Nature Communications

https://doi.org/10.1038/s41467-026-74701-w

Article in Press

Structural choreography of bacteriophage N4 ejection proteins and the giant virion-associated RNA polymerase

Received: 29 September 2025

Accepted: 12 June 2026

Cite this article as: Bellis, N.F., Lokareddy, R.K., Pavlenok, M. *et al.* Structural choreography of bacteriophage N4 ejection proteins and the giant virion-associated RNA polymerase. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-74701-w

Nathan F. Bellis, Ravi K. Lokareddy, Mikhail Pavlenok, Ruth Q. Jacobs, Stephanie L. Cooper Horton, James L. Kizziah, Francesca Forti, David A. Schneider, Michael Niederweis, Federica Briani & Gino Cingolani

Structural choreography of bacteriophage N4 ejection proteins and the giant virion-associated RNA polymerase

Nathan F. Bellis [1], Ravi K. Lokareddy [1], Mikhail Pavlenok [2], Ruth Jacobs [1], Stephanie L. (Cooper) Horton [1], James L. Kizziah [3], Francesca Forti [4], David A. Schneider [1], Michael Niederweis [2], Federica Briani [4], and Gino Cingolani [1 *]

[1] *Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, 1825 University Blvd, Birmingham, AL 35294, USA*

[2] *Department of Microbiology, University of Alabama at Birmingham, 845 19th Street South, Birmingham, AL 35294, USA*

[3] *Institutional Research Core Program (IRCP), University of Alabama at Birmingham, 845 19th Street South, Birmingham, AL 35294, USA*

[4] *Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy*

*Corresponding Author: E-mail: gcingola@uab.edu

ABSTRACT

Schitoviruses are widespread prokaryotic viruses that encapsidate a giant (~3,500-residue) virion-associated RNA polymerase (vRNAP). During infection, vRNAP is expelled into Gram-negative bacteria, along with two additional ejection proteins, to assemble a transient DNA-ejectosome that becomes transcriptionally active, initiating viral replication. Here, we present an integrative structural analysis of the coliphage N4 vRNAP (gp50). We find that this 383 kDa enzyme is a multi-domain, single-chain RNA polymerase, structurally distinct from both compact single-chain RNAPs and large multi-subunit holoenzymes. vRNAP is composed of loosely connected domains and exhibits an intramolecular mode of allosteric regulation through its C-terminal domain. Comparative analysis of intact and genome-released virions identified gp51, which forms an outer-membrane complex, and gp52, which assembles a periplasmic tunnel. These proteins generate heterogeneous pores that facilitate the release of vRNAP. We further uncover a signaling hub in the phage tail, composed of the receptor-binding protein, tail tube, and tail plug, that detects receptor engagement and orchestrates the release of ejection proteins. We propose that the beads-on-a-string architecture of vRNAP enables the translocation of megadalton-scale protein complexes through the ~35 Å channel formed by the tail and ejection proteins. These findings establish N4 as a distinctive model for protein translocation through biological channels.

INTRODUCTION

Tailed bacteriophages, unlike most eukaryotic viruses, deliver their genetic material into the host bacterium while leaving the empty viral capsid outside. Various strategies are employed to translocate genomic material across the complex and diverse bacterial cell walls [1]. Of particular interest are phages whose tail apparatus is too short to penetrate

the 300 Å-thick cell envelope of Gram-negative bacteria. Formerly known as *Podoviridae*, these phages have a short, non-contractile tail that interrupts the capsid's icosahedral symmetry [2] and provides a conduit for host attachment and genome ejection. The general ejection strategy of short-tailed phages involves encapsidating multiple copies of three specific proteins, known as ejection proteins [3], which are ejected into the host prior to DNA. In this process, the ejection proteins oligomerize and form a continuous channel that couples the DNA-filled capsid to the bacterial protoplasm, enabling successful infection and replication.

Phage T7, a prototypical *E. coli* phage and the classic model system for podoviruses, has proven invaluable in deciphering the organization and function of ejection proteins. T7 stores three proteins, gp14, gp15, and gp16, in an 8:8:4 arrangement stacked above the portal [4,5]. These proteins are released into *E. coli* to form a transenvelope channel, where gp14 creates the outer membrane channel (OMC), gp15 acts as a periplasmic tunnel (PT), and gp16, the largest subunit, forms a poorly understood inner membrane complex (IMC) that binds the bacterial membrane [6-8], while recruiting the host RNA polymerase [9-11]. In vitro reconstituted complexes of PT gp15 bound to a fragment of gp16 have been reported [7,12,13], as well as an in situ analysis of T7 infecting *E. coli* mini cells proved essential to visualize large quaternary structure rearrangements in the phage tail, which are coupled to genome ejection [14]. Collectively, the DNA ejectosome is a transient assembly that facilitates the delivery of T7 DNA (~39.9 kb) using the host RNA polymerase as a downstream motor [1], but must also somehow become inactive, or disassemble after genome ejection, to prevent permeabilizing the bacterial inner membrane [6].

While T7 has provided invaluable evidence in short-tailed phage ejection, it has become clear that there is a vast diversity of tail morphologies encompassed within this long-standing classification system. Specifically, the *E. coli* phage N4 and the growing number of N4-like bacteriophages differ significantly from small podoviruses like T7 both in genetics and structure. As such, N4-like phages have recently been reclassified into the *Schitoviridae* family [15]. N4-like phages have a genome size of about 75 kb, nearly twice that of T7, and encode around 70-90 ORFs, including a characteristic large virion-associated RNA polymerase (vRNAP) unique to this family of bacterial viruses. Recently, we characterized the *Schitovirus* DEV that infects *Pseudomonas aeruginosa* and mapped an operon consisting of three ejection proteins, roughly analogous to T7 gp14, gp15, and gp16, despite lacking detectable sequence similarity [16]. Cryo-EM analysis showed that DEV gp14-like OMC (gp73) and gp15-like PT (gp72) are entirely α-helical and form an aqueous channel about 25-35 Å wide, large enough to allow DNA passage. Additionally, we found that DEV gp14-like factor exhibits pore-forming activity in vitro, supporting its role as a channel at the OM. Interestingly, we found that the ejection operon of DEV lacks a clear gp16 analog; instead, the third ejection protein corresponds to the previously annotated vRNAP, a giant RNA polymerase [3], hallmark of the *Schitoviridae* family [15]. This protein has been extensively studied in phage N4, particularly its role in early transcription [17]. The massive 3,500-residue vRNAP [18,19], present in 4 ± 1 copies inside the mature virion [20], is ejected into the host upon infection [17]. This enzyme is responsible for an early burst of RNA synthesis observed immediately after N4 infection, even after inhibiting the host RNAP with rifampicin [21]. Interestingly, N4 vRNAP efficiently transcribes only

denatured N4 DNA in vitro but is inactive on native N4 DNA, suggesting that transcription of early promoters requires host DNA gyrases to introduce negative supercoils into the phage genome, leading to extrusion of the unique N4 single-stranded DNA (ssDNA) hairpin promoter [17]. N4 vRNAP initiates infection by transcribing the early genes *gp1* and *gp2*, as well as *gp15* and *gp16*. gp1 and gp2 act as cofactors for the heterodimeric N4 RNAPII (gp15:gp16), which, in turn, is responsible for the transcription of the middle genes [17]. Among these middle transcripts, N4SSB redirects the host RNAP to late promoters, mediating the expression of late genes involved in virion assembly, DNA replication, packaging, and host lysis.

In stark contrast to the N4 transcriptional program, which has been studied extensively [17], many key questions about the role of phage N4 vRNAP as an ejection protein and its part in the DNA ejectosome remain unanswered. vRNAP (gp50) consists of three predicted domains: a likely membrane-spanning N-terminal domain (NTD), a single-subunit RNA polymerase (RNAP) domain, and a large C-terminal domain of unknown function (CTD). Of particular interest is how multiple copies of this 3,500-residue enzyme manage to pass through a narrow tail and become a functional RNA polymerase inside the cell. Surprisingly, N4 vRNAP was not visible in a medium-resolution asymmetric reconstruction of the mature N4 virion, and no clear loss of density was observed comparing the WT N4 with a mutant lacking the vRNAP [20]. Recent high-resolution cryo-EM studies of N4 [22] and other *Schitoviruses*, such as the DEV [16] and the *Shigella* phages Moo19 and B2 [23], containing homologous vRNAP, have also failed to visualize the structure of vRNAP in its native state. The only structural information about N4 vRNAP is limited to previous crystal structures of the transcriptionally active RNAP domain [24], which makes up the middle third of the full-length (FL) protein. The active RNAP domain is a divergent single-subunit RNA polymerase fold similar to T7 [25] and mitochondrial RNAP. The structure of the remaining two-thirds of this ejection protein remains completely unknown. Furthermore, it is unknown how such a large protein can be threaded through the narrow 30–45 Å tail of phage N4, given that just the solved crystal structure of the RNAP is nearly twice the diameter of the tail channel. This led to a model whereby the vRNAP is unfolded during ejection [17].

In this work, we use a complementary top-down and bottom-up approach by determining the cryo-EM structures of both the FL recombinantly expressed vRNAP and the fully asymmetric N4 virion tail. These structures provide key insights into both the enzyme itself and the ejection process of vRNAP and other ejection proteins. We localize N4's smaller ejection proteins, gp51 and gp52, inside the virion and demonstrate their function as pore-forming channels. Lastly, we uncover a tail-gating complex that connects the N4 receptor-binding protein to the tail tube, its tail plug, and the release of ejection proteins. Overall, this study describes the specific mechanisms of ejection initiation for the prototypical *Schitovirus* N4.

RESULTS

Bottom-up analysis of N4 ejection protein gp50 (vRNAP)

Despite the large size of N4 vRNAP (gp50) and the presence of 4 ± 1 copies within the virions, attempts to visualize this giant protein inside the virion using cryo-EM and a

genetic knockdown of *gp50* have been unsuccessful [20]. A recent high-resolution asymmetric reconstruction of the mature N4 virion tentatively placed vRNAP inside the tail apparatus [22], although the internal tail volume is not large enough to accommodate 4 ± 1 copies of a ~383 kDa enzyme (see below). We performed bubblegram imaging on intact N4 virions to probe the rough localization of vRNAP inside the capsid. Bubblegramming leverages the differential rate of damage of protein compared to nucleic acid under an electron beam. Proteinaceous material will decay first, resulting in characteristic bubbling due to the release of trapped gases inside the capsid. This method has been employed to gain a deeper understanding of several phage ejection proteins, including T7 and PhiKZ [26-28]. For N4, we observed expected bubbling at the tail vertex corresponding to the portal as well as bubble clusters typically placed above the portal with distinct separation between the portal bubbling (Fig. 1a). When measuring the distance from the tail:capsid interface to the bubble center across all virions, we observed a trimodal distribution (Fig. 1b), with the first peak likely reflecting the portal and associated proteins, and the other two peaks suggesting proteinaceous material extending up to 600 Å inside the capsid. Thus, bubblegram imaging suggests a moderately stochastic placement of proteinaceous material possibly belonging to the few copies [20] of vRNAP inside the capsid with distinct loci of bubbling.

We used a bottom-up approach to decipher the architecture of N4 vRNAP (M.W. ~383 kDa). We purified the recombinant gp50 to homogeneity, which was then vitrified and subjected to cryo-EM data collection (Supplementary Fig. 1a,b). Single particle analysis (SPA) of cryomicrographs revealed a heterogeneous mix of particles with two predominant and two minor populations (Supplementary Fig. 2). The two major populations were reconstructed to 2.6 and 2.8 Å resolution, respectively (Table 1), allowing us to build atomic models for the isolated RNAP domain (res. 1007–2101) and a tightly bound complex of RNAP and CTD with an unresolved NTD, referred to as ΔN-vRNAP tight complex (res. 1007–3500) (Fig. 2a–c). N4 isolated RNAP and ΔN-vRNAP tight complex were real-space refined to a map to model correlation coefficients of CC = 0.90 and 0.84, respectively (Table 1, Supplementary Fig. 3a,b).

The minor populations present on the grid were more challenging to decipher (Supplementary Fig. 2). One population was likely consistent with the isolated CTD; however, due to its preferred orientation, we were unable to compute a meaningful 3D reconstruction. The other smaller population represents a distinct quaternary structure of RNAP and CTD referred to as ΔN-vRNAP loose complex (Fig. 2d). We were able to resolve a 4.3 Å map of this alternate quaternary arrangement of ΔN-vRNAP with a small 18,000 particle subset of a much larger stack consisting of ~325,000 particles putatively in this loose state (Supplementary Fig. 3c). This reconstruction allowed us to dock, and rigid-body refine RNAP and CTD, which revealed CTD rotates relative to RNAP by 100º, compared to the ΔN-vRNAP tight complex (Fig. 2c). Difficulty aligning all ~325,000 particles belonging to this state suggests that the resolved loose complex map using ~18,000 particles may be one stable state on a continuum of motion of the CTD with respect to the RNAP domain. No 2D classes representing NTD (Fig. 2a), either alone or as part of vRNAP, could be resolved. This is explained by a 100-amino-acid low-complexity sequence (res. 907–1006) that links the NTD and RNAP, suggesting a disconnected NTD complex. We observe no evidence of degradation, and immunoblot analysis confirmed the presence of an intact NTD in the sample used for cryo-EM

(Supplementary Fig. 1a), indicating that the NTD is present on the grid; however, both 2D and 3D reconstruction may be hindered by its lack of ordered structure or by smaller subdomains that are easily masked by the stronger signals from the heterogeneous mix of RNAP and CTD.

Quaternary organization of RNAP and C-terminal domains

The significant conformational heterogeneity captured by SPA prompted us to investigate the relative arrangement of RNAP and CTD tight complex. In the isolated RNAP and the tight complex, the overall organization of the RNAP is similar to crystal structures previously reported, with significant local variations likely caused by crystal contacts and crystallization conditions (RMSD with PDB:2PO4 ~ 4.7 Å and 2.0 Å for isolated RNAP and tight complex RNAP, respectively) [24].

The vRNAP CTD structure described herein is a large structure, quite wide in two dimensions (125 Å x 100 Å) and very narrow in the third (50 Å). The domain is roughly divided into two subdomains: C1 (res. 2124–2375, Fig. 2a–d green) and C2 (res. 2376–3500, Fig. 2a–d blue). The C1 domain possesses a conserved HExxH motif, found in zincin metalloproteases and radical S-adenosylmethionine (radical SAM) enzymes [29]. C2 is large (~110 kDa), mostly α-helical, and shows no significant structural homology to any deposited protein structures or even AlphaFold predictions of biochemically well-studied proteins lacking solved structures.

The tight complex of the RNAP and CTD (Fig. 2b,c) forms a compact structure where the CTD binds the RNAP on the opposing face of the DNA hairpin promoter binding site [24], leaving the region open for promoter binding. Despite a large binding interface of 1,659 Å^2, PISA analysis of the C2 and RNAP interface indicates a weak interaction with a ΔG of nearly zero at –0.2. Interestingly, interactions are slightly stronger at a ΔG of –2.1 for RNAP to C1 with a 725 Å^2 interface. A salt bridge between D2012 (RNAP) and R2279 (C1) (Fig. 3a) likely serves as a clasp for stabilizing the tight complex, which is not predicted to be stable in solution. Interactions between the C1 and C2 are slightly stronger, although we observe contiguous density connecting them, and the designation as separate domains is more a result of the fold than the expected dynamics.

We next focused on the intramolecular linkers connecting vRNAP domains. As described above, a ~100-residue linker between NTD and RNAP (res. 907–1007) could span as much as 250 Å, explaining the challenge of aligning this domain with the rest of vRNAP during SPA. AlphaFold 3 predicts the NTD as two globular helical subdomains, and putative membrane-spanning α-helices (MSH) are predicted between residues 224 and 278 (Supplementary Fig. 4a). This prediction was validated in vitro, where the isolated gp50-NTD can be incorporated into membrane nanodiscs (Supplementary Fig. 4b,c), consistent with vRNAP membrane localization during infection [30]. Similarly, a 23-amino-acid linker, spanning residues 2102–2123, connects RNAP to the CTD and is invisible in our structure. Thus, vRNAP can be rationalized as three globular, folded units flexibly connected by at least two extended linkers. These linkers may explain why SPA of vitrified vRNAP also captured individual RNAP and CTD domains if the two domains are separate in a split complex. All possible resolved maps and 2D classes are explained by these three putative states (Fig. 3b).

The globular structures of both RNAP and CTD do not provide any clarification on how the entire assembly manages to squeeze through the ~35 Å diameter of the tail-tube [22].

We used SWORD2 [31] to perform structural decomposition of ΔN-vRNAP and to begin unraveling how this enzyme may unfold and refold into a fully functional enzyme upon translocation through a narrow channel across the double-membrane *E. coli* cell wall. SWORD readily decomposed RNAP and CTD into nine subdomains (Fig. 3c) with a high likelihood score. This peeled topology, along with the two globular subdomains predicted by AlphaFold 3 in the NTD, results in a total of 11 roughly globular and flexibly linked subdomains, resembling beads on a string (Fig. 3c).

A bulky CTD autoinhibits RNA polymerase activity

N4 RNAP domain adopts a typical single subunit RNAP (ssRNAP) fold that resembles a right hand with three subdomains: fingers, palm, and thumb (Fig. 4a). The CTD buries a large surface area of the RNAP occupied by the fingers and palm including the palm insertion and extended foot domains, but despite the extensive interface, the overall structure of the N4 RNAP domain is relatively unchanged in complex with the CTD (RMSD 3.7 Å) (Fig. 4a), barring a few changes in subdomains unique to N4 and a subtle shift of the fingers.

N4 vRNAP is highly specific for a ssDNA template and is transcriptionally active on such a template [25]. To assess if the ssDNA template induces conformational changes in vRNAP, we also performed cryo-EM SPA on vRNAP bound to a single-stranded P1 promoter (Fig. 4b). Initial 2D classes were nearly identical to the apo structure and again resolved two populations of the isolated RNAP and ΔN-vRNAP tight complex bound to P1 that we refined to 2.8 and 2.9 Å resolution, respectively (Table 1, Supplementary Fig. 3d,e). Similar to the apoenzyme, two populations of CTD alone and the loose complex were visible in 2D classes but unresolved in 3D. In both structures, the P1 promoter inserts deeply into the RNAP active site, and the DNA hairpin binds as expected (Supplementary Fig. 5a,b) based on previous crystallographic studies [32]. The isolated RNAP makes 10 hydrogen bonds (H-bonds) and 130 van der Waals contacts with P1. In contrast, RNAP in the tight complex makes 12 H-bonds and a total of 126 van der Waals contacts (Supplementary Fig. 5c). Surprisingly, the RNAP and CTD binding interface was unchanged in our apo and P1-bound structures, suggesting that the promoter's binding is insufficient for the release of CTD from RNAP. Even if the RNAP is globally unchanged in all observed structures, the localization of CTD (either with or without P1) results in allosteric changes in three crucial motifs of the RNAP core: the intercalating β-hairpin, the B-motif loop, and the N4 plug module (Fig. 4c). The intercalating β-hairpin is a conserved domain in ssRNAPs, which helps unwind dsDNA [33]. The B-motif is a conserved motif in ssRNAPs as well and is responsible for the coordination of incoming NTPs; however, in N4, it is unique in that a portion of the motif exists on a dynamic loop, the B-motif loop [24]. The N4 plug module is entirely unique to N4 vRNAP and is suggested to inhibit polymerase activity by occupying the active site in concert with the B-motif loop [24].

Extensive structural characterization of the recombinant N4 vRNAP catalytic core (mini-vRNAP) has been reported, including crystal structures of the apoenzyme, the promoter-bound enzyme, and early transcription initiation intermediates [24,25,32,34]. To compare our four cryo-EM structures to the existing crystal structures, we focused on the apo crystal structure and the P1 promoter-bound crystal structure (2PO4 and 3C2P, respectively). As expected, the most conformationally variable elements were the plug module, the intercalating β-hairpin, and the B-motif loop. To perform an all-to-all RMSD

comparison across the six structures, we first superimposed them onto the rigid scaffold, excluding the three mobile elements. RMSD measured on this rigid scaffold (996 residues) ranged from 0.4 Å to 2.4 Å, compared to 0.5 Å to 4.4 Å when all 1,093 comparable residues were included, confirming that the mobile subdomains account for the majority of apparent structural variation. Subdomain RMSDs were then measured using the rigid scaffold superposition. Full RMSD results are in Supplementary Fig. 6. This analysis revealed that within the cryo-EM dataset, the apo-isolated RNAP and promoter-bound structures showed only minor subdomain displacements, with a maximum RMSD of 1.1 Å in the plug module, indicating that promoter binding does not induce large conformational changes in the catalytic core. By contrast, when the apo isolated cryo-EM structure was compared to the mini-vRNAP apo crystal structure, the mobile subdomains diverged substantially: RMSD values of 6.1 Å, 5.9 Å, and 23.5 Å were observed for the plug module, intercalating β-hairpin, and B-motif loop, respectively, with a rigid scaffold RMSD of 2.4 Å. Notably, the mini-vRNAP apo crystal structure most closely resembled the tight complex cryo-EM structures, both apo and P1 promoter-bound, rather than the isolated RNAP domain reconstructions. This suggests that crystal packing in the mini-vRNAP apoenzyme stabilizes subdomain conformations that, in the context of the FL enzyme, are instead imposed by allosteric contacts with CTD. Given the heterogeneous mix of states, we further analyzed our cryo-EM densities using cryoSPARC's 3D variability, which revealed minor oscillation of the plug module and intercalating β-hairpin (Supplementary Movie 1) in the tight complex of the apo vRNAP though the B-motif loop and CTD remained unaffected.

Cryo-EM reconstructions of the four high-resolution states (Fig. 4c) provide clear evidence that the CTD drives the B-motif loop into the active site in a conformation that we predict is inactive. Without the CTD, the B-motif loop is folded into a β-hairpin that sits outside the active site. Interestingly, there is no direct contact between the CTD and B-motif loop, indicating that the hairpin loop is moved and released via an allosteric mechanism. In the P1-bound tight complex, the P1 promoter lifts the intercalating β-hairpin and the N4 plug module, but the B-motif loop still occupies the polymerase active site, suggesting that promoter binding is insufficient to release the B-motif loop and CTD. The P1 promoter is relatively unchanged in overall structure when comparing the isolated RNAP and the tight complex. A few contacts are added or gained due to a more relaxed structure in the isolated RNAP. The only major difference is that the –1 nucleotide is flipped away from the active site in the tight complex (Supplementary Fig. 5b), likely due to steric hindrance of the B-motif loop conformation that occupies the active site in the tight complex conformation. To experimentally determine the role of the CTD in transcription, we compared the enzymatic properties of FL-vRNAP, a recombinantly expressed RNAP domain construct (res. 998–2103), and the isolated CTD (res. 2124–3500). Radiolabeled runoff transcription assays were performed using a native N4 ssDNA template containing the P1 hairpin promoter (Fig. 4d, lanes 1–3). To assess the potential of the CTD to autoinhibit the RNAP domain, we titrated increasing concentrations of isolated CTD into reactions containing 100 nM RNAP domain, from a 1:0.5 molar ratio (RNAP:CTD) doubling stepwise to 1:16, reaching a final CTD concentration of 1.6 μM (Fig. 4d, lanes 4–9). FL-vRNAP; RNAP domain controls at 100 nM were included, and all reactions were run for 5 minutes. A modest reduction in RNA product was observed with increasing CTD, though inhibition only became apparent at high CTD excess (Fig. 4e),

supporting the weak binding interface observed from PISA analysis of the tight complex (Fig. 3a).

Given that inhibition required a substantial molar excess of CTD, we next examined the effect of enzyme concentration on transcriptional activity. Time courses were performed for both FL-vRNAP and the RNAP domain from 8 to 60 seconds at 100 nM, 1 µM, and 10 µM. At 100 nM and 1 µM, the two constructs displayed comparable activity. However, at 10 µM, FL-vRNAP showed a markedly reduced amplitude relative to the RNAP domain (Fig. 4f, Supplementary Fig. 7). Notably, the curves were otherwise similar between constructs, suggesting no difference in nucleotide addition rate, but rather a difference in the fraction of catalytically active enzyme, consistent with a concentration-dependent shift toward the tight complex conformation, in which the CTD occludes the active site.

Top-down analysis of N4 ejection proteins gp51 and gp52

N4 vRNAP (gp50) is part of a conserved *Schitoviridae* operon that encodes three ejection proteins: gp50, gp51, and gp52 (Fig. 5a). We took a top-down approach to identify the other two putative subunits, gp51 (644 aa) and gp52 (150 aa), which are homologous to phage DEV gp73 and gp72, respectively [16]. We purified N4 virions to homogeneity (Supplementary Fig. 8a,b) and determined the high-resolution structure of phage N4 tails from filled virions (FV) and empty particles (EP) to 3.2 Å and 3.6 Å, respectively (Fig. 5b,c, Supplementary Fig. 9a–e, Table 2). The tail has a total vertical length of 450 Å, formed by the HT-adapter (gp67), tail tube (gp54), and tail plug (gp53), and is attached to the head through interactions between the HT-adapter and the portal (gp59). Surrounding the tail is a six-fold symmetric non-contractile sheath formed from a heterodimer of the minor subunit gp64 and the major subunit gp65 [22].

Tails from EPs are remarkably similar in structure to FVs, and comparison of the similarly resolved maps (Fig. 5b,c) identified three critical differences after the genome is ejected from the capsid. First, the portal undergoes two major changes: the collapse of the barrel helices beyond residue 667 [20] (Supplementary Fig. 10a) and the narrowing of the portal tunnel from 33 Å to 25 Å due to restructuring of the tunnel loop (Supplementary Fig. 10b,c). Additionally, an elongated density is visible within the portal barrel of FV that disappears in EPs, strongly indicating that this density represents dsDNA ready for ejection (Fig. 5d). The density terminates abruptly at the base of the barrel, leaving the rest of the portal protein lumen empty. Second, the ejection protein gp51, which forms an α-helical hairpin adjacent to the wingtip domain of the portal, disappears in the empty virions (Fig. 5e, Supplementary Fig. 10a). A similar density for an ejection protein was also observed around the portal crown of phages DEV [16], Moo19, B2 [23], and Pa223[35], which reinforces our assignment of N4 gp51. Third, we identified 12 fragments of approximately 16 residues inside the lumen of the HT-adaptor, which are absent in EP (Fig. 5f), comprising the N-terminus of the ejection protein gp52 [16]. The FL gp52 consists of 150 residues, and the remainder of the protein likely orients downwards in the tail, where a recent asymmetric reconstruction of the N4 virion also identified spurious density [22].

Bottom-up analysis of recombinant N4 gp51 and gp52

To further characterize N4 ejection proteins, we purified recombinant gp51 and gp52 in *E. coli*, N4's natural host, and found that both proteins (Supplementary Fig. 11a) behave like membrane proteins, soluble only in the presence of 0.05% n-dodecyl-D-maltoside (DDM). Accordingly, MemBrain [16] detected putative membrane-spanning helices at the N- and C-termini of N4 gp51 (Supplementary Fig. 11b) and a transmembrane helix at the N-terminus of gp52 (res. 72–88) (Supplementary Fig. 11c).

We focused on gp51, whose homologous protein in phage DEV, gp72, is more soluble and amenable to structural analysis [16]. We generated several constructs of N4 gp51 to identify a less hydrophobic fragment and found that a C-terminally deleted construct spanning residues 1-359 could be purified and kept in solution with a low concentration of detergent (Supplementary Fig. 12a). This fragment, ΔC-gp51 was subjected to cryo-EM analysis and, despite a tendency to aggregate, yielded well-defined 2D class averages, which we used to reconstruct a structure at 3.3 Å resolution (Supplementary Fig. 12b,c, and Table 1). The ΔC-gp51 structure revealed a decameric hollow channel-like formation (Fig. 6a), with the protomer adopting a leg-fold similar to that of phage T7 gp15 [7] (Fig. 6b). Residues 52-255 were resolved out of 1 to 361. Residues prior to 52 were disordered, probably needing interactions with gp52 to form a stable structure, and residues beyond 255 were also disordered, likely needing the continuation of the gp51 C-terminal residues to stably fold.

AlphaFold 3 predicts full-length gp51 (FL-gp51) to adopt a slender, predominantly α-helical architecture, approximately 140 Å in length and 30 Å in width, resembling phage T7 gp15 [7] (Fig. 6b). Similarly, the smaller gp52 is predicted to form an elongated α-helical protomer lacking a well-defined hydrophobic core. Based on the experimentally observed decameric stoichiometry (Fig. 6a), we also asked AlphaFold 3 to predicted the gp51:gp52 decameric assembly (Fig. 6c). With moderate confidence scores, the predicted N4 gp51 and gp52 structure formed a hetero-decameric trumpet-shaped complex ~300 Å in length with an internal diameter ranging from ~20 Å to 40 Å (Fig. 6c), which is stabilized by extensive protein-protein interactions at their N-termini. This prediction is remarkably similar to the cryo-EM reconstruction of phage DEV ejection proteins gp72:gp73 despite the negligible sequence identity (<10%) between DEV and N4 ejection proteins (Fig. 6c), which nonetheless form nonameric oligomers in vitro. The AlphaFold 3 model of gp52 suggests that the protein forms a putative channel with an internal diameter of 40 Å, primarily composed of the decameric arrangement of an amphipathic α-helix (res. 69–131) (Supplementary Fig. 11c,d). Notably, N4 gp52 has genomic synteny, bioinformatic and structural homology to DEV gp71 [16] and T7 gp14 [4,5,7], which has been experimentally shown to insert into a membrane after ejection [4], specifically the bacterial OM [36].

Gp51 and gp52 integrate into lipid membranes and form membrane-spanning channels

Phage ejectosomes must provide a continuous channel from the phage head to the infected bacterium's cytoplasm. Previously, we showed that the *Pseudomonas* phage DEV utilizes the gp73 and gp72 proteins as OMC and PT proteins, respectively [16]. AlphaFold 3 predictions indicate that phage N4 gp51 and gp52 proteins form a complex and might fulfill similar roles in DNA ejection by the N4 phage across the cell envelope of *E. coli* (Fig. 6c). To test this hypothesis, we performed lipid bilayer experiments to determine whether gp52 forms membrane-spanning channels similar to DEV gp73. To

exclude contamination with endogenous porins of *E. coli*, gp51 and gp52 were purified from the *E. coli* (DE3) omp8 strain (Supplementary Fig. 11a), which lacks all major pore-forming proteins OmpF, OmpC, LamB and OmpA [37]. Control experiments with only electrolyte, or with protein buffer containing DDM added to the lipid membrane, showed no interaction with the membrane or pore activity (Supplementary Fig. 13a,b). By contrast, the addition of the purified gp52 protein to lipid membranes resulted in current increases indicative of pore formation (Fig. 6d). Interestingly, only occasionally did we observe a steady stepwise current increase (Supplementary Fig. 13c–e), which is typical of bacterial porins [38] or pore-forming toxins [39]. These gp52 pores were prone to channel closures as indicated by downward spikes (Fig. 6d). We also observed events of a sudden, large increase in membrane currents, which ruptured the membrane (Supplementary Fig. 13f–h). In many instances, we recorded unstable pores or membrane disturbances as indicated by rapidly fluctuating currents. Overall, we tested gp52 in 24 individual lipid membranes. These experiments demonstrate that gp52 interacts with membranes and forms membrane-spanning channels with channel conductance ranging from 0.45 nS to 1.9 nS.

We also examined whether the periplasmic tunnel protein gp51 interacts with membranes. Surprisingly, lipid bilayer experiments revealed that the purified gp51 protein has channel-forming activity (Fig. 6e, Supplementary Fig. 14a–f) in contrast to its DEV counterpart, gp72 [16]. Similar to N4 gp52, we observed stable gp51 pores with conductances ranging from 0.5 to over 5 nS. We also observed rapid and large insertions that frequently led to membrane rupture (Supplementary Fig. 14d, f), as well as flickering and noisy insertions (Supplementary Fig. 14a–c,e). Overall, we tested 18 individual membranes with the purified gp51 protein and observed frequent interactions with lipid bilayers and heterogeneous populations of channels. This is in contrast to the equivalent PT protein gp72 of the DEV phage [16], indicating differences in their capacity to interact with membranes, probably due to the presence of hydrophobic α-helices (Supplementary Fig. 11b), which are not found in DEV gp72. Notably, the gp51 concentrations needed to observe channel activity were at least 10 times higher than those of gp52, probably reflecting the different propensities of the individual proteins to insert into membranes. Overall, our data show that the phage N4 gp52 and gp51 proteins form heterogeneous channels in lipid membranes with widely varying physical properties, which differ from those of bacterial pore-forming outer membrane proteins that typically create stable channels in lipid bilayer experiments.

A tail-gating complex couples receptor binding to ejection protein release

The tail tube of phage N4, consisting of 12 copies of gp54, in both FV and EP, is plugged by a large protein with a globular core and a long, feather-shaped β-sheet protrusion that we identified as gp53 (Fig. 7a,b). A focused reconstruction of the tail plug hub (Fig. 7c) revealed a convincing density that we interpreted using an AlphaFold model of gp53 (Supplementary Fig. 15a–c) that was rebuilt into the cryo-EM density. The plug identity was further supported by gp53's presence in our LC-MS analysis (Supplementary Table 1), indicating this factor is a structural component of the N4 virion. While the majority of gp53 density sits outside of the tail-tube (Fig. 7c) and was readily modeled, a non-contiguous stretch of gp53 residues (29–56 and 670–716) forms a hydrophobic knot (Fig. 7c, Supplementary Fig. 15b) comprising a complex weaving of six short-alpha helices that

corks the tail-tube base. Density continues for a few residues beyond this knot before the remaining 160 residues are lost in the tail lumen. A poorly resolved double ring structure is observed (Fig. 7c), perhaps belonging to the remaining gp53 C-terminal residues and, possibly, multiple copies of ejection protein gp52, or a combination thereof.

The focused reconstruction of the N4 tail end uncovered another feature: a single protomer of the non-contractile sheath connects to the globular core of the tail plug gp53 (Fig. 7b), indicating asymmetry in the sheath assembly. In N4, the non-contractile sheath is formed by a heterodimer of gp65 and gp64 (Supplementary Fig. 16a,b). Long known to be the N4 receptor binding protein [40], gp65 has a modular organization consisting of 10 domains (D1–D10), and in the focused reconstruction gp65 tightly binds the core of gp53 through the final 200 residues, which comprise D9–D10 (Fig. 7b, Supplementary Fig. 16b). This asymmetric interaction between two domains of the non-contractile sheath (D9–D10 of gp65) and the tail plug forms a tail-gating complex (Fig. 7b), which seals the dodecameric tail tube, resulting in a 12:6:1 symmetry mismatch among the tail tube, the sheath, and the plug.

Finally, our reconstructions revealed that the N4 tail tube appeared sealed in both the full and empty particles (Fig. 5b,c). The unique interaction between the sheath and the plug could allow the plug protein to remain in proximity and rebind the tail base after DNA ejection, facilitated by the hydrophobic knot, thereby explaining this phenomenon at least in vitro. Further supporting the asymmetric interactions of the N4 non-contractile sheath, 3D classification using a focused mask on a single sheath assembly revealed two divergent classes (Fig. 7d). One showed the protomer slightly displaced, with D8 detached from the tail tube and bound to a neighboring protomer. In the other, the protomer was absent, but inspection at low contour revealed weak density away from the tail core, which would require a 90 ° rotation about the HT-adapter anchored gp65 C-terminus. This low-contour density likely results from a subset of particles in the final particle stack, with other particles having freely moving and unanchored sheath assemblies. This suggests that while one or two sheath proteins may be tightly anchored to the tail core by gp53, the other four might be free to detach and interact with the polysaccharide-rich environment of the extracellular space.

DISCUSSION

Ejection proteins, which are essential for infectivity and encoded by most *podoviruses*, exemplify the concept of conformational plasticity and structural adaptability perhaps better than any other protein in nature [3]. These encapsidated proteins undergo significant conformational and oligomerization changes during expulsion through the phage tail, leading to the formation of a megadalton-sized DNA-ejectosome that spans the cell envelope of Gram-negative bacteria. The DNA-ejectosome is not permanently open; otherwise, it would harm the host, so it must regulate its gate or potentially disassemble after the genome is ejected. Most of what is known about ejection proteins is inferred from studies in the model system T7, which is the best-characterized phage for studying genome ejection. Nonetheless, T7 is much simpler than phage N4, which is the prototype of a rapidly expanding and surprisingly abundant family of phages recently renamed *Schitoviridae* [41]. Not only is the N4 genome nearly twice the size of T7's, but N4's largest ejection protein, analogous to gp16 in T7, is a 3,500-residue RNAP ejected from the capsid, which has been studied as a unique RNAP for over 50 years [17]. N4 requires three

RNAPs—two encoded by the phage and one from the host. In this study, we used a combination of top-down and bottom-up approaches to elucidate the structure and activity of the phage N4 ejection proteins as well as how phage attachment to a primary receptor triggers the ejection of these proteins into the bacterium. We made four discoveries.

First, cryo-EM analysis of recombinant vRNAP establishes this protein as a single-chain multi-subunit RNAP, much more complex than phage or mitochondrial ssRNAPs, yet simpler than multi-subunit eukaryotic RNAP. This three-domain protein comprises a lipid-binding NTD loosely associated with the catalytic RNAP, which, in turn, neighbors a massive CTD. Single-particle analysis identified both a tight state of RNAP and CTD, as well as individual domains, suggesting a loose quaternary structure that was not affected by promoter binding. Although the CTD makes sparse contacts with RNAP, it allosterically regulates three crucial motifs of the RNAP core: the intercalating β-hairpin, the B-motif loop, and the N4 plug module, thereby repressing transcriptional activity. In vitro transcription assays confirmed this inhibitory role of the CTD, which we speculate can only be exerted when the two domains are in direct contact. Computational decomposition analysis provides evidence for beads-on-a-string architecture characterized by as many as eleven globular sub-domains of approximately 250–300 amino acids, each large enough to fit through the phage tail channel. We propose that this topology represents the pre-ejection conformation of N4 vRNAP inside the N4 capsid (Fig. 8a), which is ejected into the host cell envelope, where it assembles and refolds.

To date, only two virion-encapsidated phage RNAPs have been structurally characterized: N4 vRNAP and the vRNAP of *Crassvirus* phi14:2 (gp66) [42]. Despite sharing the distinction of being ejected polymerases, the two enzymes appear to be fundamentally distinct at a global structural level. Phi14:2 gp66 (~2,631 residues) is more closely related to eukaryotic single-subunit RNAPs involved in RNA interference and, while harboring domains with structural similarities to multisubunit bacterial RNAPs, is a single globular domain for which nearly the complete structure was resolved by crystallography. N4 vRNAP, by contrast, is a larger (3,500 residues) single-subunit, multidomain enzyme structurally more related to T7 RNAP and eukaryotic mitochondrial RNAP, and is distinguished by its loosely connected NTD, catalytic core, and CTD as discrete structural components. Whether these two polymerases share a common evolutionary origin or represent convergent solutions to the problem of virion-encapsidated transcription remains an open question. At the level of ejection strategy, however, some parallels between *Schitoviruses* and *Crassviruses* may exist. *Crassoviruses* appear to encapsidate five proteins in total [43], including two cargo proteins that may be functionally analogous to the membrane-spanning N4 gp51 and gp52, and three RNAP-associated proteins whose relationships to the distinct NTD, catalytic core, and CTD of N4 vRNAP warrant further investigation. Resolving the structures of the *Crassvirus* ejection proteins will be necessary to determine the extent to which ejection mechanisms are conserved across these two phage families.

Second, by combining bottom-up analysis of the isolated vRNAP and top-down studies of N4 virions, we discovered that the other two ejection proteins, gp52 and gp51, which are part of the same operon as vRNAP, are located in the pre-ejection conformation inside the tail and around the portal protein crown, respectively (Fig. 8a). After ejection, these proteins form an OM channel, like phage T7 gp14 [4,5], and a PT tunnel going from the OM to the IM, like T7 gp15 [7,13], which are first ejected during infection. This finding

reconciles previous work that proposed the ejection proteins reside in the tail of phage P22 [44], while others found evidence of capsid localization, loosely bound to the portal barrel [26]. Our reconstruction of N4 ΔC-gp51 reveals a decameric structure, unlike the nonameric gp72 of phage DEV [16] but similar to the homologous ejection protein gp12 of the *Shigella* phage Sf6 [45]. Despite the different stoichiometry, both DEV and N4 create a tunnel large enough to allow dsDNA ejection. The decameric stoichiometry, likely transmitted to the OMC gp52, suggests a possible symmetry mismatch with the dodecameric tail tube, unlike DEV, where a nonamer can form symmetric binding interfaces with a dodecamer. Nonetheless, N4 is a coliphage, and the recombinant gp51 purified from *E. coli* extracts using detergents suggests that a decamer, or possibly a nonamer, may also reflect an in vitro structural polymorphism, as observed many times in structural biology with portal proteins [46,47], small terminase subunits [48], and others.

Additionally, we observed a discrepancy between the stoichiometries of N4 gp51 and gp52 inside the virion (pre-ejection), where they are present in 12 copies, and after ejection, when gp51 assembles into a decameric tunnel. As previously reported for phage DEV [16], whose PT decreases from 12 to 9, and for phage T7 [7], which drops from 8 to 6, ejection from the virion is linked to subunit loss. The presence of extra copies of the ejection protein before ejection might suggest redundancy in encapsulating ejection proteins during assembly to compensate for the low fidelity of assembly in a DNA-ejectosome. We hypothesize that the additional PT and OMC subunits either remain in the capsid or are lost in the periplasm or OM during ejection.

Third, our study provides direct and indirect evidence to decipher how ejection proteins assemble into a DNA-ejectosome. Our experimental structure of gp51 reveals a leg-fold previously observed in phage T7 gp15 [7,13]. This slender fold lacks an intramolecular hydrophobic core and is supposedly small enough to be ejected through the poral protein channel. Similarly, the gp52 protomer predicted by AlphaFold, and similar to phage DEV gp73 [16], is quite elongated and may also be ejected through the tail lumen without a significant refolding. Assembly of gp51 and gp52 upon ejection would imply the formation of a quaternary structure and, thus, an intermolecular hydrophobic core. A different scenario involves the largest ejection protein, gp50. N4 vRNAP is likely unfolded or loosely folded into subdomains inside the virion, as shown by our inability to visualize the protein with single-particle analysis and by the distribution of bubbles in the bubblegrams that appear within 600Å of the virion, far from the unique vertex. The beads-on-a-string structure of N4 vRNAP probably plays a key role in helping this giant protein be ejected through the portal channel and tail lumen into the host.

Fourth, through focused reconstructions and mass spec analysis, we deciphered the structure of the N4 tail-gating complex, which is characterized by a unique 12:6:1 symmetry mismatch. We found that asymmetry in the receptor-binding sheath subunit gp65 generates a 3-way signaling platform formed by tail plug (gp53), the receptor-binding factor (RBP, or sheath, gp65), and the ejection proteins residing inside the tail (gp52) and capsid (gp51 and gp50). We propose that the tail-gating complex links phage binding to the bacterial receptor NfrA or its exported polysaccharide NGR (Fig. 8b), or a combination of both [40,49,50], to a coordinated and synchronized release of ejection proteins into the bacterial cell envelope. Surprisingly, both the N4 mature virion and empty particle have a density for the tail plug, suggesting that the plug is not lost in the environment after receptor binding but remains tethered to the tail complex, available for reattachment after

ejection proteins and DNA have been released. Although our cryo-EM data may not accurately represent an in vivo infection because empty N4 particles have released the DNA into a test tube rather than ejecting it into a cell via receptor binding, lateral displacement of the plug protein is indeed a common feature among *Siphophages* [51]. Notably, a tail-gating mechanism similar to that identified in this study for N4 was recently discovered through cryo-EM analysis of phage DEV tail tip (Supplementary Fig. 16c) [52]. A homologous plug protein, gp74, seals the DEV tail tube, but notably lacks the β-feather domain characteristic of N4 gp53. Despite this, the tail tube and plug protein are clearly structural homologs across the two phages. More striking is the divergence in how plug positioning is stabilized. Whereas N4 relies on the non-contractile sheath heterodimer, DEV employs an entirely distinct set of proteins. A trimeric assembly of gp56 anchors to the outer tail via a trimeric knob, with a coiled-coil domain engaging the plug protein gp74. Biochemical and computational evidence further suggests that this trimer is capped by a single copy each of gp55 and gp54, with the capped trimer serving as the receptor-binding protein in DEV. This mirrors the architecture in N4, where the plug is anchored by gp65, itself the receptor-binding protein. Thus, despite the vast structural differences between N4 gp65 and DEV gp56, both phages converge on the same functional principle of coupling plug retention to host recognition.

Based on the experimental evidence presented in this paper and a substantial body of data in the N4 literature, we propose a hypothetical model to describe the steps of N4 infection. The first step (Fig. 8c) is the nonspecific reversible adsorption of N4 to the host, which likely involves the interaction of 12 N4 appendages (gp66) with an exopolysaccharide, likely NGR, a polysaccharide putatively exported by N4's secondary receptor NfrA [40,50]. This step is conceptually similar to that of phages T4 and T7 [53] and facilitates the phage association with NfrA or NGR. Once the phage gets closer to the outer membrane (Fig. 8d), the tail-gating complex is disrupted through interactions with the receptor NfrA or NGR, which results in the lateral displacement and the dislodgment of the tail-gating complex, including the plug C-terminus. Displacement of these residues could act as a ripcord for tail luminal copies of gp52, serving as the initiator for ejection and insertion of the OMC into the outer membrane. The ejection cascade (Fig. 8e), initially powered by the pressurized genome stored inside the capsid, continues with the expulsion of gp51, which forms the PT spanning from OM to IM. In N4, the last C-terminal residues of gp51 are predicted to be hydrophobic and may form the actual channel crossing the IM. Together, the OMC and PT form the structural channel spanning the bacterial cell envelope, which we envision as a decamer based on our structure of ΔC-gp51 (Fig. 6a). Peptidoglycan hydrolytic activity is required to penetrate the PG layer, although N4 ejection proteins do not appear to harbor this catalytic activity, which is presumably provided in trans by some other factors. vRNAP is the last ejection protein to be expelled from the capsid (Fig. 8f). We propose that this large protein exists in the beads-on-a-string architecture illustrated in Fig. 3c, possibly maintained in a molten globule state inside the capsid due to the pressurized genome [54]. vRNAP is ejected through the newly formed OMC:PT channel spanning the entire cell envelope. It remains anchored to the DNA-ejectosome via hydrophobic interactions between its NTD and IM (Supplementary Fig. 4b–c) and through direct association with gp51, which we demonstrate for the related DEV ejection proteins gp72 (PT) and gp71 (vRNAP)[16]. This

explains why the N4 transcriptional apparatus is associated with the cytoplasmic membrane during infection [30].

We envision that two copies of vRNAP can associate with the decameric PT, based on the observation that vRNAP is present in 4 ± 1 copies inside the capsid [20]. In the host cytoplasm, without the pressure from the genome within the capsid [55], the subdomains likely refold to generate a compacted and catalytically active RNAP and CTD. This transcriptionally active state may allow the protein to promote DNA-dependent transcription of the first three promoters, pulling the genome into the bacterium. We speculate that each promoter (P1–P3) interacts with a different copy of the RNAP, which is specifically active when the single-stranded promoter hairpin associates with the RNAP. The specific role of CTD in early transcription remains unknown. Given its large size, which is conserved throughout *Schitoviruses* [16], it likely works alongside RNAP during early gene transcription, possibly by binding to or unwinding DNA. After N4 second RNAP (RNAP II) has been translated, and the phage moves into middle transcription, intramolecular closing of gp50 RNAP and CTD (Fig. 8g) would dampen transcriptional activity due to the intermolecular inhibitor effect of the latter, reconciling the observed function of the C-terminus as autoinhibitory of transcriptional activity. We speculate that vRNAP inhibition is part of the viral mechanism that allows N4 to achieve a larger-than-average burst size, estimated at about 3,000 virions per cell [56]. N4 has a somewhat unique delayed lysis phenotype, taking approximately 3 hours to fully lyse the bacteria. vRNAP inhibition may be necessary to suppress transcriptional activity during viral genome replication and virion assembly, thus preventing off-target transcription. However, we cannot rule out indirect effects mediated by the CTD, as a previous report indicated that N4 vRNAP CTD is required for genome replication, possibly suggesting a functional coupling between early transcription and replication control [57].

In summary, this study provides a framework for understanding the structure, activity, and plasticity of the giant and mysterious vRNAP, which is ejected by phage N4 and likely all *Schitoviridae* into Gram-negative bacteria. This polymerase is a single-chain, multi-domain RNAP that exemplifies a remarkable case of phage adaptation to its host. Future research will need to determine how vRNAP unfolds during ejection and whether its transcriptional activity is also part of the genome ejection motor that enables *Schitoviridae* to deliver their genomes into Gram-negative bacteria.

METHODS

N4 phage preparation for cryo-EM

A 100 mL culture of *E. coli* MG1655 (GenBank: U00096.3) in LB was grown at 37°C up to OD_{600} = 0.2 (about 7.5 x 10^7 CFU mL^{-1}) and infected with N4 (GenBank: EF056009.1) at a multiplicity of infection (MOI) of 3. After 3 hours (h) of incubation at 37 °C with shaking, chloroform (0.5 mL) was added, and the incubation continued for 15 minutes (min). The lysate was centrifuged at 5,000 x g for 15 min, and the supernatant was recovered and filtered through a 0.45 µm filter. 58 g L^{-1} NaCl and 105 g L^{-1} polyethylene glycol (PEG) MW 6,000 were dissolved in the supernatant. The solution was incubated for 16 h at 4°C before pelleting the phage particles by centrifugation at 20,000 x g at 4 °C for 30 min. The pellets were resuspended in TM buffer (10 mM Tris-HCl, pH 8, 10 mM $MgCl_2$) and centrifuged on a $CsCl_2$ 1.3 to 1.6 g cc^{-1} step gradient, top to bottom, pre-formed in

polyallomer ultracentrifuge tubes for Beckman rotor SW41. Phages in TM (3 mL) were applied to the top of the gradient, and tubes were centrifuged at 100,000 x g for 120 min at 4 °C in a Beckmann Optima XE-90 ultracentrifuge using a SW41 rotor. The phage bands, which usually sediment in the 1.5 g cc^{-1} step, were extracted from the tubes with a syringe, transferred into polyallomer tubes for the SW60 Beckman rotor, and centrifuged for ca. 16 h at 150,000 x g using a SW60 rotor. The phage bands were collected as above, dialyzed 2 x for 20 min against water and 16 h against TM buffer, filtered through 0.22 µm filters, and stored at 4 °C.

Mass spectrometry and SDS-PAGE analysis of N4 virion proteins

Virion proteins were extracted from 7–8 x 10^{11} pfu prepared for cryo-EM analysis as described by mixing 1.5 mL of the phage suspension in TM with 1.5 mL of methanol and 1.125 mL of chloroform. After vigorous mixing, the sample was centrifuged for 5 min at 16,873 x g in a microfuge, and the upper fraction was discarded. The lower fraction and the interface were mixed with 1.3 mL of methanol and centrifuged as above. The protein pellet was dried and resuspended in 6 M urea dissolved in 10 mM Tris-HCl, pH 7.4. The viral proteins were analyzed both by 15% SDS-PAGE and Coomassie staining and by mass spectrometry at UNITECH OMICs (University of Milano, Italy) using Dionex Ultimate 3000 nano-LC system (Sunnyvale CA, USA) connected to Orbitrap Exploris™ 240™ Mass Spectrometer (Thermo Scientific, Bremen, Germany) equipped with nano electrospray ion source. Peptide mixtures were pre-concentrated onto a PepMap 100 – 0.3 x 5 mm C18 (Thermo Scientific) and separated on EASY-Spray column ES902, 25 cm x 75 µm ID packed with Thermo Scientific Acclaim PepMap RSLC C18, 3 µm, 100 Å using mobile phase A (0.1 % (v/v) formic acid in water) and mobile phase B (0.1 % (v/v) formic acid in acetonitrile 20/80, (v/v)) at a flow rate of 0.300 µL min^{-1}. The temperature was set to 35°C, and 5 µL samples were injected in triplicate. MS spectra were collected over an m/z range of 375–1500 Da at 120,000x resolution, operating in data-dependent mode, cycle time 3 sec between master scans. HCD was performed with a collision energy set at 35 eV. Polarity: positive. Data were processed using Proteome Discoverer 2.5 software (Thermo Scientific, USA) with the search database set as Escherichia phage N4 (sp_tr_incl_isoforms TaxID=2886925_and_subtaxonomies) (v2024-03-27) and trypsin as the digestion enzyme. The following filters were applied: Protein level, ≥ 2 peptides; Peptide level, Xcorr ≥ 2.2, Rank = 1, Confidence = high; PSMs level: Xcorr ≥ 2.2.

Molecular cloning, expression, and purification of recombinant proteins

The genes encoding gp50 full length (FL-gp50), gp50-N (1–906), RNAP (998–2103), gp50 CTD (2124–3500), gp51 (1–644), and gp52 (1–150) were amplified by PCR from N4 DNA and cloned between XhoI and BamHI restriction sites in pET-16b. ΔC-gp51 was generated by introducing a stop codon at residue 360 of the gp51 gene using site-directed mutagenesis. The constructs were expressed in BL21-AI (Invitrogen), NiCo21(DE3) *E. coli* expression strain (New England Biolabs), or *E. coli* (DE3) omp8 strain, supplemented with appropriate antibiotics.

The cultures were grown in L.B. medium at 37 ºC until an optical density at 600 nm (OD_{600}) = ~0.3, when the temperature was dropped to 28 ºC until an OD600 = ~0.6, and were induced with 0.5 mM IPTG and 0.2% L-arabinose for 3-4 h. In the case of gp52, the

culture was induced at OD_{600} = ~1 with 0.5 mM IPTG and 0.2% L-arabinose for 2 h at 37 ºC. For FL-gp50, the cell pellet was lysed by sonication in lysis buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM $MgCl_2$, 5% glycerol, 1 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, 20 µg mL^{-1} DNase). After centrifugation at 39,000 × g for 30 min, 4 ºC, the soluble fraction containing the protein was subjected to HiTrap Heparin HP column and washed with wash buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM EDTA) and eluted with buffer B (10 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 1 mM EDTA). Fractions containing protein were pooled and buffer-exchanged to wash buffer using an Econo-Pac 10DG column (BioRad). Gp50 tends to reversibly salt out in a buffer containing less than 50 mM NaCl. Next, the protein was subjected to HiTrap DEAE FF (Cytiva) and eluted with buffer B listed above. Finally, the protein was polished by SEC using a Superdex 200 16/60 column (Cytiva) equilibrated with gel filtration buffer (20 mM Tris-HCl, pH 8.0, 75 mM NaCl, 2.5% glycerol, 0.5 mM EDTA, 0.0075% DDM, 0.1 mM PMSF). For other gp50 constructs, the cell pellets were lysed by sonication in lysis buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 2 mM $MgCl_2$, 5% glycerol, 0.1% Triton X-100, 1 mM PMSF, 20 µg mL^{-1} DNase). After centrifugation at 39,000 × g for 30 min at 4 °C, the soluble fraction was incubated with Low Density Nickel Agarose beads (GoldBio) for 2 h with rotation at 4 °C. The beads were washed with a wash buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 2 mM $MgCl_2$, 2.5% glycerol, 1 mM PMSF, 5 mM imidazole) and eluted with wash buffer containing 20–320 mM imidazole. The fractions containing protein complex were further purified by SEC using a Superdex 200 16/60 column (Cytivia) equilibrated with gel filtration buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2.5% glycerol, and 0.1 mM PMSF.

For gp50-N (1–906), gp50-RNAP (res. 998–2103), gp51 (1–644), ΔC-gp51 (1–359), and gp52 (1–150), the protein was isolated from inclusion bodies. The cell pellet was lysed by sonication in lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 4 mM $MgCl_2$, 2.5% glycerol, 2 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, 20 µg mL^{-1} DNase). After centrifugation at 39,000 × g for 30 min, 4 ºC, the pellet fraction containing the protein was resuspended in solubilization buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1% glycerol, 0.35% N-Lauroylsarcosine) and allowed to solubilize by rotating for 1:30 h at room temperature. The supernatant, after centrifugation at 39,000 × g for 30 min at 4 ºC, was incubated with Ni-NTA resin (GoldBio) for 2 h. The resin was packed in a column and washed with 50–100 CV refolding buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM $MgCl_2$, 1% glycerol, 0.05% DDM, 1 mM PMSF) supplemented with 5 mM imidazole. The proteins were eluted stepwise with 20–600 mM imidazole. The eluted protein was placed in Snakeskin™ Dialysis Tubing 10k MWCO (Thermo Scientific) and gently agitated in dialysis buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 1% glycerol, 0.05% DDM, 1 mM PMSF overnight at 4°C. After dialysis, the proteins were gel-filtrated on a Superose 6 10/300 column (Cytiva) containing dialysis buffer. ΔC-gp51 was dialyzed and gel filtrated in buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM MgCl2, 1% glycerol, 0.01% LMNG, and 1 mM PMSF. For lipid bilayer experiments, gp51 and gp52 were expressed in *E. coli* (DE3) omp8 strain. The final gel filtration buffer was 20 mM HEPES, pH 7.4, 100 mM NaCl, 0.05% DDM. Oligonucleotide sequences used for cloning are listed in Supplementary Data 1.

In vitro transcription assays

For the CTD titration, transcription reactions comprised of reaction buffer (40 mM Tris-Acetate, pH 7.9, 50 mM KCl, 0.2 mg mL^{-1} BSA) and 100 nM of FL-gp50, RNAP, or CTD were incubated with 10 µM DNA template at 4 °C for 10 min, 25 °C for 10 min, and 37 °C for 10 min. For the RNAP:CTD titration, RNAP remained constant at 100 nM, and CTD concentration doubled stepwise beginning at 50 nM and ending at 1.6 µM. The DNA template consisted of the P1 promoter (underlined) with 30 bp of native DNA sequence on either end (5'-TAGTTGATTGATAAGATACGGACAACATGTCCATAAGTTGCGAAGCAACGTGTAAC GTGTACAAGGTGGGGTAGGCTAAGG-3'). Transcription was initiated with NTP mix (40 mM Tris-Acetate, pH 7.9, 50 mM KCl, 0.2 mg mL^{-1} BSA, 6 mM Mg-Acetate, 0.05 mg mL^{-1} heparin, 1 mM ATP, 1 mM GTP, 1 mM UTP, 20 µM CTP, 5 nM α-32P-CTP) at t=0. Reactions were quenched by adding RNA loading dye (90 % formamide, 25 mM EDTA-Na pH 8.5, 10 mg mL^{-1} bromophenol blue) at t=5 min. Reactions were vortexed, boiled (95 °C), and loaded on a TBE-urea-polyacrylamide (28%) gel. Gels were run at a constant 25 W until the dye front ran into the bottom reservoir of the TBE buffer. Gels were exposed to phosphor screens for 16 h and scanned using an Amersham Typhoon imager. Band density was determined with ImageQuant 1D gel analysis. Experiments were performed in triplicate. Background was subtracted using an empty signal from the CTD lane and then normalized to the signal from the FL-vRNAP lane on a per gel basis.

For the concentration-dependent time courses, transcription reactions comprised reaction buffer (40 mM Tris-Acetate, pH 7.9, 50 mM KCl, 0.2 mg mL^{-1} BSA) and varying concentrations of FL or RNAP (100 nM, 1 µM, and 10 µM) were incubated with 10 µM DNA template at 4 °C for 10 min, 25 °C for 10 min, and 37 °C for 10 min. Transcription was initiated with NTP mix (40 mM Tris-Acetate, pH 7.9, 50 mM KCl, 0.2 mg mL^{-1} BSA, 6 mM Mg-Acetate, 0.05 mg mL^{-1} heparin, 1 mM ATP, 1 mM GTP, 1 mM UTP, 20 µM CTP, 5 nM α-32P-CTP) at t=0. Reactions were quenched over time (8 s, 15 s, 30 s, 60 s) by adding RNA loading dye (90 % formamide, 25 mM EDTA-Na pH 8.5, 10 mg mL^{-1} bromophenol blue). Gels were run and imaged, and band density was measured as described above. Gels were run in triplicate and plotted using the raw, unnormalized band density.

Immunoblotting with anti-His antibody

Gel-filtrated fractions of FL-gp50 were resolved by SDS-PAGE on 10% polyacrylamide gel and transferred onto PVDF membrane (Immobilon®-P, Millipore®) using wet transfer overnight at 30 V in the Transfer buffer containing 25 mM Tris, 192 mM glycine, 20% methanol, 0.05% SDS. Membranes were rinsed three times for 5 min each in TBST (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.1% (v/v) Tween-20) and blocked in 5% (w/v) non-fat dry milk in TBST for 1 h at room temperature with gentle shaking. Following blocking, the membrane was incubated overnight at 4 °C with mouse anti His-Tag primary antibody (ABclonal, AE003) diluted 1:5,000 in blocking buffer. Membrane was washed three times for 5 min each in TBST and then incubated with HRP-conjugated anti-mouse IgG (Promega, W4028) diluted 1:10,000 in blocking buffer for 1 h at room temperature. After three additional washes in TBST, signal was detected using SuperSignal™ West Pico PLUS chemiluminescent substrate (Thermo Scientific™, 34577) and visualized using BIO-RAD ChemiDoc™ MP imaging system. gp50 CTD was used as a positive control.

Reconstitution in membrane nanodiscs

Membrane scaffolding protein 1E3D1 (MSP1E3D1) was expressed in NiCo21(DE3) *E. coli* expression strain (New England Biolabs) supplemented with 35 μg mL^{-1} kanamycin. The cultures were grown in 2xYT medium at 37 ºC until OD_{600} = ~0.5, when the temperature was dropped to 28 ºC until an OD600 = ~1, and were induced with 0.5 mM IPTG for 3–4 h. The cell pellet was lysed by sonication in lysis buffer containing 20 mM HEPES, pH 7.5, 200 mM NaCl, 20 mM $MgSO_4$, 1 mM PMSF, and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). The MSP1E3D1 lysate was clarified by ultracentrifugation at 39,000 × g for 30 min at 4 °C and the recovered supernatant was supplemented with 50 mM sodium cholate and incubated with Ni-NTA resin (GoldBio) for 2 h with rotation at 4 °C. The resin, after packing into a column, was washed with a wash buffer containing 20 mM HEPES, pH 7.5, 200 mM NaCl, and 20 mM Imidazole, and MSP1E3D1 was eluted with a buffer containing 20 mM HEPES, pH 7.5, 200 mM NaCl, and 400 mM Imidazole. The eluted protein was placed in Snakeskin Dialysis Tubing (10 k MWCO; Thermo Scientific) containing TEV protease (GenScript) and then in dialysis buffer consisting of 20 mM HEPES, pH 7.5, and 150 mM NaCl. After dialysis, the mixture was repassed through Ni-NTA resin to recapture cleaved His-tag and TEV protease, and the flow-through was collected. Untagged MSP1E3D1 was concentrated to 10 mg mL^{-1} and stored in -80 °C freezer.

Nanodiscs were assembled by mixing purified gp50-N at 5.7 mg mL^{-1} with 16:0-18:1 PC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) (POPC) (Avanti Polar Lipids, Inc) at 15 mg mL^{-1} and untagged scaffolding protein MSP1E3D1 at 10 mg mL^{-1}, resulting in a molar ratio of 1:500:5. The mixture was incubated for 1 h in the dark at 4 °C before adding 100 mg Bio-Beads SM-2 adsorbents (BioRad) and proceeded for overnight incubation in the dark. The mixture without the Bio-Beads was incubated with Low-Density Nickel Agarose beads (GoldBio) for 30 minutes at 4 °C with rotation. The beads were washed with wash buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM imidazole) and eluted with wash buffer containing 300 mM imidazole. The mixture was injected onto a Superdex 200 10/300 (Cytiva) column, which was equilibrated with a buffer containing 20 mM HEPES, pH 7.5, and 75 mM NaCl. Throughout the purification process, protein samples were set aside to run on 12.7% SDS-PAGE and 1.5% native agarose gels.

Vitrification, cryo-EM screening, and data collection

Recombinantly expressed FL-gp50 was vitrified on 300-mesh copper R 1.2/1.3 Quantifoil holey carbon grids. Both the apo and the P1 promoter-bound vRNAP were frozen together. Grids were glow-discharged using a Tergeo-EM Plasma Cleaner for 30 s with power at 20 W. 3 μL of purified protein at 0.5 mg mL^{-1} was applied to grid. Grids were blotted using a Vitrobot Mark IV (FEI) for 5.5 sec using a blot force of 4 before being vitrified immediately in liquid ethane. Grids of ΔC-gp51 were prepared in the same way as the N4 vRNAP grids. Purified N4 Virion was vitrified on 300-mesh copper R 2/1 C-Flat holey carbon grids. Grids were glow-discharged with identical settings as N4-vRNAP. A total of 6 μL of N4 mature virions at 1 x 10^{12} pfu was applied. A double blotting strategy was used first pipetting 3 μL onto the grid and manually blotting while simultaneously adding the next 3 μL. Following this application, grids were blotted using a Vitrobot Mark IV (FEI), for 5 sec using a blot force of 4 before being vitrified immediately in liquid ethane.

Both purified gp50 and N4 grids were screened in house at the University of Alabama at Birmingham Cryo-EM Core, on a 200 kV Glacios 2 equipped with a Falcon 4i detector. Grids for the Apo and P1 gp50 were shipped to SLAC for full data collection on a 300 kV Titan Krios with a Falcon 4i detector. Micrographs were collected with a pixel size of 1.217 Å at 120,000x magnification, using a total dose of 50 e-/Å^2 and a defocus range of -0.5 µm to -2.5 µm. A total of 13,007 and 5,323 movies were collected for the Apo and P1 gp50, respectively. For the N4 virion, three complete data collections were carried out and merged at the UAB Cryo-EM Core. Identical preparation and settings were used for vitrification and data collection of the virion grids, except for the final collection, where the sample was incubated at 42 °C for 10 minutes before vitrification, in an effort to enrich for empty particles. EPU software was used for data collection using the fast-positioning mode. Micrographs were collected with a pixel size of 1.19 Å at 120,000x magnification, using a total dose of 40 e-/Å^2 and a defocus range of -0.5 µm to -2.5 µm. For ΔC-gp51, 9,068 Micrographs were collected on a 200 kV Glacios 2 microscope using a Falcon 4i detector using a pixel size of 0.71 and a total dose of 60 e-/Å. Further collection parameters are in Tables 1 and 2.

N4 virion bubblegrams

Virions were vitrified as described in the previous section. Grids were imaged at the UAB Cryo-EM core on a 200 kV Glacios 2 equipped with a Falcon 4i detector. Nine movies were collected with a total dose of 250 e-/Å^2, capturing 39 FVs and 1 EP. Cryo-SPARC was used to fractionate the movies during Patch Motion Correction into 9 static overlapping frames each with a total dose of 50 e-/Å^2 with 25 e-/Å^2 size steps between each frame (1st frame: 50 e-/Å^2 total dose, 50 e-/Å^2 accumulated dose, 2nd frame: 50 e-/Å^2 total dose, 75 e-/Å^2 accumulated dose, 3rd frame: 50 e-/Å^2 total dose, 100 e-/Å^2 accumulated dose, etc.). Out of 39 virions, 37 had identifiable tails, and a point was chosen at the interface between the tail and capsid using ImageJ. Next, the center of all visible bubbles was chosen, and the distance between the tail:capsid interface and bubble center was measured on a per virion basis. The Mclust package was used in R to fit the data to a trimodal distribution, and ggplot2 was used to overlay the distribution and the bubble histogram.

Cryo-EM single particle analysis

All steps of SPA were carried out using cryoSPARC software [58] using Patch Motion Correction and Patch CTF Estimation. For the vRNAP reconstruction, blob picking followed by 2D classing and template picking resulted in 2D classes representing four putative species: the RNAP domain by itself, the CTD by itself, and two conformations of the RNAP and CTD in complex. The isolated RNAP domain was resolved with the standard ab initio to non-uniform pipeline. Initial attempts to resolve the three remaining states were difficult due to preferred orientation. A two-class ab initio using particles of all states resulted in an isotropic RNAP map and an anisotropic CTD map. A map was reconstructed using Homogenous Reconstruction only, using the RNAP-CTD tight complex particles determined by 2D classification, but with alignments from the isotropic RNAP map generated from the 2-class ab initio. The resulting map from homogeneous reconstruction only had isotropic density for the RNAP domain and anisotropic density for the CTD. Heterogeneous Refinement was performed with six maps. An isotropic RNAP

map, an isotropic RNAP-anisotropic CTD map, an anisotropic CTD map, and three junk classes generated by running a multi-class ab initio for a single iteration. This Heterogeneous Refinement resulted in maps of similar quality. The mixed iso-RNAP/aniso-CTD map was further subjected to a 25-class 3D classification, where a single class representing 8.7% of the input particles (~12,000 out of ~141,000) had an isotropically resolved CTD that further resolved to 3.8 Å. The RNAP-CTD loose complex was represented by a single class, though it could be refined to high resolution with particles that belonged to the RNAP domain only classes. With a medium-resolution map of the RNAP-CTD complex, templates were simulated using the Create Templates job and then used for template picking, followed by 2D classing to remove obvious junk (carbon edge, ice, etc). The resulting stack of 1.2 million particles was then used for another Heterogeneous Refinement with nine input maps: 3 RNAP domain focused states (High-Res, Med-Res, RNAP-CTD loose complex, 2 RNAP-CTD tight complex states (Isotropic, Anisotropic), 1 CTD isolated map, and three junk classes, which resulted in final particle stacks for the RNAP domain map and the RNAP-CTD tight complex map. During the processing of the P1-bound
(5'- TGCCTCCCAGGCAGTCAAAAGTTGCGAAGCAAC-3') gp50, the similarity to the apoenzyme was apparent, and 2D classes and initial maps were nearly identical with the addition of the hairpin DNA promoter. As this was the case, P1 states were solved using a Heterogeneous Refinement job with the input models from the apoenzyme processing and the resulting maps had clear DNA density that was absent from the input models. With the final particle stacks in hand, Reference-Based Motion Correction, Local CTF Refinement, and 3D Classification were used to improve the resolution further, resulting in four final high-resolution maps. For the RNAP-CTD loose complex, we leveraged the high-resolution CTD from the tight complex. Any attempts to resolve this conformation resulted in an isotropic RNAP and an anisotropic CTD. The map of the loose complex from the 9-class Heterogeneous Refinement was aligned and resampled to the tight complex map in ChimeraX [59] with respect to the tight complex CTD. A mask was then generated around the CTD in ChimeraX [59]. The realigned loose complex map was then imported into CryoSPARC [58], and the original map and particles were aligned to the imported map using the Align 3D Maps job. The mask around the CTD was used to locally refine the CTD of both the tight conformation and the loose conformation particles, with hopes that the anisotropy of the loose complex CTD could be resolved by the isotropy of the tight complex. The resulting loose complex particles were reconstructed without the tight complex particles, although this time the CTD was isotropic; however, the RNAP was poorly resolved. 3D classing on this particle stack with the CTD-focused alignments resulted in a single class of ~18,000 particles that had the RNAP and CTD equally resolved. The appearance of density missing from the high-resolution CTD from the tight complex (res. 2643–2663), which appears to be critical for the stabilization of the loose complex, gives us high confidence in the reality of the density.

For the N4, virion particles were selected using Blob Picker, followed by the removal of junk particles through 2D classification. High-quality picks were used to train a Topaz model for both full particles and empty particles. The results of Topaz picking yielded 40,673 and 25,795 total full and empty phage particles, respectively. An initial 3D map was generated through ab initio reconstruction with icosahedral(I) symmetry enforced. With refined icosahedral capsids, the Symmetry Expansion job was used to expand I

symmetry. A cylindrical mask was created in ChimeraX [59] and resampled to the capsid map to loosely cover the pentamer at an icosahedral vertex. Using this as a focus mask, 3D Classification was used to sort out the symmetry-breaking vertex with the phage tail. cryoSPARC [58] employs only one icosahedral symmetry, I in cryoSPARC [58] and I2 by RELION convention (the symmetry axis placed on the two-fold axis at the icosahedral edge-midpoint). Volume Alignment Tool was used to move the five-fold axis into the z-axis, whereby another round of 3D classification was used with C12 symmetry enforced to separate out the symmetry-mismatched phage tails based on portal symmetry. Duplicates were removed, yielding a total particle number of 30,374 and 19,338 for full and empty particles. Using Volume Alignment Tools, both empty and full particles were recentered and re-extracted to the portal to be used for the C12 portal reconstructions, as well as recentered and re-extracted further down the tail to be used for the C6 tail reconstructions. Resolution of the asymmetric tail-gating complex began after a focused 3D classification of a single tail sheath heterodimer. This classification led to two states (detached and displaced), and in both of these states, the poorly resolved density of the plug appeared slightly improved. As we observed no major differences in these features when comparing full and empty particles, both particle sets were combined during these steps. To ensure that the plug was present in both full and empty particles, the stacks were separated and reconstructed using Homogeneous Reconstruction Only, yielding strong density for the plug and all tail components in both full and empty maps. Local CTF refinement and Ewald Sphere correction were used to improve resolution. Subsequent refinements and particle cleanings through 3D Classification led to final particle stacks. For ΔC-gp51, after initial blob picking, iterative rounds of template picking and Topaz picking were used to generate the initial particle stack. 2D classing narrowed this particle stack to roughly 200,000 particles total. The initial model was generated using ab initio refinement without symmetry applied. Using the initial asymmetric model as well as 2D classes, the correct symmetry was determined to be C10. Subsequent rounds of Non-Uniform Refinement and Local Refinement, followed by Reference Based Motion Correction and a final round of 3D classification, led to a final particle stack of 77,357 with a resolution of 3.3 Å. Final particle numbers for all maps referenced are indicated in Tables 1 and 2. The final densities were sharpened using *phenix.auto_sharpen* [60]. Electron density maps were displayed using ChimeraX [59].

Model building and refinement

All de novo atomic models presented in this paper were built using Coot [61] or ChimeraX [59], including the ISOLDE plugin [62]. AlphaFold 3 [63] models guided by residue placements from ModelAngelo [64] were used as starting templates, and gaps or poorly predicted regions were filled in de novo when present. Five maps and resulting models were generated from the FL-vRNAP datasets (Table 1): (i) a 2.8 Å map of the recombinant N4 vRNAP with the RNAP and CTD in the tight complex. Residues 1007–3500 (full length 3,500 aa) were modeled using this map. Density for this conformation was continuous, barring a 23-residue linker connecting the RNAP and CTD (2102–2124), and two loops in the CTD: 21 residues from 2643-2663 and 7 residues from 2522–2528. (ii). a 2.6 Å map of the recombinant N4 vRNAP with the RNAP domain only, residues 1007–2101. (iii) a 4.3 Å map of the recombinant N4 vRNAP with the RNAP and CTD in the loose complex. (iv) a 2.9 Å map of the recombinant N4 vRNAP with the RNAP and CTD in a

tight complex with the P1 promoter bound, with identical residues modeled in (i). (v) a 2.8 Å map of the recombinant N4 vRNAP with the RNAP domain only (1007–2101).

Reconstruction of the N4 virion yielded five maps and corresponding relative atomic models (Table 2). (vi) a 3.2 Å C6 averaged map of the full particle tail. Models build using this map were the tail tube (gp54: Full Length), the adaptor (gp67: Full Length), the sheath large subunit (gp65: 1–1184), the sheath small subunit (gp64: Full Length), the N-terminus of the appendage trimer (gp66: ~1–80), and the portal (gp59: 19–693). Inside the tail tube at the portal tail interface, we observed a density, which, through de novo modeling, we identified as the putative outer membrane pore (gp52: 1–16). (vii) a 3.6 Å C6 averaged map of the empty particle tail. Models build using this map were the tail tube (gp54: Full Length), the adaptor (gp67: Full Length), the sheath large subunit (gp65: 1–1184), the sheath small subunit (gp64: Full Length), the N-terminus of the appendage trimer (gp66: ~1–80), and the portal (gp59: 19–693). (viii) a 2.9 Å C12 averaged map of the full particle portal. The portal (gp59: 19–693) was the sole model refined against this map. (ix) a 3.2 Å C12 averaged map of the empty particle portal. As with the prior map, the portal (gp59: 19–693) was the sole model refined against this map. (x) A locally refined C1 map of the N4 tail, including the tail-gating complex. Models built from this map include the plug gp53 (res. 9–724) and the final residues of gp65 (res. 1183–1382). All atomic models were refined using several rounds of rigid-body, real-space, and B-factor refinement using phenix.real_space_refinement [65] and validated using MolProbity [66]. Refinement statistics are in Table 2.

Structure analysis, AlphaFold prediction, and bioinformatics tools

All renderings of protein structures and map surface representations were generated using ChimeraX [59]. Model analysis and inspection were carried out in PyMol [67]. Structural comparison and identification were done using the Foldseek server [68]. Binding interfaces were analyzed using PISA [69] and PDBsum [70]. All RMSDs in the Cα position between superimposed structures were calculated using Coot [71]. Decomposition of vRNAP cryo-EM structure into protein units was carried out using the web server SWORD2 [31]. Ejection proteins gp51 and gp52 nonameric models were generated using AlphaFold 3 [63], whereas other monomeric models used in this study were generated with AlphaFold [72] and AlphaFold2 [73]. Membrane insertion was predicted using MemBrain [74].

Lipid bilayer measurements

Lipid bilayer experiments were performed in a custom-made lipid bilayer apparatus [75]. Briefly, a Teflon cuvette with 10 mL volume is separated into two compartments (cis- and trans-) by a wall with apertures of approximately 0.3 mm or 1 mm in diameter. Ag/AgCl electrodes were bathed in a 1 M KCl, 10 mM HEPES, pH 7.4, electrolyte solution. The cuvette was primed on both sides of the aperture with 2% diphytanoylphosphatidylcholine (DPhPC; Avanti Polar Lipids) in chloroform. Then the cuvette was filled with 10 mL of electrolyte solution. Lipid membranes were painted across the aperture from a solution of 1% DPhPC in n-decane with a Teflon loop. Baseline currents were recorded using only electrolyte and detergent-containing buffer to exclude interference from contaminants and membrane-active detergents. The purified gp51 or gp52 proteins were added to both sides of the cuvette, and the reaction was recorded for 15 minutes. Currents were recorded at –10 mV applied potential using a Keithley 428 current amplifier with a rise

time of 30 ms and digitized by a computer equipped with Keithley Metrabyte STA 1800 U interface. The data were recorded with Test Point 4.0 software (Keithley). The data were analyzed in SigmaPlot 11.0 (Systat Software) to generate the graphs shown in this paper. The raw data were analyzed using IGOR Pro 5.03 (WaveMetrics) using a macro provided by Dr. Harald Engelhardt.

DATA AVAILABILITY

The atomic coordinates generated in this study have been deposited in the Protein Data Bank database under accession codes:
9PNR, [https://doi.org/10.2210/pdb9PNR/pdb] (N4 ΔN-vRNAP Tight Complex)
9PNQ, [https://doi.org/10.2210/pdb9PNQ/pdb] (Isolated N4 RNAP)
9PNT, [https://doi.org/10.2210/pdb9PNT/pdb] (N4 ΔN-vRNAP Loose Complex)
9PNW, [https://doi.org/10.2210/pdb9PNW/pdb] (P1 bound ΔN-vRNAP Tight Complex)
9PNV, [https://doi.org/10.2210/pdb9PNV/pdb] (P1 bound isolated RNAP)
9YF4, [https://doi.org/10.2210/pdb9YF4/pdb] (C6 averaged N4 Full Virion Tail)
9YF5, [https://doi.org/10.2210/pdb9YF5/pdb] (C6 averaged N4 Empty Particle Tail)
9YF8, [https://doi.org/10.2210/pdb9YF8/pdb] (C12 averaged Full Virion Portal)
9YF9, [https://doi.org/10.2210/pdb9YF9/pdb] (C12 averaged Empty Particle Portal)
9YFT, [https://doi.org/10.2210/pdb9YFT/pdb] (asymmetric N4 Tail-Gating Complex)
11MN, [https://doi.org/10.2210/pdb11MN/pdb] (ΔC-gp51)

The cryo-EM density maps generated in this study have been deposited in the Electron Microscopy Data Bank under accession codes:
EMD-71769, [https://www.emdataresource.org/EMD-71769] (ΔN-vRNAP Tight Complex)
EMD-71768, [https://www.emdataresource.org/EMD-71768] (Isolated N4 RNAP)
EMD-71771, [https://www.emdataresource.org/EMD-71771] (N4 ΔN-vRNAP Loose Complex)
EMD-71774, [https://www.emdataresource.org/EMD-71774] (P1 bound ΔN-vRNAP Tight Complex)
EMD-71773, [https://www.emdataresource.org/EMD-71773] (P1 bound isolated RNAP)
EMD-72876, [https://www.emdataresource.org/EMD-72876] (C6 N4 Full Virion Tail)
EMD-72877, [https://www.emdataresource.org/EMD-72877] (C6 N4 Empty Particle Tail)
EMD-72880, [https://www.emdataresource.org/EMD-72880] (C12 N4 Full Virion Portal)
EMD-72881, [https://www.emdataresource.org/EMD-72881] (C12 N4 Empty Particle Portal)
EMD-72905, [https://www.emdataresource.org/EMD-72905] (asymmetric N4 Tail-Gating Complex)
EMD-75838, [https://www.emdataresource.org/EMD-75838] (ΔC-gp51)

The minimum datasets (e.g., motion corrected micrographs) necessary to interpret and verify all eleven cryo-EM reconstructions presented in this paper can be downloaded from the University of Alabama at Birmingham (UAB) Cheaha Supercomputer (https://www.uab.edu/it/home/research-computing/cheaha) upon request to the authors. All AlphaFold models used as initial templates for the interpretation and model building of the N4 cryo-EM reconstructions have been deposited in Figshare and are available at

https://doi.org/10.6084/m9.figshare.32592447. The source data underlying Fig. 4 and Supplementary Figs. 1, 4, 8, 11, and 12 are provided as a Source Data file.

REFERENCES

1 Iglesias, S. M., Li, F., Briani, F. & Cingolani, G. Viral Genome Delivery Across Bacterial Cell Surfaces. *Annu Rev Microbiol* **78**, 125-145 (2024). https://doi.org/10.1146/annurev-micro-041222-124727

2 Parent, K. N., Schrad, J. R. & Cingolani, G. Breaking Symmetry in Viral Icosahedral Capsids as Seen through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy. *Viruses* **10** (2018). https://doi.org/v10020067 [pii]
10.3390/v10020067

3 Swanson, N. A., Hou, C. D. & Cingolani, G. Viral Ejection Proteins: Mosaically Conserved, Conformational Gymnasts. *Microorganisms* **10** (2022). https://doi.org/10.3390/microorganisms10030504

4 Chen, W. *et al.* Structural changes in bacteriophage T7 upon receptor-induced genome ejection. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org/10.1073/pnas.2102003118

5 Zheng, J. *et al.* Conformational changes in and translocation of small proteins: insights into the ejection mechanism of podophages. *J Virol* **99**, e0124924 (2025). https://doi.org/10.1128/jvi.01249-24

6 Lupo, D. *et al.* The T7 ejection nanomachine components gp15-gp16 form a spiral ring complex that binds DNA and a lipid membrane. *Virology* **486**, 263-271 (2016). https://doi.org/S0042-6822(15)00413-4 [pii]
10.1016/j.virol.2015.09.022

7 Swanson, N. A. *et al.* Cryo-EM structure of the periplasmic tunnel of T7 DNA-ejectosome at 2.7 A resolution. *Mol Cell* **81**, 3145-3159 e3147 (2021). https://doi.org/10.1016/j.molcel.2021.06.001

8 Perez, G. L., Huynh, B., Slater, M. & Maloy, S. Transport of phage P22 DNA across the cytoplasmic membrane. *J Bacteriol* **191**, 135-140 (2009). https://doi.org/JB.00778-08 [pii]
10.1128/JB.00778-08

9 Kemp, P., Gupta, M. & Molineux, I. J. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Mol Microbiol* **53**, 1251-1265 (2004). https://doi.org/10.1111/j.1365-2958.2004.04204.x
MMI4204 [pii]

10 Struthers-Schlinke, J. S., Robins, W. P., Kemp, P. & Molineux, I. J. The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. *J Mol Biol* **301**, 35-45 (2000). https://doi.org/10.1006/jmbi.2000.3940
S0022-2836(00)93940-8 [pii]

11 Molineux, I. J. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol* **40**, 1-8 (2001). https://doi.org/mmi2357 [pii]
10.1046/j.1365-2958.2001.02357.x

12 Swanson, N. A. *et al.* Expression and purification of phage T7 ejection proteins for cryo-EM analysis. *STAR Protoc* **2**, 100960 (2021). https://doi.org/10.1016/j.xpro.2021.100960

13 Perez-Ruiz, M. *et al.* Assisted assembly of bacteriophage T7 core components for genome translocation across the bacterial envelope. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org/10.1073/pnas.2026719118

14 Hu, B., Margolin, W., Molineux, I. J. & Liu, J. The bacteriophage t7 virion undergoes extensive structural remodeling during infection. *Science* **339**, 576-579 (2013). https://doi.org/science.1231887 [pii]
10.1126/science.1231887
15 Wittmann, J. *et al.* From Orphan Phage to a Proposed New Family-the Diversity of N4-Like Viruses. *Antibiotics (Basel)* **9** (2020). https://doi.org/10.3390/antibiotics9100663
16 Lokareddy, R. K. *et al.* Integrative structural analysis of Pseudomonas phage DEV reveals a genome ejection motor. *Nat Commun* **15**, 8482 (2024). https://doi.org/10.1038/s41467-024-52752-1
17 Lenneman, B. R. & Rothman-Denes, L. B. Structural and biochemical investigation of bacteriophage N4-encoded RNA polymerases. *Biomolecules* **5**, 647-667 (2015). https://doi.org/10.3390/biom5020647
18 Falco, S. C., Zehring, W. & Rothman-Denes, L. B. DNA-dependent RNA polymerase from bacteriophage N4 virions. Purification and characterization. *J Biol Chem* **255**, 4339-4347 (1980).
19 Falco, S. C., Laan, K. V. & Rothman-Denes, L. B. Virion-associated RNA polymerase required for bacteriophage N4 development. *Proc Natl Acad Sci U S A* **74**, 520-523 (1977). https://doi.org/10.1073/pnas.74.2.520
20 Choi, K. H. *et al.* Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol* **378**, 726-736 (2008). https://doi.org/10.1016/j.jmb.2008.02.059
21 Rothman-Denes, L. B. & Schito, G. C. Novel transcribing activities in N4-infected Escherichia coli. *Virology* **60**, 65-72 (1974). https://doi.org/10.1016/0042-6822(74)90366-3
22 Zheng, J. *et al.* Asymmetric structure of podophage N4 from the Schitoviridae family reveals a type of tube-sheath short-tail architecture. *Structure* (2025). https://doi.org/10.1016/j.str.2025.04.008
23 Subramanian, S., Bergland Drarvik, S. M., Tinney, K. R., Doore, S. M. & Parent, K. N. Moo19 and B2: Structures of Schitoviridae podophages with T = 9 geometry and tailspikes with esterase activity. *Sci Adv* **10**, eadt0022 (2024). https://doi.org/10.1126/sciadv.adt0022
24 Murakami, K. S., Davydova, E. K. & Rothman-Denes, L. B. X-ray crystal structure of the polymerase domain of the bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci U S A* **105**, 5046-5051 (2008). https://doi.org/10.1073/pnas.0712325105
25 Kazmierczak, K. M., Davydova, E. K., Mustaev, A. A. & Rothman-Denes, L. B. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J* **21**, 5815-5823 (2002). https://doi.org/10.1093/emboj/cdf584
26 Wu, W. *et al.* Localization of the Houdinisome (Ejection Proteins) inside the Bacteriophage P22 Virion by Bubblegram Imaging. *MBio* **7** (2016). https://doi.org/mBio.01152-16 [pii]
10.1128/mBio.01152-16
27 Cheng, N., Wu, W., Watts, N. R. & Steven, A. C. Exploiting radiation damage to map proteins in nucleoprotein complexes: the internal structure of bacteriophage T7. *J Struct Biol* **185**, 250-256 (2014). https://doi.org/S1047-8477(13)00319-5 [pii]
10.1016/j.jsb.2013.12.004
28 Wu, W., Thomas, J. A., Cheng, N., Black, L. W. & Steven, A. C. Bubblegrams reveal the inner body of bacteriophage phiKZ. *Science* **335**, 182 (2012). https://doi.org/335/6065/182 [pii]

10.1126/science.1214120
29 Morishita, Y. *et al.* Fused radical SAM and alphaKG-HExxH domain proteins contain a distinct structural fold and catalyse cyclophane formation and beta-hydroxylation. *Nat Chem* **16**, 1882-1893 (2024). https://doi.org/10.1038/s41557-024-01596-9
30 Falco, S. C. & Rothman-Denes, L. B. Bacteriophage N4-induced transcribing activities in Escherichia coli. II. Association of the N4 transcriptional apparatus with the cytoplasmic membrane. *Virology* **95**, 466-475 (1979). https://doi.org/10.1016/0042-6822(79)90501-4
31 Cretin, G. *et al.* SWORD2: hierarchical analysis of protein 3D structures. *Nucleic Acids Res* **50**, W732-W738 (2022). https://doi.org/10.1093/nar/gkac370
32 Gleghorn, M. L., Davydova, E. K., Rothman-Denes, L. B. & Murakami, K. S. Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* **32**, 707-717 (2008). https://doi.org/10.1016/j.molcel.2008.11.010
33 Brieba, L. G. & Sousa, R. The T7 RNA polymerase intercalating hairpin is important for promoter opening during initiation but not for RNA displacement or transcription bubble stability during elongation. *Biochemistry* **40**, 3882-3890 (2001). https://doi.org/10.1021/bi002716c
34 Davydova, E. K., Kaganman, I., Kazmierczak, K. M. & Rothman-Denes, L. B. Identification of bacteriophage N4 virion RNA polymerase-nucleic acid interactions in transcription complexes. *J Biol Chem* **284**, 1962-1970 (2009). https://doi.org/10.1074/jbc.M807785200
35 Hou, C. D. *et al.* High-resolution Cryo-EM Analysis of the Therapeutic Pseudomonas Phage Pa223. *J Mol Biol* **437**, 169386 (2025). https://doi.org/10.1016/j.jmb.2025.169386
36 Chang, C. Y., Kemp, P. & Molineux, I. J. Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology* **398**, 176-186 (2010). https://doi.org/10.1016/j.virol.2009.12.002
37 Prilipov, A., Phale, P. S., Van Gelder, P., Rosenbusch, J. P. & Koebnik, R. Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli. *FEMS Microbiol Lett* **163**, 65-72 (1998). https://doi.org/10.1111/j.1574-6968.1998.tb13027.x
38 Niederweis, M. *et al.* Cloning of the *mspA* gene encoding a porin from *Mycobacterium smegmatis*. *Mol. Microbiol.* **33**, 933-945 (1999).
39 Manich, M. *et al.* Clostridium perfringens delta toxin is sequence related to beta toxin, NetB, and Staphylococcus pore-forming toxins, but shows functional differences. *PLoS One* **3**, e3764 (2008). https://doi.org/10.1371/journal.pone.0003764
40 McPartland, J. & Rothman-Denes, L. B. The tail sheath of bacteriophage N4 interacts with the Escherichia coli receptor. *J Bacteriol* **191**, 525-532 (2009). https://doi.org/10.1128/JB.01423-08
41 Zheng, K. *et al.* Identification of hidden N4-like viruses and their interactions with hosts. *mSystems* **8**, e0019723 (2023). https://doi.org/10.1128/msystems.00197-23
42 Drobysheva, A. V. *et al.* Structure and function of virion RNA polymerase of a crAss-like phage. *Nature* **589**, 306-309 (2021). https://doi.org/10.1038/s41586-020-2921-5
43 Bayfield, O. W. *et al.* Structural atlas of a human gut crassvirus. *Nature* **617**, 409-416 (2023). https://doi.org/10.1038/s41586-023-06019-2
44 Wang, C., Tu, J., Liu, J. & Molineux, I. J. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nat Microbiol* **4**, 1049-1056 (2019). https://doi.org/10.1038/s41564-019-0403-z

10.1038/s41564-019-0403-z [pii]
45 Zhao, H. *et al.* Structure of a Bacterial Virus DNA-Injection Protein Complex Reveals a Decameric Assembly with a Constricted Molecular Channel. *PLoS One* **11**, e0149337 (2016). https://doi.org/10.1371/journal.pone.0149337
46 David Hou, C. F. *et al.* Cryo-EM Structure of a Kinetically Trapped Dodecameric Portal Protein from the Pseudomonas-phage PaP3. *J Mol Biol* **434**, 167537 (2022). https://doi.org/10.1016/j.jmb.2022.167537
47 Dedeo, C. L., Cingolani, G. & Teschke, C. M. Portal Protein: The Orchestrator of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses. *Annu Rev Virol* **6**, 141-160 (2019). https://doi.org/10.1146/annurev-virology-092818-015819
48 Lokareddy, R. K., Hou, C. D., Li, F., Yang, R. & Cingolani, G. Viral Small Terminase: A Divergent Structural Framework for a Conserved Biological Function. *Viruses* **14** (2022). https://doi.org/10.3390/v14102215
49 Junkermeier, E. H. & Hengge, R. A Novel Locally c-di-GMP-Controlled Exopolysaccharide Synthase Required for Bacteriophage N4 Infection of Escherichia coli. *mBio* **12**, e0324921 (2021). https://doi.org/10.1128/mbio.03249-21
50 Sellner, B. *et al.* A New Sugar for an Old Phage: a c-di-GMP-Dependent Polysaccharide Pathway Sensitizes Escherichia coli for Bacteriophage Infection. *mBio* **12**, e0324621 (2021). https://doi.org/10.1128/mbio.03246-21
51 d'Acapito, A., Decombe, A., Arnaud, C. A. & Breyton, C. Comparative anatomy of siphophage tails before and after interaction with their receptor. *Curr Opin Struct Biol* **92**, 103045 (2025). https://doi.org/10.1016/j.sbi.2025.103045
52 Nieto Noblecia, J. *et al.* Pseudomonas aeruginosa DEV phage exploits the essential LptD outer membrane protein as receptor for adsorption. *mBio*, e0356125 (2026). https://doi.org/10.1128/mbio.03561-25
53 Kiino, D. R. & Rothman-Denes, L. B. Genetic analysis of bacteriophage N4 adsorption. *J Bacteriol* **171**, 4595-4602 (1989). https://doi.org/10.1128/jb.171.9.4595-4602.1989
54 Grayson, P. & Molineux, I. J. Is phage DNA 'injected' into cells--biologists and physicists can agree. *Curr Opin Microbiol* **10**, 401-409 (2007). https://doi.org/10.1016/j.mib.2007.04.004
55 Molineux, I. J. & Panja, D. Popping the cork: mechanisms of phage genome ejection. *Nat Rev Microbiol* **11**, 194-204 (2013). https://doi.org/10.1038/nrmicro2988
56 Schito, G. C. Dvelopment of coliphage N4: ultrastructural studies. *J Virol* **13**, 186-196 (1974). https://doi.org/10.1128/JVI.13.1.186-196.1974
57 Rothman-Denes, L. B. in *The Bacteriophages* (ed Richard Calendar) 200-214 (Oxford University Press, 2006).
58 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017). https://doi.org/10.1038/nmeth.4169
59 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018). https://doi.org/10.1002/pro.3235
60 Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr D Struct Biol* **74**, 545-559 (2018). https://doi.org/10.1107/S2059798318004655
61 Yu, I. *et al.* Building atomic models based on near atomic resolution cryoEM maps with existing tools. *J Struct Biol* **204**, 313-318 (2018). https://doi.org/10.1016/j.jsb.2018.08.004

62 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018). https://doi.org/10.1107/S2059798318002425
63 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024). https://doi.org/10.1038/s41586-024-07487-w
64 Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024). https://doi.org/10.1038/s41586-024-07215-4
65 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018). https://doi.org/10.1107/S2059798318006551
66 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010). https://doi.org/S0907444909042073 [pii]
10.1107/S0907444909042073
67 DeLano, W. L. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC. (2002).
68 van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243-246 (2024). https://doi.org/10.1038/s41587-023-01773-0
69 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007). https://doi.org/10.1016/j.jmb.2007.05.022
70 Laskowski, R. A. PDBsum new things. *Nucleic Acids Res* **37**, D355-359 (2009). https://doi.org/10.1093/nar/gkn860
71 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
72 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2
73 Skolnick, J., Gao, M., Zhou, H. & Singh, S. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *J Chem Inf Model* **61**, 4827-4831 (2021). https://doi.org/10.1021/acs.jcim.1c01114
74 Yin, X., Yang, J., Xiao, F., Yang, Y. & Shen, H. B. MemBrain: An Easy-to-Use Online Webserver for Transmembrane Protein Structure Prediction. *Nanomicro Lett* **10**, 2 (2018). https://doi.org/10.1007/s40820-017-0156-2
75 Heinz, C. & Niederweis, M. Selective extraction and purification of a mycobacterial outer membrane protein. *Anal Biochem* **285**, 113-120 (2000). https://doi.org/10.1006/abio.2000.4728

ACKNOWLEDGMENTS

We thank Dr. Regine Hengge for providing the N4 phage. We also thank Dr. Peter Prevelige for assistance with the mass spectrometry identification of gp51. Electron microscopy was carried out at the UAB Cryo-EM Facility (RRID:SCR_025450), supported by the Institutional Research Core Program and the O'Neal Comprehensive Cancer Center (NIH grant P30 CA013148), with additional support from NIH grant S10 OD024978. We thank the staff at NCEF and the Stanford-SLAC CryoEM Center (S2C2) for assistance with cryo-EM data collection. NCEF is supported by contract 75N91019D00024, and S2C2 is supported by the NIH Common Fund Transformative

High-Resolution Cryo-Electron Microscopy program (U24 GM129541). Part of this work was carried out at the UNITECH OMICs mass spectrometry platform of the Università degli Studi di Milano.

FUNDING STATEMENT
This work was supported by the National Institutes of Health grants R01 AI191107 and R35 GM140733 to G.C.; R35 GM140710 to D.A.S.; and R01 AI184596 and R01 AI175106 to M.N. Additional support was provided by the Fondazione per la Ricerca sulla Fibrosi Cistica–Delegazioni di Milano, Sondrio Valchiavenna, and Saviano (grant FFC#16/2023 to F.B.).

AUTHOR CONTRIBUTIONS STATEMENT
N.B., R.K.L, and G.C. performed biochemical analysis, cryo-EM, structural analysis, deposition of atomic coordinates and maps, and figure preparation. J.L.K. assisted with vitrification, cryo-EM screening, and data collection. F.F. and F.B. purified the N4 virion and analyzed virion proteins by SDS-PAGE. M.P. and M.N. carried out all steps of lipid bilayer experiments, including data collection and interpretation. S.L.H., R.J., and D.A.S. performed all transcription assays, including data collection and interpretation. G.C. supervised the entire project. N.B. and G.C. wrote the paper. All authors contributed to the writing and editing of the manuscript.

COMPETING INTERESTS STATEMENT
The authors declare no competing interests.

Table 1. Map and model refinement statistics for vRNAP

Data Collection Statistics						
Specimen	N4 vRNAP - Apo			N4 vRNAP w / P1 Promoter		ΔC-gp51
Facility/Microscope	S2C2 /Titan Krios					UAB / Glacios 2
Detector	Falcon 4					Falcon 4i
Imaging Software	EPU					EPU
Magnification	120,000x					190,000x
Voltage (kV)	300 kV					200 kV
Exposure ($e^-/Å^2$)	50 e-/$Å^2$					60 e/$Å^2$
Exposure Time (sec)	6.99 sec					3.81 sec
Defocus range/step (μm)	-0.5 μm to -2.0 μm					-0.75 μm to -2.0 μm
Pixel size (Å/px)	1.217 Å / px					0.7159 Å / px
Total movies	13,007			5,323		9,068
Frames/movie	50					26
Refinement Statistics						
Entry	ΔN-vRNAP Tight Complex (1008–3500)	Isolated RNAP (1008–2101)	ΔN-vRNAP Loose Complex (1008–3500)	P1:ΔN-vRNAP Tight Complex (1008–3500)	P1:Isolated RNAP (1008–2101)	PT-fragment ΔC-gp51
PDB / EMDB entry	9PNR / EMD-71769	9PNQ / EMD-71768	9PNT / EMD-71771	9PNW / EMD-71774	9PNV / EMD-71773	11MN / EMD-75838
Initial particle number	238,654	531,502	326,271	206,327	239,323	
Final particle number	221,813	305,193	18,762	128,569	133,659	77,357
Map Resolution (Å) at FSC 0.143	2.8	2.6	4.3	2.9	2.8	214,562
Map Symmetry	C1	C1	C1	C1	C1	C10
Initial Model	4FF3 AlphaFold 3 / de novo	3C2P AlphaFold 3 / de novo	9PNR	4FF3 AlphaFold 3 / de novo	3C2P	AlphaFold 3
Chains (Residues/DNA)	1 (2443)	1 (1095)	1 (2443)	2 (2443/20)	2 (1095/20)	10 (2040)
Map-to-Model Correlation Coefficient (CC)	0.84	0.90	0.53	0.86	0.89	0.85
MolProbity / Clash	1.17 / 3.60	0.99 / 2.19	1.18 / 3.02	1.54 / 4.83	1.24 / 2.71	1.42 / 6.21
R.M.S. deviations						
Bond Length (Å) / Angles (°)	0.0 / 0.4	0.002 / 0.4	0.003 / 0.9	0.002 / 0.4	0.002 / 0.4	0.003 / 0.8
Rotamer outliers (%)	1.3	0.8	1.3	2.4	1.8	1.3
Ramachandran (%) Fav / Allow / Outlier	98.1 / 1.9 / 0.00	99.1 / 0.9 / 0.0	98.6 / 1.4 / 0.0	98.2 / 1.8 / 0.0	98.9 / 1.1 / 0.0	99.5 / 0.5 / 0.0

Table 2. Map and model refinement statistics for N4 virion

Data Collection Statistics					
Specimen	N4 Virion				
Facility / Microscope	UAB / Glacios 2				
Detector	Falcon 4i				
Data Collection Software	EPU				
Magnification	120,000 x				
Voltage (kV)	200 kV				
Exposure ($e^-/Å^2$)	40 $e^-/Å^2$				
Exposure Time (sec)	7.05 sec				
Defocus range / step (µm)	0.75-2.0 (0.25) µm				
Pixel size (Å/px)	1.19 Å/px				
Total movies	24,989				
Frames / movie	40				
Refinement Statistics					
Entry	C6 Tail Full	C6 Tail Empty	C12 Portal Full	C12 Portal Empty	Tail-Gating Complex
Modeled Proteins	gp54/gp59/gp64/ gp65/gp66/gp67	gp54/gp59/gp64/ gp65/gp66/gp67	gp59	gp59	gp54/gp53/gp65
PDB / EMDB entry	9YF4 / EMD-72876	9YF5 / EMD-72877	9YF8 / EMD-72880	9YF9 / EMD-72881	9YFT / EMD-72905
Initial particle number	30,374	19,338	30,374	19,338	49,712
Final particle number	16,320	10,265	16,320	10,265	18,434
Map Resolution (Å) at $FSC_{0.143}$	3.28	3.65	2.99	3.24	3.61
Map Symmetry	C6	C6	C12	C12	C1
Initial Model	AlphaFold 3 / de novo	AlphaFold 3 / de novo	AlphaFold 3 / de novo	AlphaFold 3 / de novo	AlphaFold 3 / de novo
Chains (Residues)	16 (ASU)	16 (ASU)	12	12	15
Map-to-Model Correlation Coefficient (CC)	0.85	0.85	0.88	0.88	0.68
MolProbity / ClashScore	1.48 / 7.29	1.57 / 7.12	1.49 / 5.71	1.18 / 3.93	1.84 / 6.70
R.M.S. deviations					
Bond Length (Å) /Angles (°)	0.005 / 0.9	0.002 / 0.5	0.005 / 0.9	0.004 / 0.9	0.002 / 0.4
Rotamer outliers (%)	1	1.7	1.7	0.9	1.9
Ramachandran (%) Fav / Allow / Outlier	97.6 / 2.3 / 0.0	98.1 / 1.9 / 0.0	98.5 / 1.5 / 0.0	98.6 / 1.4 / 0.0	96.1 / 3.9 / 0.0

FIGURE LEGENDS

Figure 1 | Bubblegram analysis of bacteriophage N4 virions.
(**a**) Representative dose series of 5 full virions exposed to a total dose ranging from 50 e-/Å to 250 e-/Å out of a total of 39 full virions. Bubbles emerging from inside the capsid appear after an accumulated dose of approximately 175 e-/Å. The last row shows empty particles lacking bubbles inside the virion due to the loss of ejection proteins and DNA.
(**b**) Quantification of all 37 virions with visible tails. The histogram shows the distance of bubbles trapped inside the capsid from the unique vertex occupied by the portal protein and tail.

Figure 2 | Cryo-EM reconstruction of N4 vRNAP.
(**a**) Linear schematics of vRNAP domain organization.
(**b**) Experimental density of ΔN-vRNAP tight complex colored by domain: RNAP (red), CTD-C1 (green), and CTD-C2 (blue).
(**c**) Ribbon diagram of ΔN-vRNAP tight complex colored as in panel (**b**).
(**d**) Ribbon diagram of ΔN-vRNAP loose complex colored as in panel (**b**).

Figure 3 | Plastic quaternary organization of vRNAP domains.
(**a**) Map of binding interfaces and interactions that stabilize vRNAP tight conformation. Yellow indicates salt bridges with the corresponding residue linked by a dashed line. Other colored residues indicate hydrogen-bonding residues interacting with subunits of the corresponding color. RNAP is colored red, CTD-C1 green, and CTD-C2 blue.
(**b**) General schematics of all putative domain organizations of vRNAP observed on grid with supporting 2D classes. RNAP is colored red, CTD-C1 green, CTD-C2 blue, and AlphaFold3-predicted NTD is colored orange.
(**c**) Putative model of beads-on-a-string pre-ejection conformation. NTD subdomains were predicted by AlphaFold 3; RNAP and CTD subdomains were predicted by the protein peeling structural decomposition algorithm SWORD2. Domains are colored as in panel (b).

Figure 4 | CTD allosterically regulates vRNAP transcriptional activity.
(**a**) ΔN-vRNAP tight complex atomic model with RNAP colored by subdomains. RNAP subdomains are colored as follows: N-terminal extension (gray), intercalating β-hairpin (red), N4 plug module (orange), B-motif loop (yellow), specificity loop (cyan), thumb (green), fingers (blue), palm (pink), palm insertion (magenta), and extended foot (purple).
(**b**) Surface representation of the ΔN-vRNAP tight complex bound to the P1 promoter DNA hairpin. P1 DNA is colored purple.
(**c**) Conformational changes in RNAP subdomains due to P1 and CTD binding. Structural elements are colored as in panel (a), with P1 DNA shown in purple and active-site magnesium shown as a gray sphere.
(**d**) Radiolabeled transcription runoff assay of full-length vRNAP (FL-vRNAP; res. 1–3500), recombinant mini-vRNAP (res. 998–2103), and isolated CTD (res. 2124–3500) (lanes 1–3). Lanes 4–9 show CTD titration against RNAP at molar ratios (RNAP:CTD) from 1:0.5 to 1:16.

(**e**) Quantification of the runoff assay in (**d**) from three independent replicates (n=3) shows a progressive decrease in product formation with increasing CTD concentration. (**f**) Radiolabeled transcription runoff assay of FL-vRNAP (FL) and mini-vRNAP (RNAP) at final concentrations of 100 nM, 1 μM, and 100 μM. Representative cropped gels are shown (top), with corresponding densitometric quantification from three independent replicates (bottom). Error bars in panels 4e-f represent the standard deviation (SD), and data are presented as mean ± SD. Source data are provided as a Source Data file.

Figure 5 | Asymmetric cryo-EM reconstructions of N4 full virion and empty particle identify ejection proteins gp51 and gp52.
(**a**) Linear schematic of genomic organization of the N4 ejection cassette (gp50:gp51:gp52) with adjacent ejection-related proteins (gp53:gp54).
(**b**) Cross-section of N4 FV density with main differences from empty particles boxed in red. Proteins are colored as follows: gp59 (maroon), gp67 (pink), gp66 (orange), gp52 (purple), gp54 (yellow), gp64 (green), gp65 (light green), and gp53 (tan).
(**c**) Cross-section of N4 EP density with main differences from full virions boxed in red. The color coding is the same as in panel (**b**), except for the portal, which is colored teal.
(**d**) The FV barrel lumen is filled with density consistent with dsDNA.
(**e**) Isolated density of α-helical hairpin surrounding the portal is assigned to gp51.
(**f**) Isolated density of gp52 fragment (res. 1–16) localized to the tail lumen.

Figure 6 | Channel activity of N4 ejection proteins gp51 and gp52.
(**a**) Cryo-EM reconstruction of ΔC-gp51. Experimental density (left) and corresponding ribbon model (right) are shown in side and top views. The inner diameter ($Ø_i$) is approximately 22 Å.
(**b**) Protomer structures of gp52 and gp51 predicted by AlphaFold 3 (AF; gray). The experimental structure of the ΔC-gp51 protomer (light blue) is overlaid on the AF model.
(**c**) AlphaFold 3 model of the decameric gp52:gp51 assembly of phage N4 (left) compared with the nonameric gp73:gp72 assembly of phage DEV (PDB: 8VXQ). Gp52 (purple) and gp73 (gold) form the putative OMC, whereas gp51 (blue) and gp72 (cyan) form the predicted PT.
(**d, e**) Lipid bilayer recordings of recombinant gp52 (**d**) and gp51 (**e**). Experiments were performed using diphytanoyl phosphatidylcholine (DPhPC) membranes in 10 mM HEPES, pH 7.4, and 1 M KCl at an applied potential of −10 mV. Proteins were added to both sides of the chamber. Recordings were collected for up to 15 min or until membrane rupture. Channel activity was observed at 8 ng mL^{-1} for gp52 and 320 ng mL^{-1} for gp51. In panels (**d, e**), blue arrows indicate channel opening events and red arrows indicate channel closure events. Abbreviations used in the figure: OMC, outer membrane complex; PT, periplasmic tunnel.

Figure 7 | Structure of the N4 asymmetric tail-gating complex.
(**a**) Ribbon diagram of the asymmetric N4 tail. All sheath heterodimers (gp64–gp65) are shown transparent and outlined, except for the gp53-interacting sheath heterodimer, which forms the tail-gating complex. Proteins are colored as follows: gp59 (maroon), gp67 (pink), gp66 (orange), gp52 (purple), gp54 (yellow), gp64 (green), gp65 (light green), and gp53 (tan).

(**b**) Atomic model of asymmetric tail-gating complex.
(**c**) Cross-section of experimental density colored by component showing unique internal features. Density for sheaths and front tail-tube protomers has been removed. The tail tube is yellow, and the hydrophobic knot tan.
(**d**) Densities showing the different conformations of the N4 sheath. Detached conformation shown with an additional viewpoint to show weak density of the sheath (boxed in red).

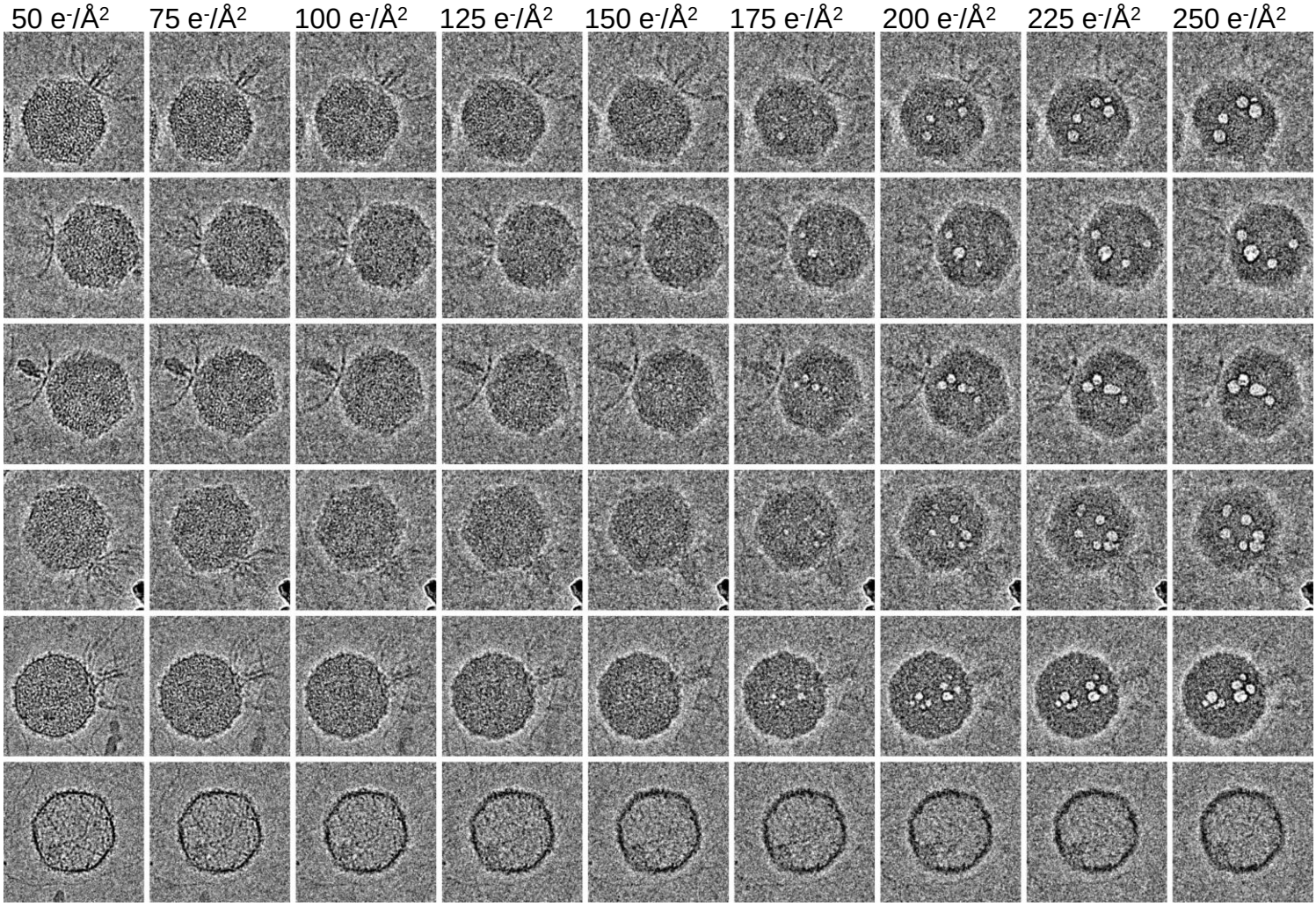
Figure 8 | Composite model of N4 ejection protein-facilitated genome delivery.
(**a**) Diagram of the infectious N4 virion with ejection proteins gp50, gp51, and gp52 poised for expulsion. Gp50/vRNAP is colored (from N- to C-termini) in orange, red, green, and blue; gp51 light blue, gp52 purple, and gp53 tan. In pre-ejection conformation, gp52 localizes inside the tail, gp51 surrounds the portal, gp50-NTD associates with gp51, and gp50 RNAP-CTD folds as beads-on-a-string inside the capsid.
(**b**) Schematic diagram of phage N4 receptors at the E. coli surface: the N4 glycan receptor (NGR), an exopolysaccharide exported by the proteinaceous receptor NfrA.
(**c**) N4 appendages (gp66) reversibly bind NGR.
(**d**) Irreversible binding of the RBP gp65 with the receptor NfrA, which displaces the plug (gp53), letting several luminal copies of gp52 implant into the membrane as the OMC. (**e**) The ejection cascade continues, with gp51 being expelled through the tail and OMC to form a tunnel through the periplasm.
(**f**) Gp50 follows in ejection through the interaction between gp51 and gp50-NTD. The inner membrane complex is likely formed by both the gp51 C-terminus and the gp50-NTD. gp50-RNAP and CTD refold within the bacterial cytoplasm but stay anchored to the bacterial membrane by the linker.
(**g**) Once transcription of early promoters is complete, RNAP activity is dampened via CTD autoinhibition. Abbreviations used in the figure: OMC, outer membrane complex; IMC, inner membrane complex; RBP, receptor-binding protein; NGR, N4 glycan receptor.

Editorial Summary:

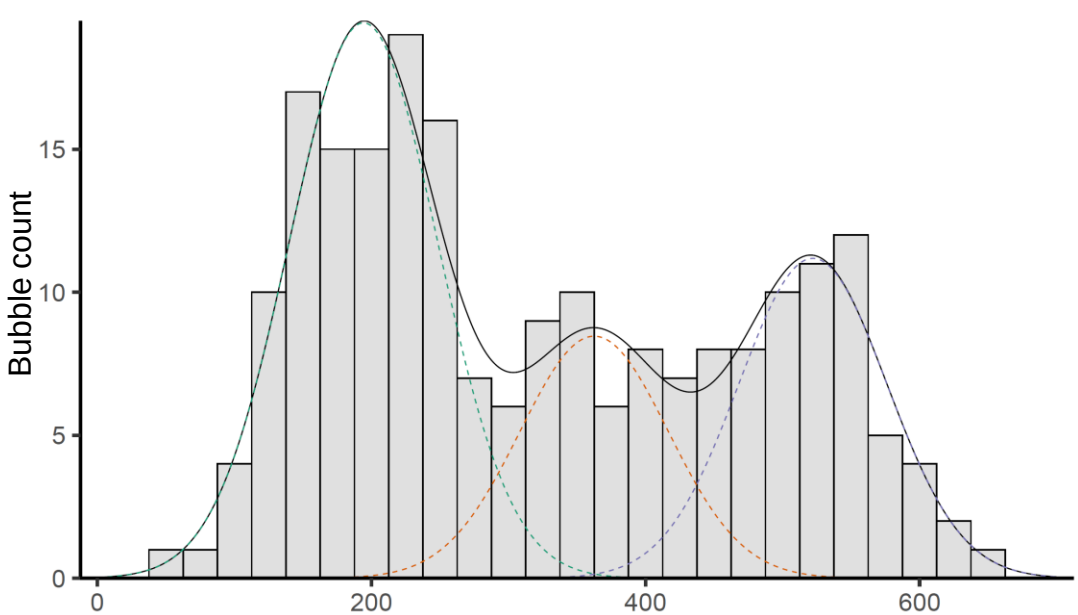
Schitoviruses inject a giant RNA polymerase into bacteria to initiate infection. Here, Bellis et al. reveal the architecture of the coliphage N4 vRNAP and its crosstalk with phage ejection proteins that form the DNA-ejectosome for viral genome delivery.

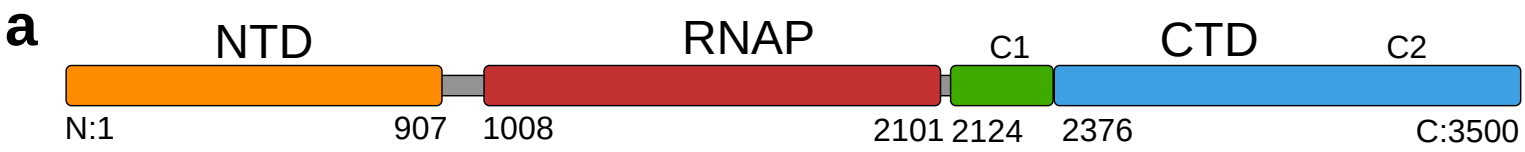
Peer review information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

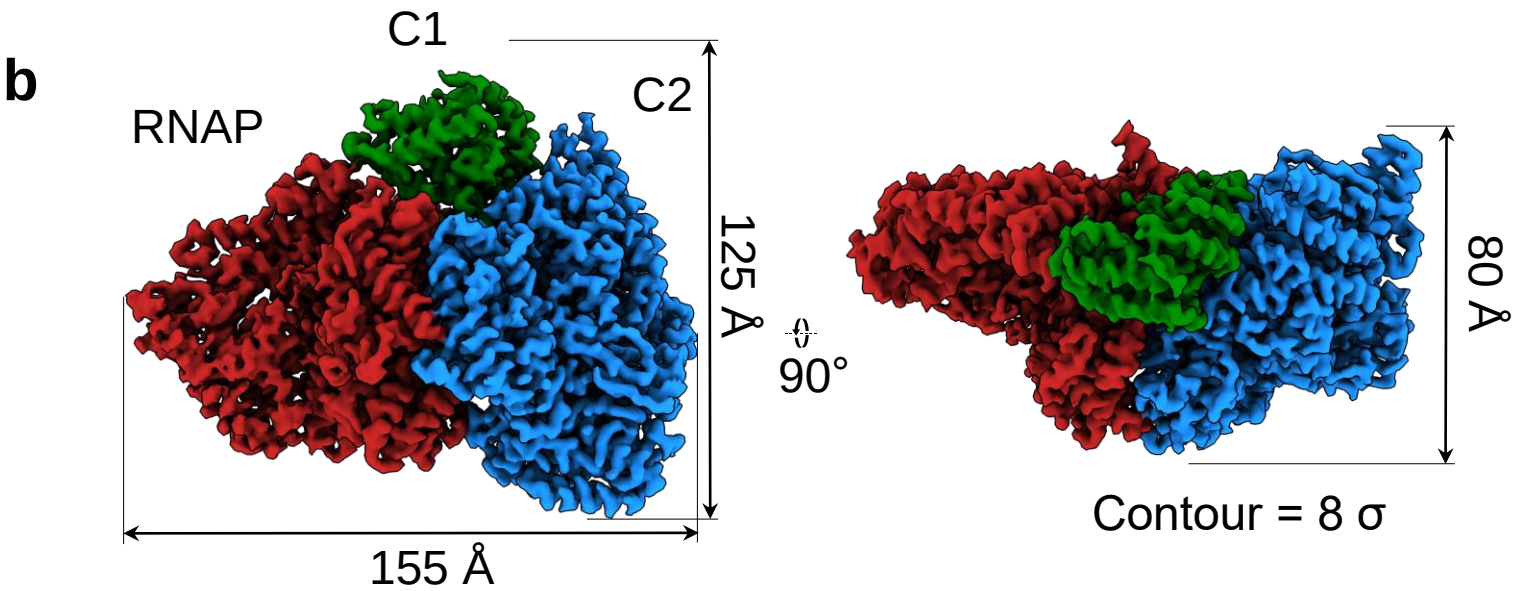
a


Accumulated Dose:
50 e-/Å2
75 e-/Å2
100 e-/Å2
125 e-/Å2
150 e-/Å2
175 e-/Å2
200 e-/Å2
225 e-/Å2
250 e-/Å2

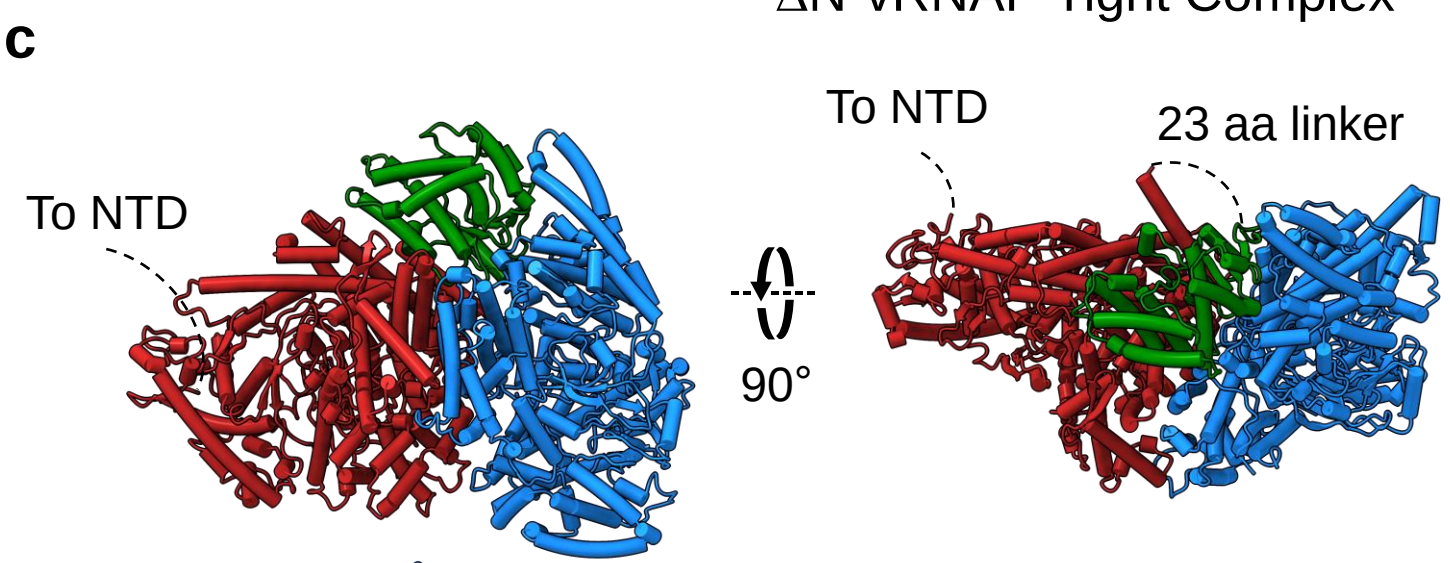
b

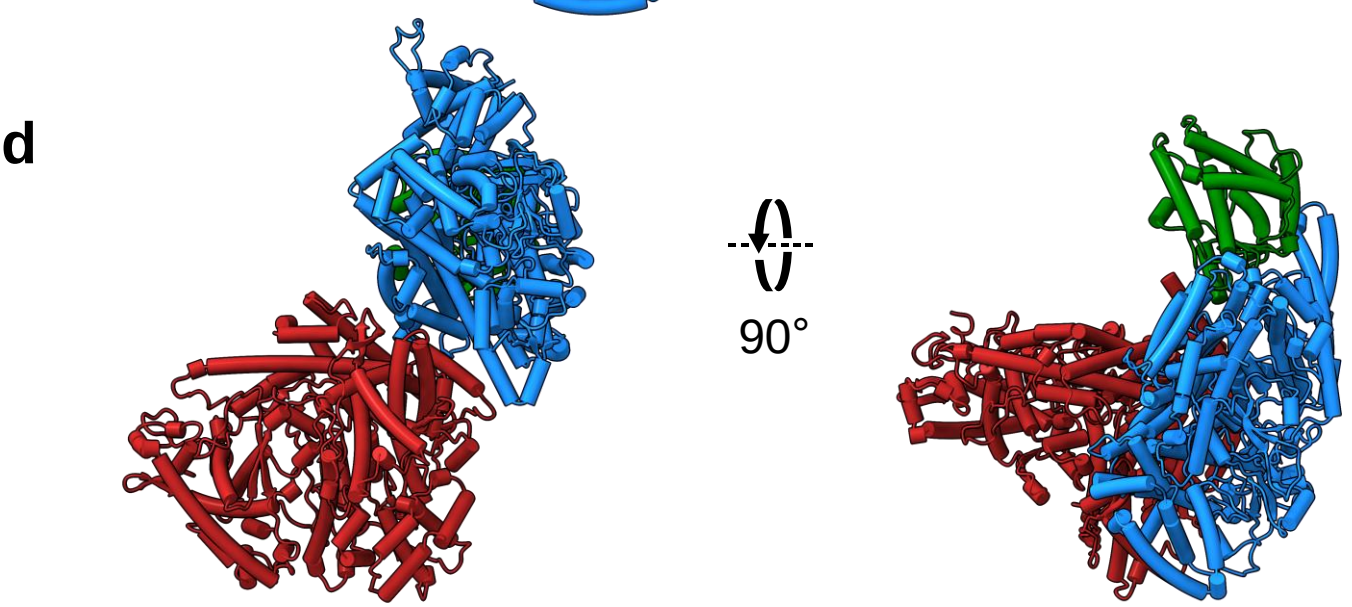

Bubble count
0
5
10
15
0
200
400
600
Distance from tail/ capsid interface to bubble center (2D)
Average 5.7 bubbles per virion/ 37 virions/ 212 bubbles



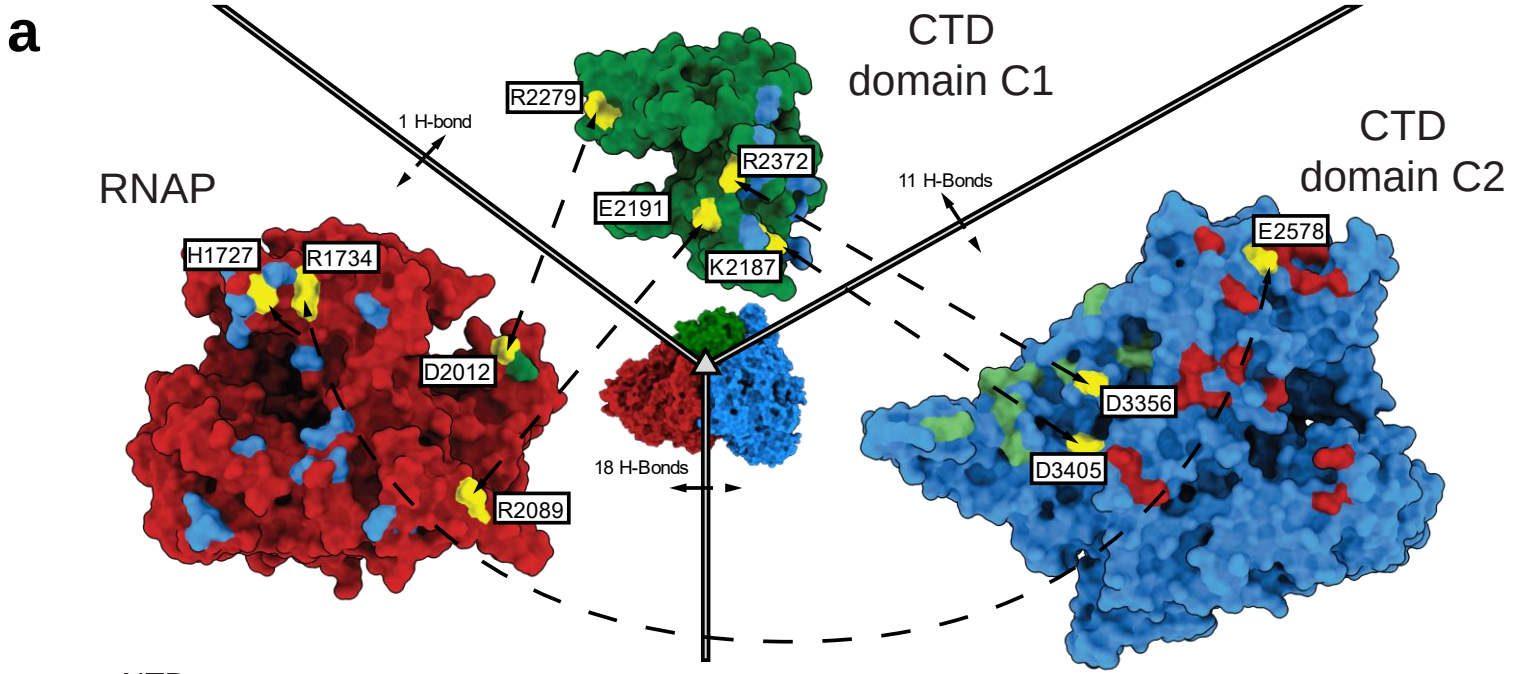


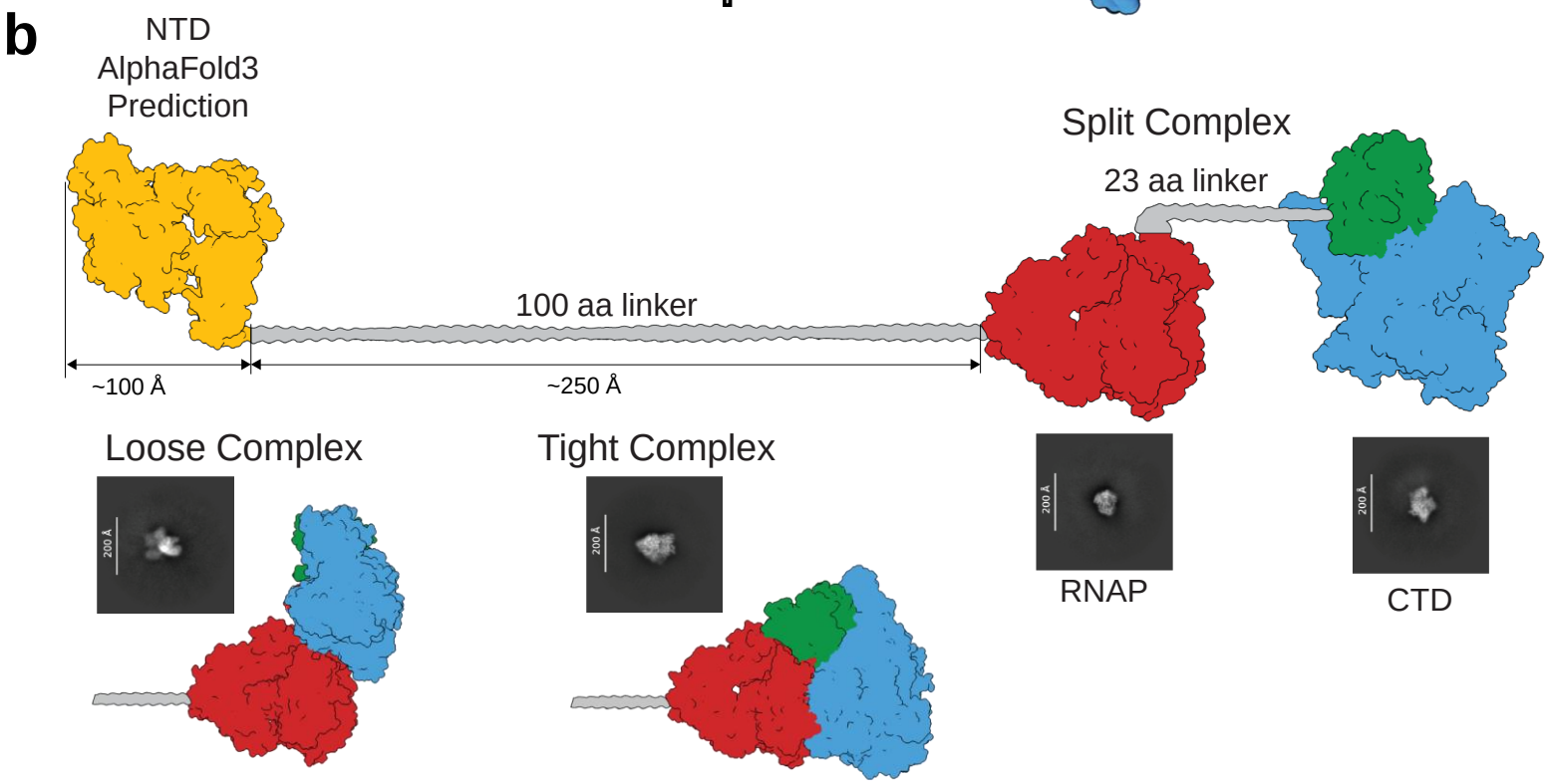
ΔN-vRNAP Tight Complex

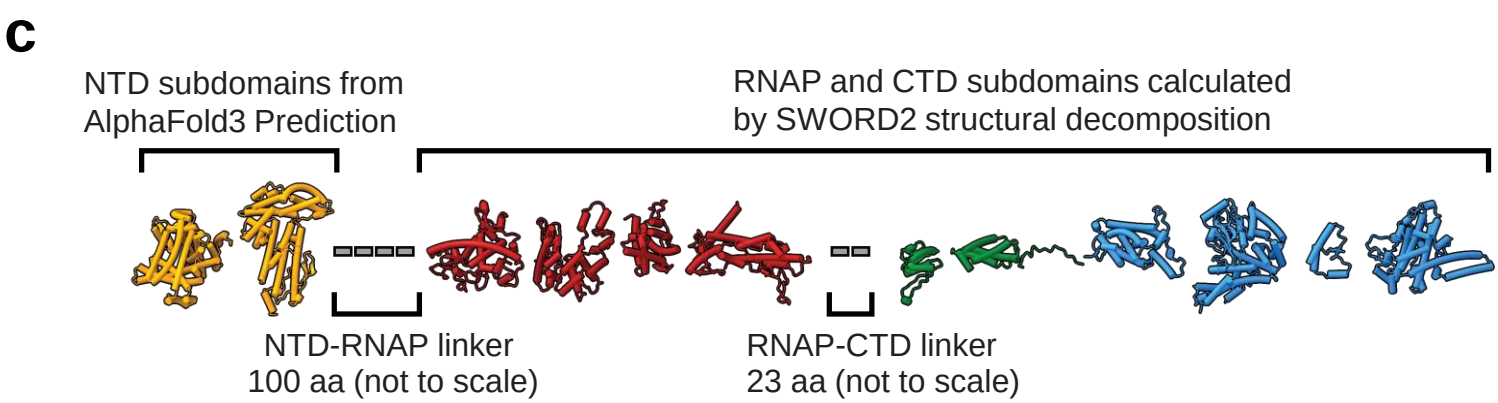


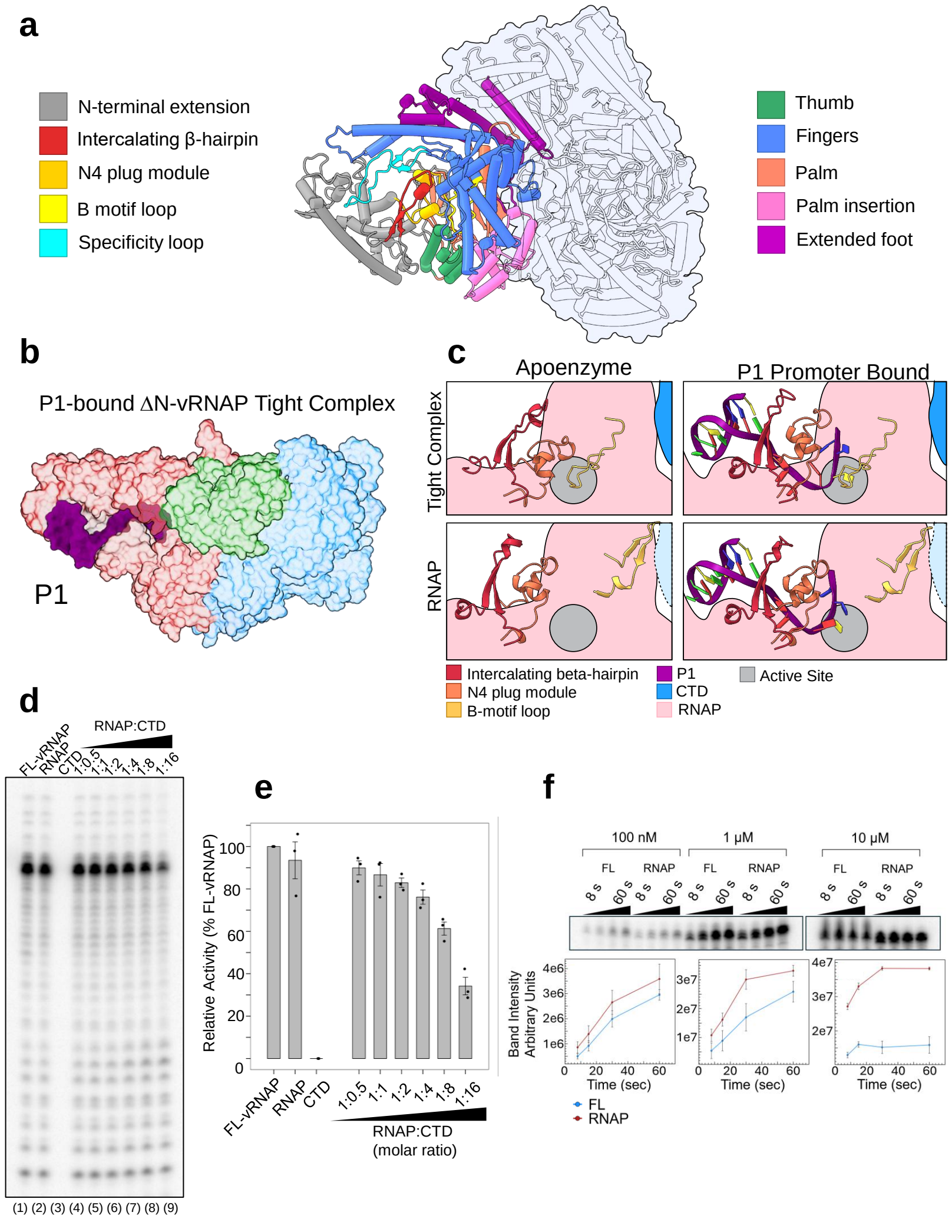


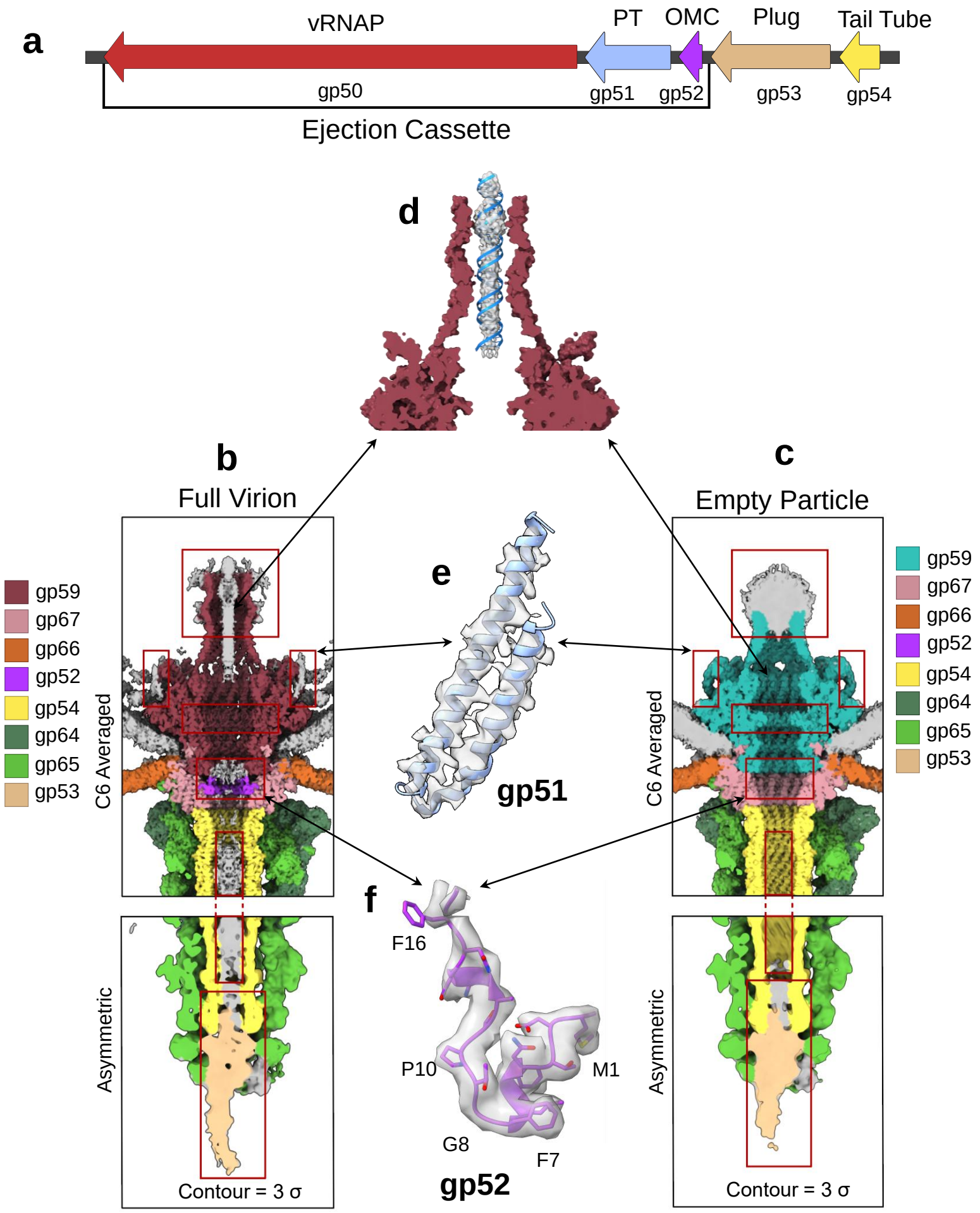
ΔN-vRNAP Loose Complex

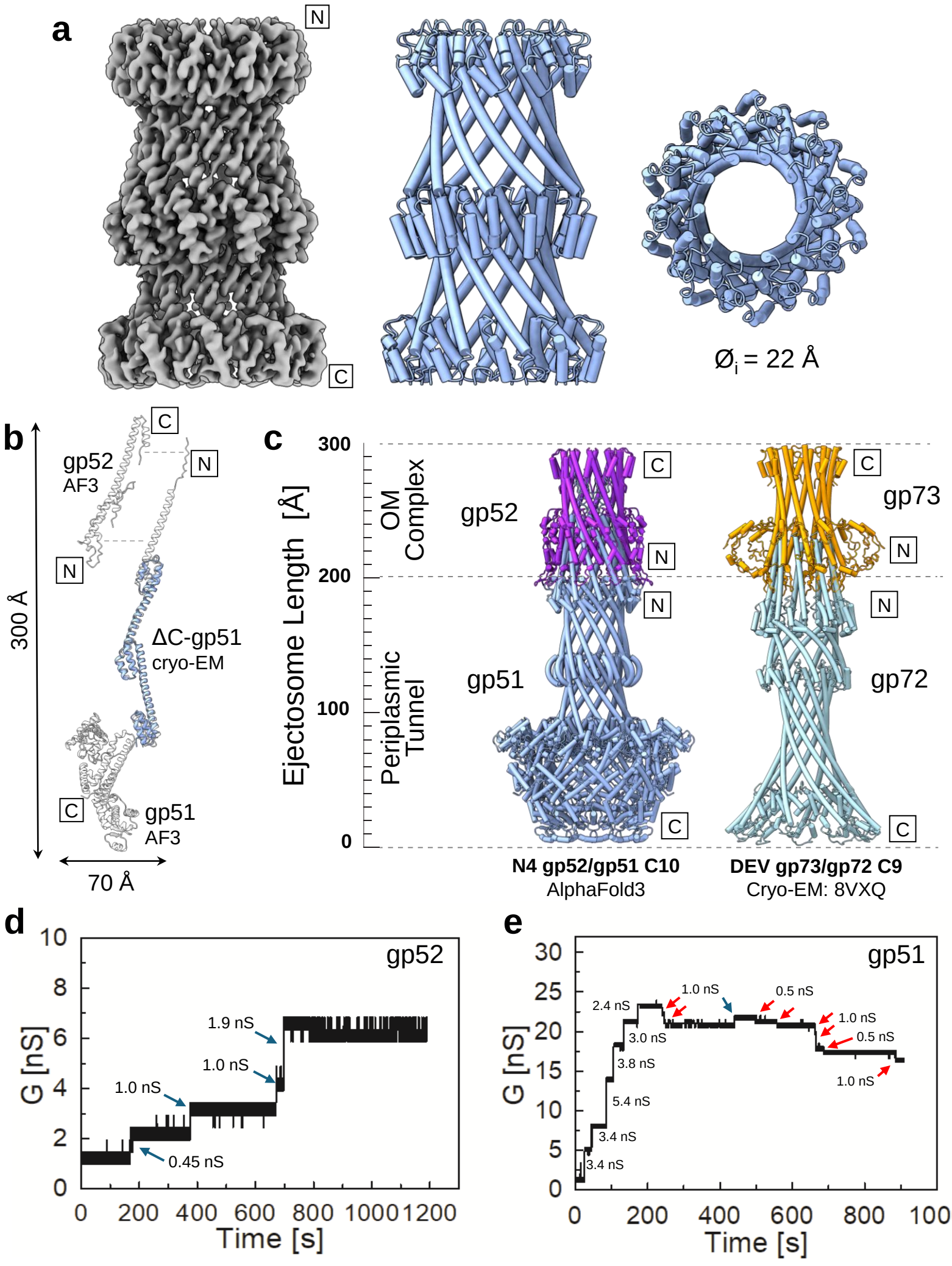

a
CTD
domain C1
CTD
domain C2
RNAP
1 H-bond
11 H-Bonds
18 H-Bonds
R2279
R2372
E2191
K2187
H1727
R1734
D2012
R2089
E2578
D3356
D3405

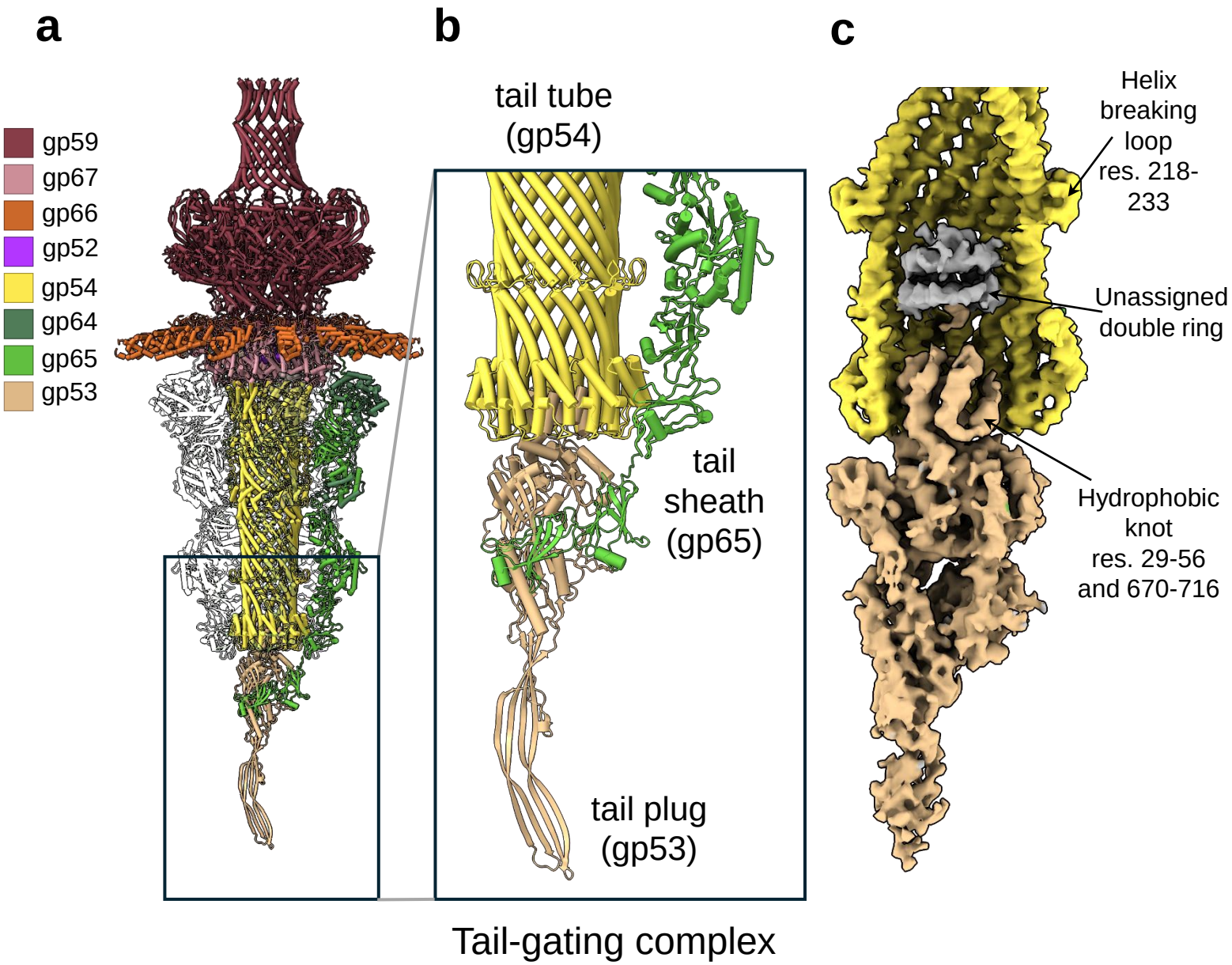

b
NTD
AlphaFold3
Prediction
Split Complex
23 aa linker
100 aa linker
~100 Å
~250 Å
Loose Complex
Tight Complex
RNAP
CTD


c
NTD subdomains from
AlphaFold3 Prediction
RNAP and CTD subdomains calculated
by SWORD2 structural decomposition
NTD-RNAP linker
100 aa (not to scale)
RNAP-CTD linker
23 aa (not to scale)

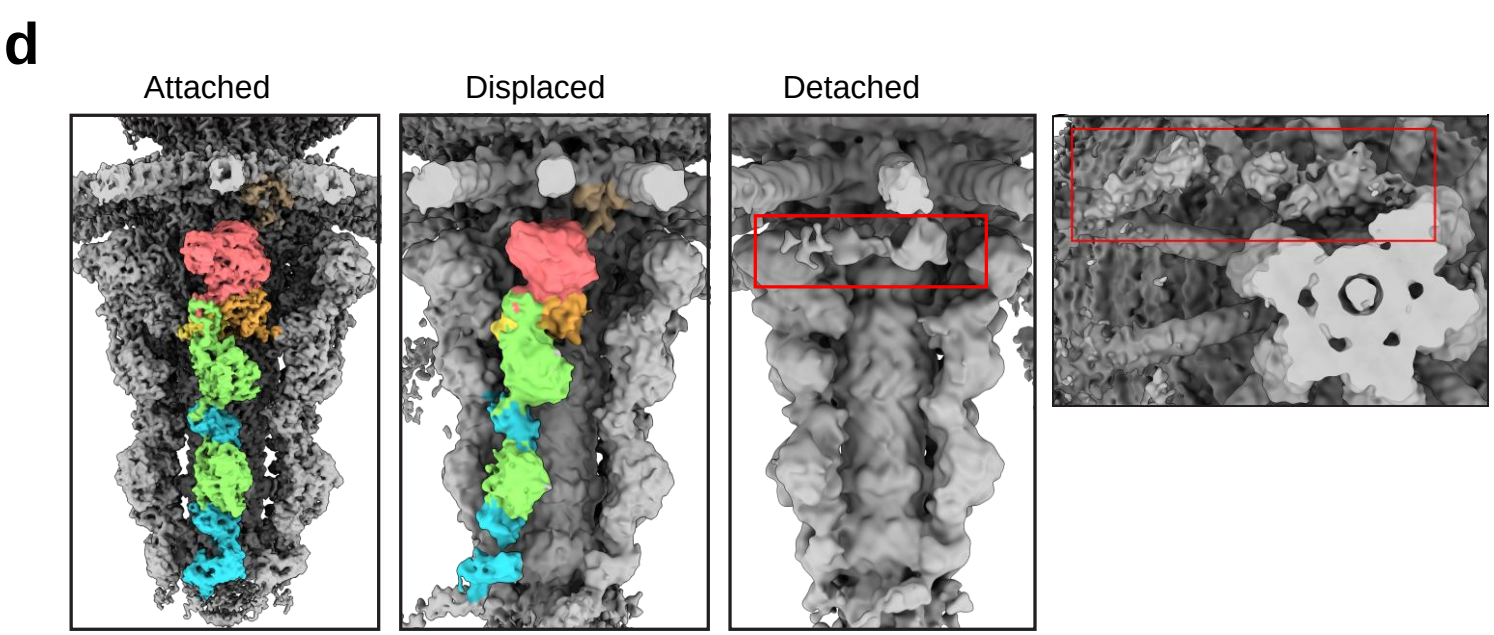

a
N-terminal extension
Intercalating β-hairpin
N4 plug module
B motif loop
Specificity loop
Thumb
Fingers
Palm
Palm insertion
Extended foot
b
P1-bound ΔN-vRNAP Tight Complex
P1
c
Apoenzyme
P1 Promoter Bound
Tight Complex
RNAP
Intercalating beta-hairpin
N4 plug module
B-motif loop
P1
CTD
RNAP
Active Site
d
FL-vRNAP
RNAP
CTD
RNAP:CTD
1:0.5
1:1
1:2
1:4
1:8
1:16
(1) (2) (3) (4) (5) (6) (7) (8) (9)
e
Relative Activity (% FL-vRNAP)
100
80
60
40
20
0
FL-vRNAP
RNAP
CTD
1:0.5
1:1
1:2
1:4
1:8
1:16
RNAP:CTD
(molar ratio)
f
100 nM
1 μM
10 μM
FL
RNAP
8 s
60 s
Band Intensity
Arbitrary Units
4e6
3e6
2e6
1e6
3e7
2e7
1e7
0
20
40
60
Time (sec)
FL
RNAP

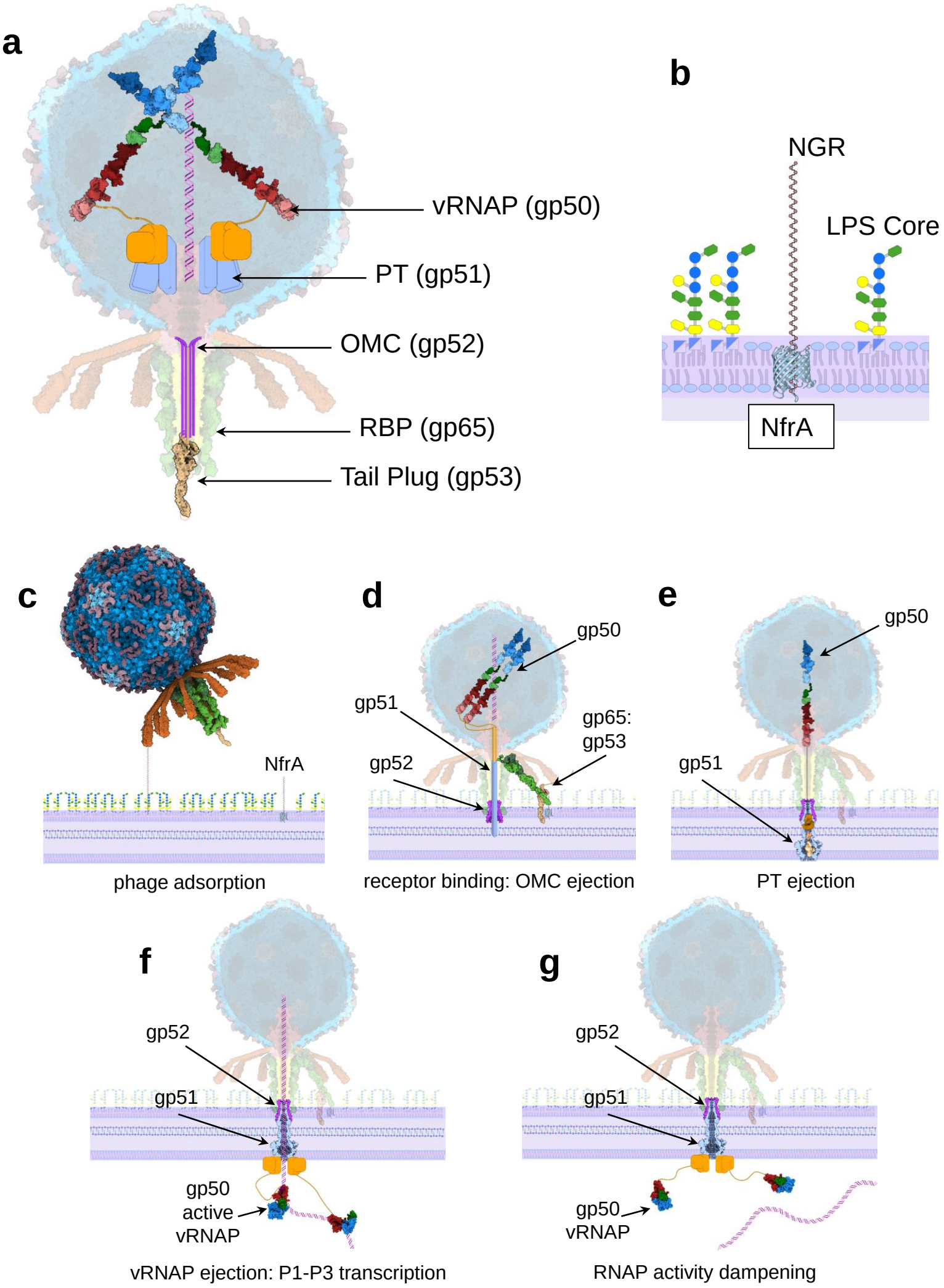

a
vRNAP
PT
OMC
Plug
Tail Tube
gp50
gp51
gp52
gp53
gp54
Ejection Cassette
d
b
Full Virion
c
Empty Particle
e
gp51
f
F16
P10
M1
G8
F7
gp52
C6 Averaged
Asymmetric
Contour = 3 σ
gp59
gp67
gp66
gp52
gp54
gp64
gp65
gp53


a
N
C
Ø$_i$ = 22 Å
b
C
gp52
AF3
N
N
ΔC-gp51
cryo-EM
C
gp51
AF3
300 Å
70 Å
c
300
200
100
0
Ejectosome Length [Å]
OM Complex
Periplasmic Tunnel
gp52
gp51
gp73
gp72
C
N
N
C
C
N
N
C
N4 gp52/gp51 C10
AlphaFold3
DEV gp73/gp72 C9
Cryo-EM: 8VXQ
d
gp52
G [nS]
Time [s]
0.45 nS
1.0 nS
1.0 nS
1.9 nS
e
gp51
G [nS]
Time [s]
3.4 nS
3.4 nS
5.4 nS
3.8 nS
3.0 nS
2.4 nS
1.0 nS
0.5 nS
1.0 nS
0.5 nS
1.0 nS



Tail-gating complex




a
vRNAP (gp50)
PT (gp51)
OMC (gp52)
RBP (gp65)
Tail Plug (gp53)
b
NGR
LPS Core
NfrA
c
NfrA
phage adsorption
d
gp50
gp51
gp65:
gp53
gp52
receptor binding: OMC ejection
e
gp50
gp51
PT ejection
f
gp52
gp51
gp50
active
vRNAP
vRNAP ejection: P1-P3 transcription
g
gp52
gp51
gp50
vRNAP
RNAP activity dampening