP and 70GR)[42] as depicted in Fig.5H. We
807 reached a template modeling (TM) score of 0.503 using US-Align[44] and an
808 average RMSD of 1.016 Å using MatchMaker[45] between our proposed ϕ PA3
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828

vRNAP and the ϕ KZ 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 ϕ PA3 and gp68 in ϕ KZ). 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 ϕ PA3 gp63 following AlphaFold prediction.

829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874

875 Discovery and validation of novel injected phage proteins

876

877

878

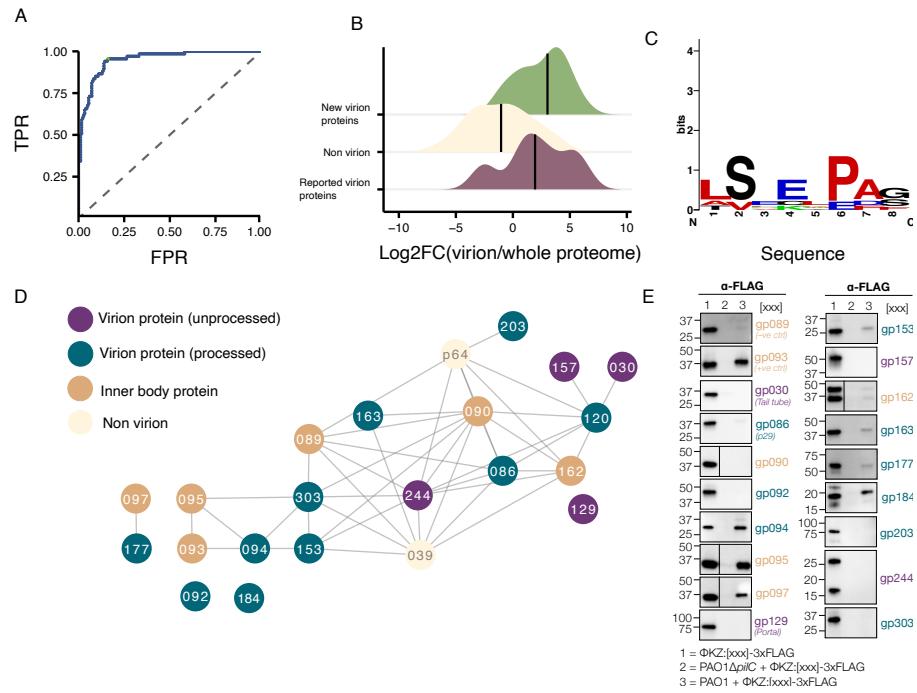


Fig. 6. Data-driven analysis of injected inner body proteins **A.** ROC curve for virion MS ($AUC \approx 0.94$) using as ground truth the prior reported virion proteins. Green highlight selected threshold for maximum sensitivity at the lowest FPR. **B.** Density plot representing the enrichment of virion 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.

912

913

914

915 Although it is well known that ϕKZ phages guard their genome from
 916 nucleolytic host-immune systems by building a proteinaceous shell[13], this
 917 structure is only visible after 20 minutes of infection. Little is known about
 918 how the phage genome is protected or packaged prior to shell assembly. In
 919

920

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 ϕ KZ coupled with deep peptide fractionation and long chromatographic acquisition (see Supplementary Methods for details). The 245 ϕ KZ 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 921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966

967 sequences within our set of IB interactors we performed motif enrichment
968 analysis using STREME[47]. In this analysis, we employed as input the
969 reported IB proteins (gp93/gp95/gp97) and the SEC-MS derived interactors
970 after filtering using the virion proteome (gp080, gp094, gp177, gp235, gp237,
971 and gp303). We identified the consensus motif LSxE as enriched (BH-adjusted
972 $p = 6e^{-5}$) which confirms the previously reported one S/A/G-X-E motif,
973 while providing additional specificity in the P2 position (Fig6C).

978

979

980 We then queried the ϕ KZ interaction network using as input the virion pro-
981 teome, to identify putative injected proteins (Fig. 6D). Building on this data,
982 we selected the interactors of the previously reported proteins (gp94, gp153,
983 gp162, gp163, and gp177) for further validation using our previously reported
984 assay for evaluating injection[17]. By further lowering the interaction thresh-
985 olds to all positive predicted interactions (i.e PPI probability ≥ 0.5 instead of
986 0.75 utilized to select high-confidence interactors), we further identified gp184
987 as IB interactor and validated it as injected proteins. These experiments, con-
988 firmed injection of the previously reported IB proteins (gp93, gp95, gp97) and
989 furthermore validated the injection of all their interactors as showcased in Fig.
990 6E. Prior assignment of IB proteins was based on an arbitrary cutoff[46]. Here
991 by using more sensitive MS of the virion combined with SEC-MS, we identify
992 a full new set of proteins that are highly abundant, found in the virion, inter-
993 act with the previously reported IB proteins and herein we validate injection
994 of 8 proteins (3 previously reported). Overall, these proteins give us a start-
995 ing point to unravel the interactome of the ejected phage genome and identify
996 proteins that protect the genome from host nucleases.

997

998

999

1000

1001

1002

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 ϕ KZ and ϕ PA3) 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.

1059 Moreover, while the lack of immediate genome organization hinders the pre-
1060
1061 diction of functions for phage proteins, the deep coverage and unbiased nature
1062 of SEC-MS data offers a straightforward approach to identify novel complexes
1063
1064 and propose putative functions. As example, by using SEC-derived interac-
1065 tors of a de-novo predicted ϕ PA3 protein (ORF 56450-58417), we identified an
1066 heterotetrameric assembly which we demonstrated bearing strong structural
1067 similarity to the reported nvRNAP in ϕ KZ. This suggests that the unbiased
1068
1069 nature of SEC-MS data allows to not only to identify uncharacterized, novel
1070 proteins, but also to probe their putative function through their interactions
1071
1072 in a structurally defined complex that could be further investigated with struc-
1073 tural biology approaches. Beside identification of interactions, these maps offer
1074
1075 the opportunity to further quantify host interactome remodelling and dis-
1076 entangle variation in expression from assembly state. By comparing the *P.*
1077
1078 *aeruginosa* interactome between infected and uninfected, we observed a large
1079 degree of changes during infection, with perturbation of similar complexes
1080
1081 between the two jumbophages suggesting conserved mechanisms of phage pre-
1082 dation. While we provided a first draft of the KZ-like jumbophage interactome
1083
1084 it is important to point out the trade-off between specificity and through-
1085
1086 put in interaction identification, which we tried to mitigate by utilizing only
1087 high-confidence interactions for analysis. Advances in deep learning models for
1088 prediction of interactions from co-fractionation mass spectrometry data and
1089 integration of orthogonal features (beside the coelution itself) such as pre-
1090
1091 dicted structure or function is expected to improve prediction accuracy and
1092 reduce false discovery rate for uncharacterized proteomes. Overall, the char-
1093
1094 acterization of host-pathogen molecular networks remains challenging, but we
1095
1096 provided the first interactome-wide study of infection progression using two
1097
1098 models ϕ KZ-like phages in *Pseudomonas aeruginosa*.
1099
1100
1101
1102
1103
1104

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 ϕ PA3 and ϕ KZ 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

1151 prediction DM. CK. BG: Phage infection experiments, virion enrichment,
1152
1153 microscopy and WB for injected proteins EM, MM: Performed shell enrich-
1154 ment YP: Critical input in revising the manuscript All co-authors contributed
1155
1156 in reviewing and editing the manuscript.

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187

1188

1189

1190

1191

1192

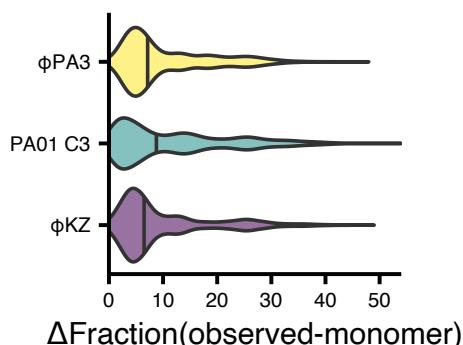
1193

1194

1195

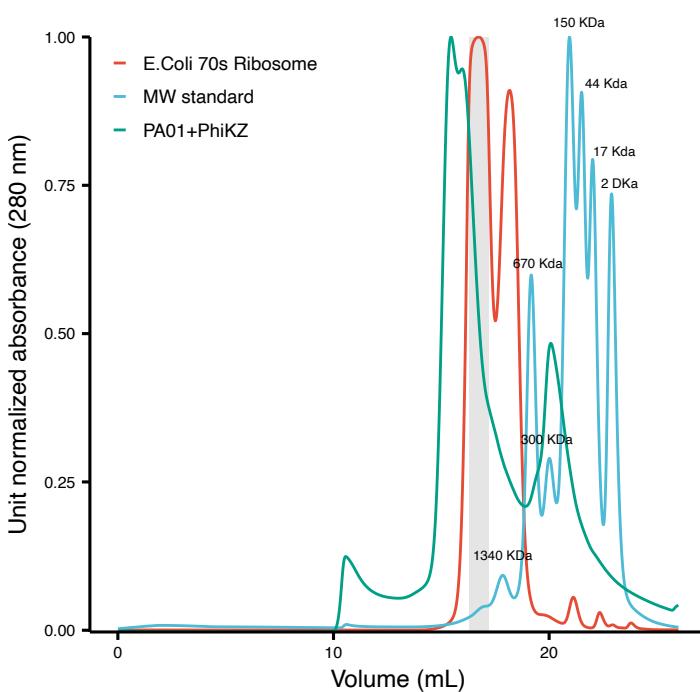
1196

Supplementary figures

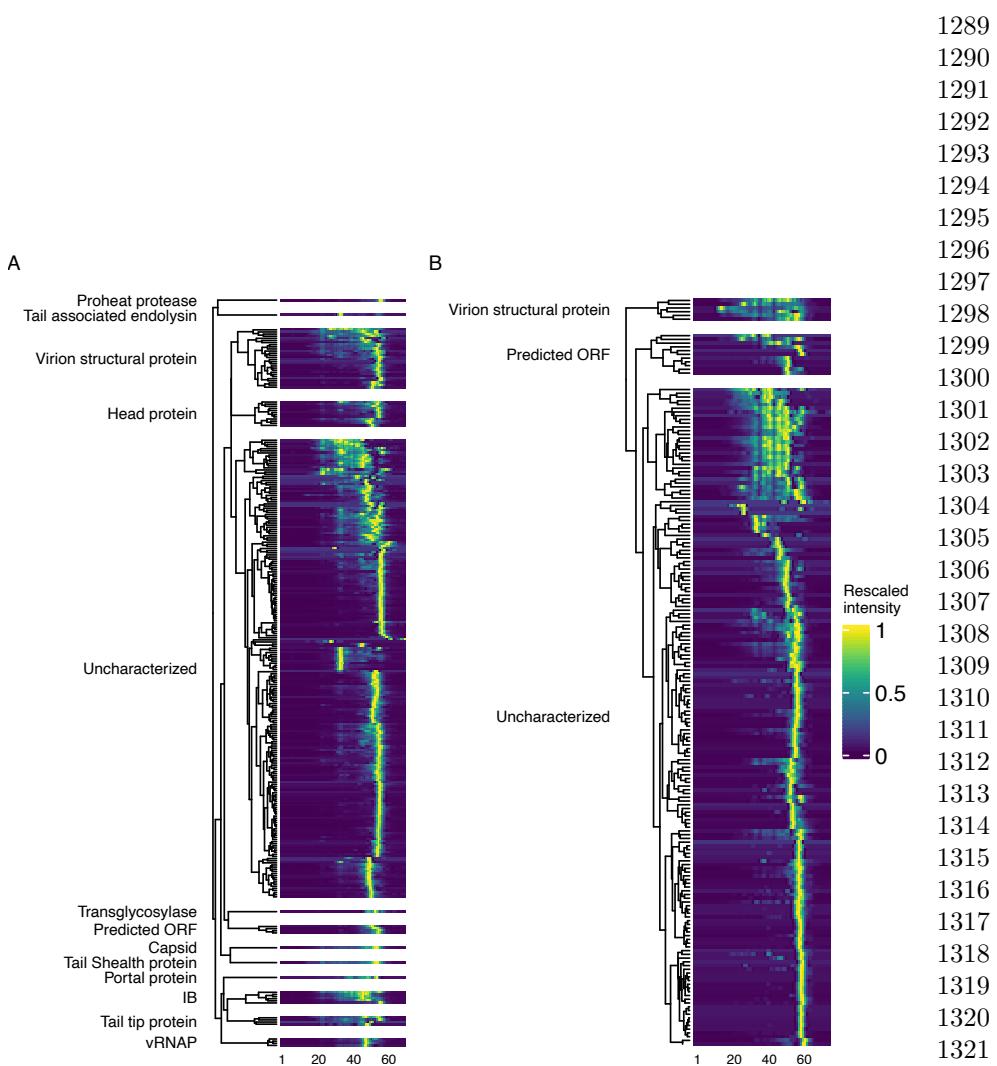


Supplementary Fig. S 1. Violin plot showing the distance between the observed protein SEC peak and their predicted molecular weight expressed as fraction number for the single-peak proteins. Black line represents the mean.

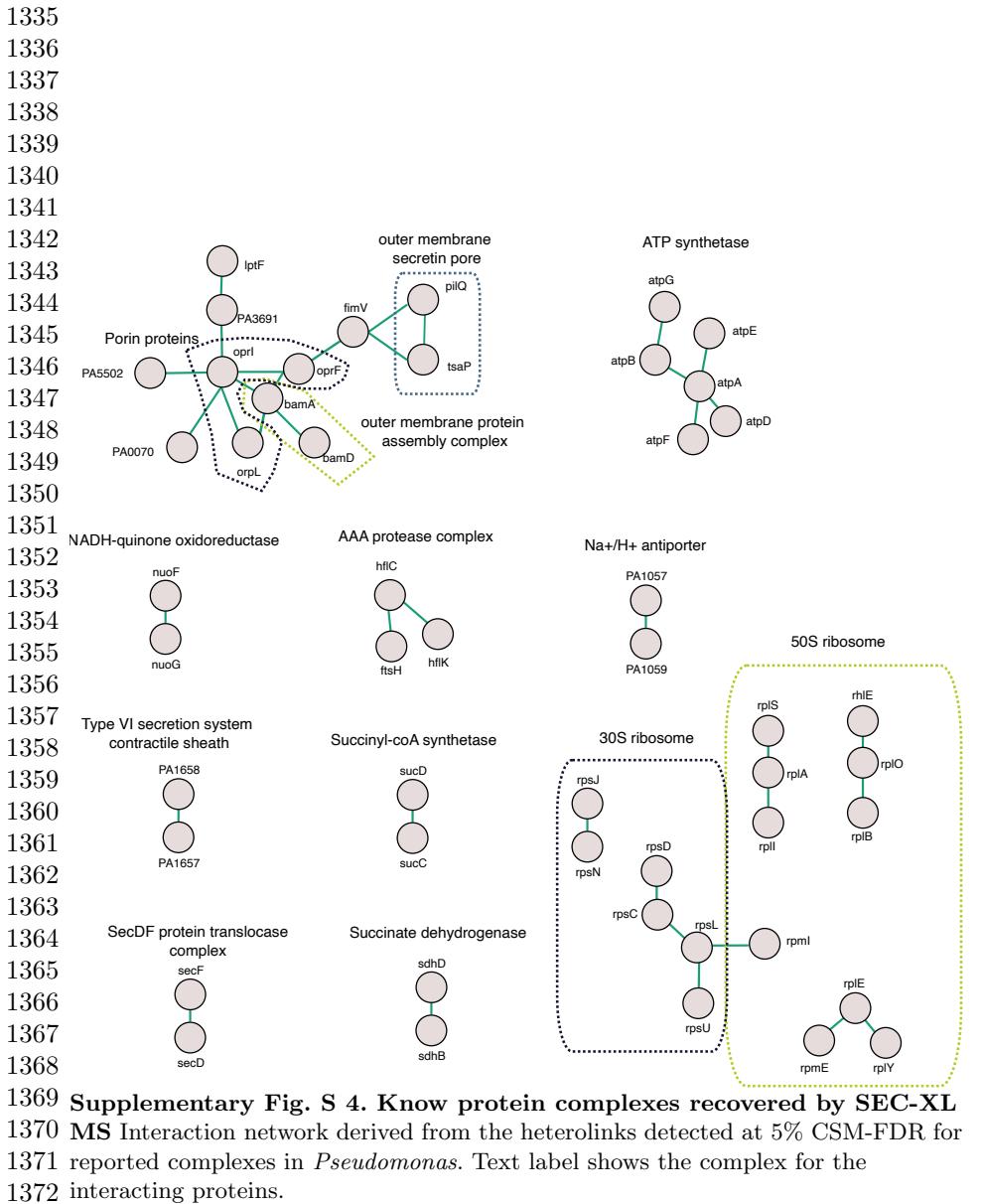
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242



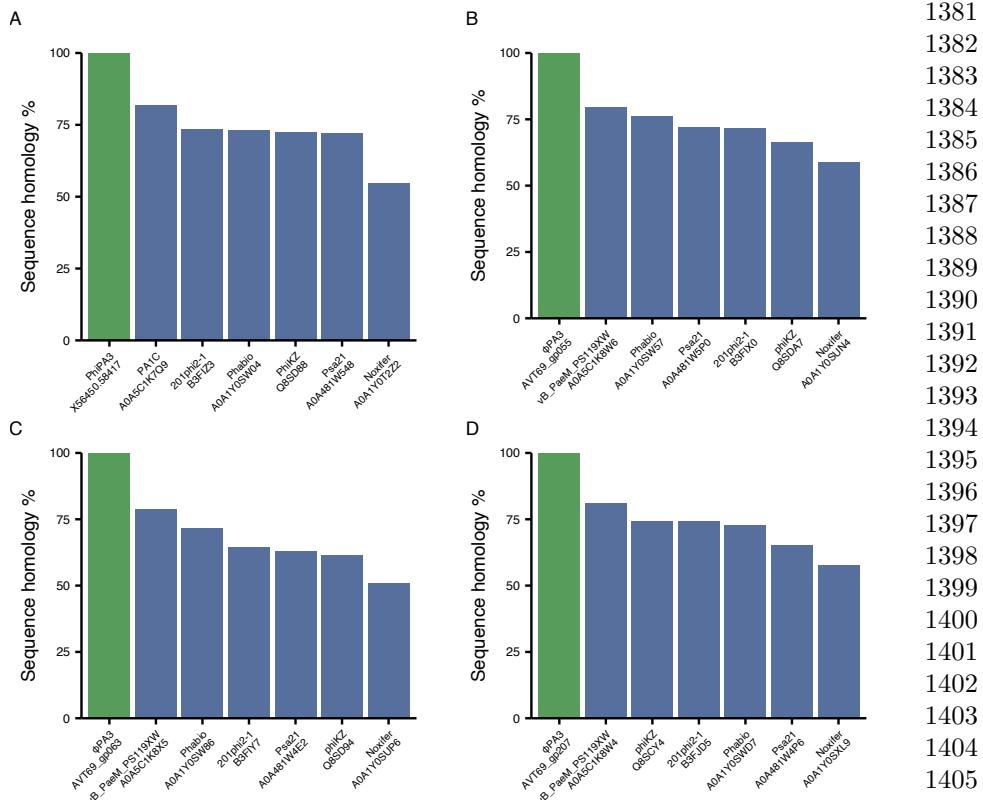
Supplementary Fig. S 2. Ribosome absorption profile and MW from SEC Chromatographic traces collected at 280 nm for the purified 70S ribosome (red), the protein mixture used as molecular weight standards (cyan) and the ϕ KZ infected sample (green). The grey area shows the fractions used for the ribosomal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



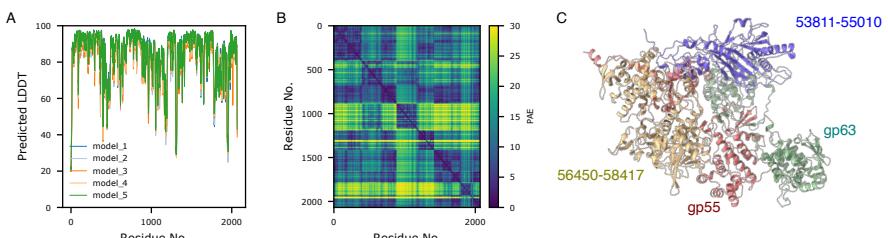
Supplementary Fig. S 3. Distribution of the KZ-like phages proteomes into discrete assemblies A-B. Coelution heatmap for all the phage proteins identified in ϕ KZ (**A**) and ϕ PA3 (**B**). The dendrogram branches are labelled based on manual literature curation for the corresponding proteins in the peak group. Color represents the unit-rescaled intensity. X axis represent the fraction number.



Supplementary Fig. S 4. Known 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 ($iPTM + TM = 0.826$)

1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426

1427

1428

1429

1430

1431

1432

1433

1434

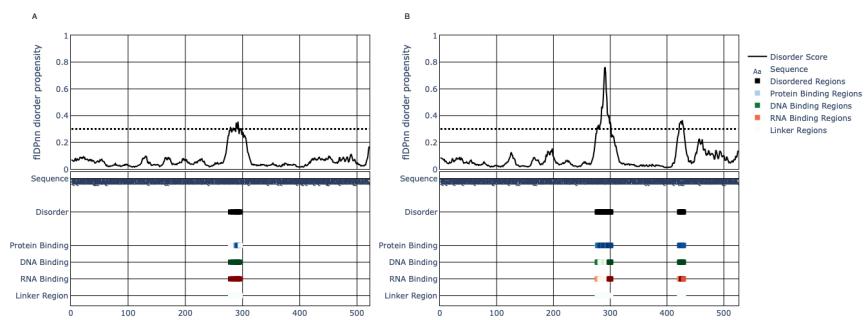
1435

1436

1437

1438

1439



Supplementary Fig. S 7. Intrinsically disordered region in ϕ KZ gp68 and ϕ PA3 gp63 Prediction of disordered regions using fIDPnn[48]. X axis represent sequence, while different rows shows different local predicted properties between ϕ KZ gp68 (**A**) and ϕ PA3 gp63 (**B**)

1440

1441

1442

1443

1444

1445

1446

1447

1448

1449

1450

1451

1452

1453

1454

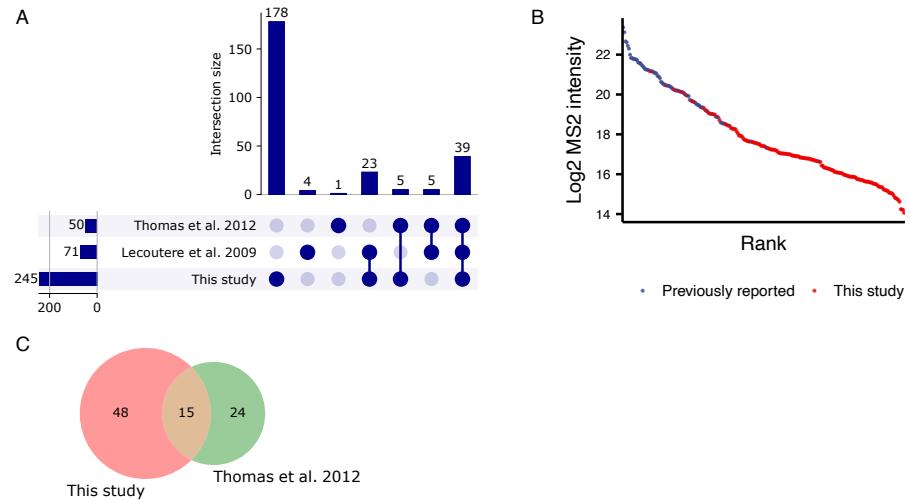
1455

1456

1457

1458

1459



Supplementary Fig. S 8. Benchmark of purified virion MS versus other reported studies of ϕ KZ 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]

1471

1472

Supplementary tables

Table 1. Gene duplication in ϕ KZ-like Jumbophages

ϕ KZ gene	ϕ PA3 gene	Virion protein	
PHIKZ083	AVT69_gp236, AVT69_gp079, AVT69_gp081	YES	1479
PHIKZ093	AVT69_gp097, AVT69_gp099	YES	1480
PHIKZ094	AVT69_gp096, AVT69_gp097, AVT69_gp098	YES	1481
PHIKZ118	AVT69_gp130, AVT69_gp132	NO	1482
PHIKZ131	AVT69_148, AVT69_149	NO	1483
PHIKZ133	AVT69_gp150, AVT69_gp152	NO	1484
PHIKZ134	AVT69_gp150, AVT69_gp151, AVT69_gp152	NO	1485
PHIKZ135	AVT69_gp150, AVT69_gp151, AVT69_gp152	NO	1486
PHIKZ144	AVT69_gp161, AVT69_gp211	YES	1487
PHIKZ145	AVT69_gp162, AVT69_gp163	YES	1488
PHIKZ146	AVT69_gp164, AVT69_gp165, AVT69_gp167	YES	1489
PHIKZ178	AVT69_gp207, AVT69_gp209	YES	1490
PHIKZ179	AVT69_gp131, AVT69_gp208	NO	1491
PHIKZ184	AVT69_gp215, AVT69_gp095	YES	1492
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 Δ fliC) 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 MgSO₄ 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 (ϕ KZ or

1519 ϕ PA3; MOI \approx 1) on ice for 10 minutes (to allow complete adsorption of virions
1520 onto cells) and then incubated at 30°C for 50 minutes (total time of infection
1521 60 minutes). Thereafter, the cell cultures were transferred to pre-chilled 50
1523 mL falcon tubes, centrifuged at 6000xg, 0°C for 5 minutes. The supernatant
1524 was discarded and cell pellets were washed twice with 5 mL ice-cold LB and
1525 combined. After the final wash, the bacterial pellets were resuspended in 5 mL
1526 ice-cold LB. The concentrated cell culture was flash frozen in liquid nitrogen
1527 and subsequently mechanically lysed using a SPEX-freezer mill.
1528
1529
1530
1531

1532 **Cesium gradient purification of phage virions**

1533
1534

1535 Bacteriophages (ϕ KZ or ϕ PA3) were propagated in LB at 37°C with PAO1 as
1536 a host. Liquid growth curve experiments were used to ascertain the MOI of
1537 bacteriophage stock needed to ensure complete lysis of the bacteria following
1538 a substantial growth as ascertained by OD600 measurement. Growth curve
1539 experiments were carried out in a Synergy H1 micro-plate reader (BioTek,
1540 with Gen5 software). Cells were diluted 1:100 from a saturated overnight cul-
1541 ture with 10 mM MgSO₄. Diluted culture (140 μ L) was added together with
1542 10 μ L of 10X serial dilutions of bacteriophage stocks to wells in a 96-well plate.
1543 This plate was cultured with maximum double orbital rotation at 37°C for
1544 24 h with OD600 nm measurements every 5 minutes. Thereafter, the bacte-
1545 riophage stock was added at the appropriate MOI to a 1:100 back-dilution of
1546 a saturated PAO1 overnight culture in 100 mL LB with 10mM MgSO₄ and
1547 the bacterial culture incubated for 24 hours (37°C with aeration, 175 rpm). 5
1548 mL of chloroform was added to the cultures in a fume-hood and the cultures
1549 were incubated to with chloroform for 15 minutes (37°C, 175 rpm) to ensure
1550 maximum lysis of bacterial cells. The cell cultures were transferred to 50 mL
1551 falcon tubes and centrifuged at 6000 xg for 15 min to pellet bacterial debris.
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564

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 MgSO₄, 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 MgCl₂ 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

1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610

1611 deplete monomeric proteins. The concentrated sample was centrifuged once
1612 more at 10'000 g at 4°C to remove particles.
1613

1614

1615

1616 Size-exclusion chromatography

1617

1618 Approx 1000 µg per sample ($\approx 80 - 90\mu\text{L}$ as estimated by Bradford's assay)
1619 were separated on a Agilent Infinity 1260 HPLC operating at 0.5 mL/minute
1620 in SEC running buffer with a Phenomenex SRT-C1000 column connected
1621 and cooled at 4 °C. 72 fractions of 125 ul were collected after 3.75 ml until 13
1622 ml and the column was then washed with 2 column volumes (18 mL) of SEC
1623 buffer. The MW was estimated using a protein mixture (Phenomenex AL0-
1624 3042), while a *E.Coli* 70s ribosome (NEB, cat nr P0763S) was used to estimate
1625 which fractions to use for ribosome XL-MS.

1630

1631

1632

1633

1634 SEC-MS proteomics sample preparation

1635

1636 The SEC samples were prepared as we previously reported[51] using a 96 well
1637 filter-aided sample preparation (FASP). The FASP-filters were conditioned by
1638 washing twice with 100 µL of ddH₂O. SEC buffer was removed by centrifuga-
1639 tion (1800 g 1 h) and proteins were resuspended in 50 µL of TUA buffer (TCEP
1640 5 mM, Urea 8M, 20 mM ammonium bicarbonate) and incubated on a thermos
1641 shaker (37°C, 400 rpm) for 30 minutes. Cysteine residues were then alkylated
1642 by addition of 20 µL CAA buffer (Chloroacetamide 35 mM, 20 mM ammo-
1643 nium bicarbonate) for 1 h at 25°C in the dark. TCEP and IAA were removed
1644 by centrifugation (1800 g, 30 min) and filters were washed 3 times with 100 µL
1645 of 20 mM ammonium bicarbonate. Proteins were digested in 50 µL of 20 mM
1651 ammonium bicarbonate with 1 µg of trypsin per fraction. A 96 well receiver
1653 plate (Nucleon, Thermo-Fisher) was used to collect the peptides by centrifu-
1654 gation for 30 minutes at 1800g. The filter plates were washed once with 100 µL
1655

of ddH₂O 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 ddH₂O 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. Elution Peptides were loaded twice and then washed with 3 CV of 0.1% FA. Elution

1703 was done using 0.5 CV of 50% ACN 0.1% FA and repeated twice. Samples
1704 were dried under vacuum and stored at -80°C until acquisition.

1706

1707

1708

1709 Crosslinking MS sample preparation

1710

1711 ϕ KZ infection and SEC-separation were performed as described above. Fol-
1712 lowing separation, the SEC-fractions corresponding to the 70S ribosome peak
1713 (F33-F38) were pooled. The was crosslinked for 1 hr at RT using 5 mM DSSO
1715 from a freshly prepared 30 mM stock in water-free DMF. The reaction was
1716 quenched by addition of ABC to 50 mM for 30 minutes at RT and the pro-
1718 teins were precipitated using 8 volumes of ice-cold acetone. Following overnight
1720 incubation, pellets were washed 5 times with 8x volumes of ice-cold acetone
1722 and briefly dried under vacuum. The pools were reconstituted in 8M urea, 100
1724 mM ABC and 5 mM TCEP and incubated for 30 minutes at RT. CAA was
1726 added to 10 mM final concentration and the samples were incubated in the
1727 dark for 1 hr. Urea was diluted to 1 M by addition of 100 mM ABC and the
1729 proteins were digested overnight with 2 ug of trypsin in a thermo shaker at 30
1730 °C. Samples were acidified with 10% TFA and high-ph tip fractionation was
1732 performed as we previously described[51]. Briefly, following activation, equili-
1734 bration and washing of the C18 resin, the elution was done using a step-wise
1735 gradient of ACN from 10 to 40 (5% increases) in 0.1% triethylamine to account
1737 for the increased hydrophobicity of the XL peptides compared to not modified
1739 ones. Resulting fractions were dried under vacuum.

1741

1742

1743

1744 SEC-MS and spectral library acquisition

1745

1746 Samples were resuspended in buffer A (0.1% FA) and approximately 200 ng
1747 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 μ l/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 μ l/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 μ m 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

1795 inclusion was enabled ($3 + \text{to}8+$ precursors). Precursors having nominal inten-
1796
1797 sity above 20'000 were selected for fragmentation using an inverted collision
1798 energy of 23 eV at 0.73 1/ k_0 and 95 eV at 1.6 1/ k_0 .
1799

1800

1801

1802

1803

1804

1805

1806 SEC-MS data analysis

1807

1808 The DDA files were searched within the Fragpipe framework using
1809 MSfragger[54] and the 'DIA-speclib-quant' workflow using the *Pseudomonas*
1811 *aeruginosa* pan proteome FASTA (5564 entries, proteome ID UP000002438,
1812 downloaded on the 05/22). For each phage, the correspondent FASTA
1813 nucleotide file was downloaded from GenBank (NC_004629.1 for ϕ KZ and
1814 NC_028999.1 for ϕ PA3) and EMBOSS was used for novel ORFs prediction
1815 (see 'Prediction of novel ORFs' section for details). The GenBank files were
1816 translated to protein level using BioPython and supplemented to the *Pseu-*
1817 *domonas* FASTA. Carbamylation of cysteines was set as fixed modification
1818 while oxidation of methionine, N-term acetylation (peptide level) and pyro-
1819 glu formation were set as variable modifications. EasyPQP (<https://github.com/grosenberger/easypqp>) was used to generate a spectral library. Follow-
1820 ing phage-specific library generation, PAO1 precursors from all libraries were
1821 transferred to ensure the presence of the same PAO1 proteins with the same
1822 peptides across all DIA experiments using lowess for RT realignment. The
1823 DIA-PASEF data was searched with DIA-NN[55] v.1.7.1 using a library-centric
1824 approach. Identified spectrum with MS1 precursors within 10 ppm and MS2
1825 precursor within 15 ppm were selected and a second library was generated
1826 (double-pass mode). Quantification was set to robust (high-accuracy) and
1827 cross-run normalization was disabled.
1828

XL-MS data analysis	1841
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.	1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886
Data analysis for DDA purified virion samples	1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886
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, <i>Pseudomonas aeruginosa</i> plus contaminants. Decoys were generated by pseudo-inversion. Percolator was used for FDR-control at 1% PSM.	1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886
Protein-protein interaction prediction from SEC-MS data	1877 1878 1879 1880 1881 1882 1883 1884 1885 1886
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	1877 1878 1879 1880 1881 1882 1883 1884 1885 1886

1887 filter and rescaled in a 0-1 range. A dot product matrix between all pro-
1888 teins was calculated and protein showing $r^2 \geq 0.3$ were selected as putative
1889 interactors for prediction. For every pair we calculated 5 features: (i) slid-
1890 ing window ($q=6$) correlation, (ii) fraction-wide intensity difference, (iii) peak
1891 shift, (iv) Euclidean distance and (v) contrast angle dot-product.
1892

1893 For prediction, we utilized a fully-connected neural network implemented
1894 in Tensorflow (<https://www.tensorflow.org>). Briefly, we set the input layer as
1895 number of features (147) followed by a fully connected layer with 100 neurons
1896 and a dropout layer (0.2 %) and a fully connected layer with 72 neurons. A final
1897 output layer using sigmoid as activation function was used for classifying co-
1898 eluting and not-coeluting proteins. For training, a previously reported dataset
1899 was used[31]. To select for positive we utilized protein pairs in STRING using a
1900 combined score of 0.9 and experimental evidence, while negative were randomly
1901 selected. The DNN model was trained for 100 epochs using ADAM (learn-
1902 ing rate = 0.001) and binary cross-entropy as loss function. Early stopping
1903 (patience = 20) to avoid overfitting. To further removed spuriously co-eluting
1904 PPIs after the prediction step, we calculated an equal number of decoy PPIs
1905 by randomly sampling the remaining proteins and utilized the DNN model to
1906 predict their coelution probability. We then utilized these two distributions to
1907 perform target-decoy competition (TDC) using posterior probabilities.
1908
1909

1910

1911

1912

1913

1914 **ORFs prediction from nucleotide FASTA**

1915

1916

1917 EMBOSS v6.6.0.0 subroutine getorf was used to predict open reading frames
1918 (ORFs) with a minimum size of 50 AA. Existing annotated genes were removed
1919 from the predicted ORFs using bedtools subroutine subtract, allowing us to
1920 differentiate between existing and novel ORFs.

Structural prediction and alignment for ϕPA3 vRNAP	1933
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.	1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978
Alignments of predicted complex structures (ϕ KZ vRNAP and 4 proteins ϕ PA3 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 ϕ PA3, was used as template in MatchMaker.	
Generation of ϕKZ particles packaged with 3xFLAG fusions of ϕKZ virion proteins.	
ϕ KZ particles packaged with virion proteins bearing a C-terminal 3xFLAG-tag were generated by adapting a protocol used to generate ϕ KZ 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 μ g/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 μ g/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	

1979 were infected with ϕ KZ (WT, MOI \approx 1) for 2.5 hours. Thereafter 1 mL of
1980 chloroform was added to the cultures in a fume-hood and the cultures were
1981 incubated with chloroform for 15 minutes (37°C, 175 rpm). The cell cul-
1982 tures were transferred to 15 mL falcon tubes and centrifuged at 6000xg for
1983 15 min to pellet bacterial debris. The supernatant (containing bacteriophages
1984 in high titer) was carefully transferred to a fresh set of 50mL falcon tubes
1985 and centrifuged at 6000xg for 15 min to pellet any residual bacterial debris.
1986 Thereafter, 4 mL the supernatant was filtered and concentrated (\approx 10x) using
1987 Amicon-100 centrifugal filters to remove excess 3xFLAG-tagged proteins. The
1988 concentrated supernatant was used for western blot experiments.
1989
1990
1991
1992
1993
1994
1995
1996
1997

1998 Western blot and blot analysis

1999

2000 PA01 cells were grown as previously described and upon reaching 0.5 OD (600
2001 nm), gentamicin was added (50 μ g/ml) and the celles were chilled on ice for 5
2002 minutes to stall translation. Thereafter PAO1 cells (\approx 1 OD equivalent) were
2003 infected with ϕ KZ particles packaged with virion proteins bearing a C-terminal
2004 3xFLAG-tag (MOI \approx 1) on ice for 10 minutes (to allow complete adsorption
2005 of virions onto cells) and then incubated at 30°C for 15 minutes. Thereafter,
2006 the cell cultures were transferred to pre-chilled 15mL falcon tubes, centrifuged
2007 at 6000xg, 0°C for 5 minutes. The supernatant was discarded and the cell
2008 pellet was washed twice with 2 mL of pre-chilled (0°C) LB to remove excess
2009 unbound virions. The cell pellet was lysed in 100 μ L of lysis buffer (20 mM
2010 Tris, pH 7.5, 150 mM NaCl, 2% glycerol, 1% TTX-100, CompleteMini EDTA-
2011 free protease inhibitor cocktail). The lysed suspension was further sonicated
2012 on ice using a Q125 sonicator (10 pulses, 1s ON, 1s OFF, 30% amplitude). The
2013 cell lysate was centrifuged at 15000xg (15min, 0°C) to remove cellular debris.
2014 The clarified cellular lysate (100 μ L) was boiled with 33 μ LL of 4X Laemmli

Buffer (with Beta-mercaptoethanol) for 10 min. 14 μ LL of lysate samples was loaded. For virion control samples, 10 μ LL of purified virions were boiled with 3.3 μ LL of 4X Laemmli Buffer (with Beta-mercaptoethanol) for 10 min and 2 μ LL 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 μ M 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% Omnipro 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/s41467-022-10937-0>

- 2071 [org/10.1038/s41467-022-29346-w](https://doi.org/10.1038/s41467-022-29346-w) .
- 2072
- 2073 [3] Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan,
2074 N. J. A systems approach to infectious disease. *Nature reviews. Genetics*
2075 **21** (6), 339–354 (2020). <https://doi.org/10.1038/s41576-020-0212-5> .
- 2076
- 2077 [4] Batra, J. *et al.* Protein Interaction Mapping Identifies RBBP6 as a
2078 Negative Regulator of Ebola Virus Replication. *Cell* **175** (7), 1917–
2079 1930.e13 (2018). URL <https://doi.org/10.1016/j.cell.2018.08.044>. <https://doi.org/10.1016/j.cell.2018.08.044> .
- 2080
- 2081
- 2082
- 2083
- 2084 [5] Hashimoto, Y., Sheng, X., Murray-Nerger, L. A. & Cristea, I. M. Tem-
2085 poral dynamics of protein complex formation and dissociation during
2086 human cytomegalovirus infection. *Nature Communications* **11** (1), 806
2087 (2020). URL <https://doi.org/10.1038/s41467-020-14586-5>. <https://doi.org/10.1038/s41467-020-14586-5> .
- 2088
- 2089
- 2090
- 2091
- 2092
- 2093
- 2094
- 2095 [6] Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals
2096 targets for drug repurposing. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2286-9> .
- 2097
- 2098
- 2099
- 2100
- 2101 [7] Stukalov, A. *et al.* Multilevel proteomics reveals host perturbations
2102 by SARS-CoV-2 and SARS-CoV. *Nature* **594** (7862), 246–252 (2021).
2103 URL <https://doi.org/10.1038/s41586-021-03493-4>. <https://doi.org/10.1038/s41586-021-03493-4> .
- 2104
- 2105
- 2106
- 2107
- 2108 [8] Meyers, J. M. *et al.* The proximal proteome of 17 SARS-CoV-2 pro-
2109 teins links to disrupted antiviral signaling and host translation. *PLoS
2110 Pathogens* **17** (10), 1–30 (2021). URL <http://dx.doi.org/10.1371/journal.ppat.1009412>. <https://doi.org/10.1371/journal.ppat.1009412> .
- 2111
- 2112
- 2113
- 2114
- 2115
- 2116

- [9] Dadgostar, P. Antimicrobial resistance: implications and costs. *Infection and Drug Resistance* **12**, 3903–3910 (2019). <https://doi.org/10.2147/IDR.S234610> .
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
- [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> .
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
- [11] Krylov, V. *et al.* Phage phikz—the first of giants. *Viruses* **13** (2), 1–18 (2021). <https://doi.org/10.3390/v13020149> .
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
- [12] Monson, R., Foulds, I., Foweraker, J., Welch, M. & Salmond, G. P. The Pseudomonas aeruginosa generalized transducing phage ϕ PA3 is a new member of the ϕ KZ-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> .
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
- [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> .
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
- [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> .
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162

- 2163 [15] Chaikeratisak, V., Birkholz, E. A. & Pogliano, J. The Phage Nucleus
2164 and PhuZ Spindle: Defining Features of the Subcellular Organization and
2165 Speciation of Nucleus-Forming Jumbo Phages. *Frontiers in Microbiology*
2166 **12** (July), 1–8 (2021). <https://doi.org/10.3389/fmicb.2021.641317> .
2167
- 2168 [16] Chaikeratisak, V. *et al.* Subcellular organization of viral particles during
2169 maturation of nucleus-forming jumbo phage. *Science Advances* **8** (18),
2170 8–9 (2022). <https://doi.org/10.1126/sciadv.abj9670> .
2171
- 2172 [17] Li, Y. *et al.* A family of novel immune systems targets early infec-
2173 tion of nucleus-forming jumbo phages. *bioRxiv* 2022.09.17.508391
2174 (2022). URL <http://biorxiv.org/content/early/2022/09/18/2022.09.17.508391.abstract>.
2175 <https://doi.org/10.1101/2022.09.17.508391> .
2176
- 2177 [18] Fossati, A. *et al.* System-Wide Profiling of Protein Complexes Via Size
2178 Exclusion Chromatography-Mass Spectrometry (SEC-MS). *Methods in
2179 molecular biology (Clifton, N.J.)* **2259**, 269–294 (2021). https://doi.org/10.1007/978-1-0716-1178-4_18 .
2180
- 2181 [19] Skinnider, M. A. & Foster, L. J. Meta-analysis defines principles for the
2182 design and analysis of co-fractionation mass spectrometry experiments.
2183 *Nature Methods* **18** (7), 806–815 (2021). URL <https://doi.org/10.1038/s41592-021-01194-4>.
2184 <https://doi.org/10.1038/s41592-021-01194-4> .
2185
- 2186 [20] Caufield, J. H., Abreu, M., Wimble, C. & Uetz, P. Protein Complexes
2187 in Bacteria. *PLOS Computational Biology* **11** (2), 1–23 (2015). URL
2188 <https://doi.org/10.1371/journal.pcbi.1004107>. <https://doi.org/10.1371/journal.pcbi.1004107> .
2189
- 2190
- 2191
- 2192
- 2193
- 2194
- 2195
- 2196
- 2197
- 2198
- 2199
- 2200
- 2201
- 2202
- 2203
- 2204
- 2205
- 2206
- 2207
- 2208

- [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](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](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](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>.
- .

- 2255 [27] Das, T. *et al.* Phenazine virulence factor binding to extracellular DNA
2256 is important for *Pseudomonas aeruginosa* biofilm formation. *Scientific*
2257 *reports* **5**, 8398 (2015). <https://doi.org/10.1038/srep08398> .
2259
- 2260
- 2261 [28] Andrésen, C. *et al.* Critical biophysical properties in the *Pseudomonas*
2262 *aeruginosa* efflux gene regulator MexR are targeted by mutations con-
2263 ferring multidrug resistance. *Protein Science* **19** (4), 680–692 (2010).
2264
- 2265 <https://doi.org/10.1002/pro.343> .
2266
- 2267
- 2268 [29] Nazir, A., Ali, A., Qing, H. & Tong, Y. Emerging Aspects of Jumbo
2269 Bacteriophages. *Infection and drug resistance* **14**, 5041–5055 (2021).
2270
- 2271 <https://doi.org/10.2147/IDR.S330560> .
2272
- 2273
- 2274 [30] Chaikeeratisak, V. *et al.* The Phage Nucleus and Tubulin Spindle Are
2275 Conserved among Large *Pseudomonas* Phages. *Cell Reports* **20** (7), 1563–
2276 1571 (2017). URL <http://dx.doi.org/10.1016/j.celrep.2017.07.064>. <https://doi.org/10.1016/j.celrep.2017.07.064> .
2277
- 2278
- 2279
- 2280
- 2281 [31] Salas, D., Stacey, R. G., Akinlaja, M. & Foster, L. J. Next-generation
2282 interactomics: Considerations for the use of co-elution to measure protein
2283 interaction networks. *Molecular and Cellular Proteomics* **19** (1), 1–10
2284 (2020). <https://doi.org/10.1074/mcp.R119.001803> .
2285
- 2286
- 2287
- 2288
- 2289 [32] Havugimana, P. C. *et al.* A census of human soluble pro-
2290 tein complexes. *Cell* **150** (5), 1068–1081 (2012). URL
2291 <http://www.ncbi.nlm.nih.gov/pubmed/22939629>{%}0A<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&term=PMC3477804>.
2292
- 2293
- 2294 <https://doi.org/10.1016/j.cell.2012.08.011> .
2295
- 2296
- 2297
- 2298
- 2299
- 2300

- [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> .
2301
2302
2303
2304
2305
2306
[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> .
2307
2308
2309
2310
2311
2312
[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> .
2313
2314
2315
2316
2317
2318
2319
[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> .
2320
2321
2322
2323
2324
2325
[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> .
2326
2327
2328
2329
2330
2331
2332
2333
[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> .
2334
2335
2336
2337
2338
[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> .
2339
2340
2341
2342
2343
2344
2345
2346

- 2347 [40] Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra gener-
2348 ated by data-independent acquisition: A new concept for consistent and
2349 accurate proteome analysis. *Molecular and Cellular Proteomics* **11** (6),
2350 O111.016717 (2012). URL <http://www.mcponline.org/lookup/doi/10.1074/mcp.O111.016717>.
2351
2352
2353
2354
2355
- 2356 [41] Orekhova, M., Koreshova, A., Artamonova, T., Khodorkovskii, M. &
2357 Yakunina, M. The study of the phiKZ phage non-canonical non-virion
2358 RNA polymerase. *Biochemical and Biophysical Research Communications*
2359 **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>.
- 2360
2361
2362
2363
2364
2365
2366
- 2367 [42] de Martín Garrido, N. *et al.* Structure of the bacteriophage PhiKZ non-
2368 virion RNA polymerase. *Nucleic Acids Research* **49** (13), 7732–7739
2369 (2021). URL <https://doi.org/10.1093/nar/gkab539>. <https://doi.org/10.1093/nar/gkab539>.
- 2370
2371
2372
2373
2374
- 2375 [43] Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer.
2376 *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>.
- 2377
2378
2379
2380
- 2381 [44] Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: univer-
2382 sal structure alignments of proteins, nucleic acids, and macromolecular
2383 complexes. *Nature Methods* (2022). URL <https://doi.org/10.1038/s41592-022-01585-1>.
- 2384
2385
2386
2387
- 2388 [45] Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E.
2389 Tools for integrated sequence-structure analysis with UCSF Chimera.
2390 *BMC Bioinformatics* **7** (1), 339 (2006). URL <https://doi.org/10.1186/1471-2105-7-339>.
- 2391
2392

- 1471-2105-7-339. <https://doi.org/10.1186/1471-2105-7-339> . 2393
2394
- [46] Thomas, J. A. *et al.* Extensive proteolysis of head and inner body proteins 2395
by a morphogenetic protease in the giant *Pseudomonas aeruginosa* phage 2396
 ϕ KZ. *Molecular Microbiology* **84** (2), 324–339 (2012). <https://doi.org/10.1111/j.1365-2958.2012.08025.x> . 2397
2398
- [47] Bailey, T. L. STREME: Accurate and versatile sequence motif discovery. 2399
Bioinformatics (Oxford, England) **37** (18), 2834–2840 (2021). <https://doi.org/10.1093/bioinformatics/btab203> . 2400
2401
- [48] Hu, G. *et al.* fIDPnn: Accurate intrinsic disorder prediction with putative 2402
propensities of disorder functions. *Nature Communications* **12** (1), 4438 2403
(2021). URL <https://doi.org/10.1038/s41467-021-24773-7>. <https://doi.org/10.1038/s41467-021-24773-7> . 2404
2405
- [49] Lecoutere, E. *et al.* Identification and comparative analysis of the struc- 2406
tural proteomes of phiKZ and EL, two giant *Pseudomonas aeruginosa* 2407
bacteriophages. *Proteomics* **9** (11), 3215–3219 (2009). <https://doi.org/10.1002/pmic.200800727> . 2408
2409
- [50] Wu, W., Thomas, J. A., Cheng, N., Black, L. W. & Steven, A. C. Bubble- 2410
grams reveal the inner body of bacteriophage ϕ KZ. *Science (New York, 2411
N.Y.)* **335** (6065), 182 (2012). <https://doi.org/10.1126/science.1214120> . 2412
2413
- [51] Fossati, A. *et al.* Toward Comprehensive Plasma Proteomics by Orthog- 2414
onal Protease Digestion. *Journal of proteome research* (20), 4031–4040 2415
(2021). <https://doi.org/10.1021/acs.jproteome.1c00357> . 2416
2417
- [52] Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation 2418
combined with data-independent acquisition. *Nature Methods* **17** (12), 2419
2420
- 2421
- 2422
- 2423
- 2424
- 2425
- 2426
- 2427
- 2428
- 2429
- 2430
- 2431
- 2432
- 2433
- 2434
- 2435
- 2436
- 2437
- 2438

- 2439 1229–1236 (2020). URL <https://doi.org/10.1038/s41592-020-00998-0>.
2440
2441
2442
2443 [53] Steigenberger, B. *et al.* Benefits of Collisional Cross Section Assisted
2444 Precursor Selection (caps-PASEF) for Cross-linking Mass Spectrometry.
2445 *Molecular & cellular proteomics : MCP* **19** (10), 1677–1687 (2020). <https://doi.org/10.1074/mcp.RA120.002094>.
2446
2447
2448
2449
2450 [54] Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. &
2451 Nesvizhskii, A. I. MSFragger: Ultrafast and comprehensive peptide iden-
2452 tification in mass spectrometry-based proteomics. *Nature Methods* **14** (5),
2453 513–520 (2017). <https://doi.org/10.1038/nmeth.4256>.
2454
2455
2456
2457
2458 [55] Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser,
2459 M. DIA-NN: neural networks and interference correction enable deep
2460 proteome coverage in high throughput. *Nature Methods* **17** (1), 41–44
2461 (2020). URL <http://dx.doi.org/10.1038/s41592-019-0638-x>. <https://doi.org/10.1038/s41592-019-0638-x>.
2462
2463
2464
2465
2466
2467 [56] Mendes, M. L. *et al.* An integrated workflow for crosslinking mass
2468 spectrometry. *Molecular Systems Biology* **15** (9), e8994 (2019). URL
2469 <https://www.embopress.org/doi/abs/10.15252/msb.20198994>. <https://doi.org/https://doi.org/10.15252/msb.20198994>.
2470
2471
2472
2473
2474
2475 [57] Guan, J. *et al.* Bacteriophage genome engineering with CRISPR-Cas13a.
2476 *Nature microbiology* **7** (12), 1956–1966 (2022). <https://doi.org/10.1038/s41564-022-01243-4>.
2477
2478
2479
2480
2481
2482
2483
2484