Volff J-N (ed): Gene and Protein Evolution.
Genome Dyn. Basel, Karger, 2007, vol 3, pp 48–65

Comparative Genomics and Evolutionary Trajectories of Viral ATP Dependent DNA-Packaging Systems

A.M. Burroughs[a,b], *L.M. Iyer*[a], *L. Aravind*[a]

[a]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, [b]Bioinformatics Program, Boston University, Boston, MA, USA

Abstract

We present an overview of comparative genomics of ATP-dependent DNA packaging systems of viruses. Several distinct ATPase motors and accessory proteins have been identified in DNA-packaging systems of viruses such as terminase-portal systems, the φ29-like packaging apparatus, and packaging systems of lipid inner-membrane-containing viruses. Sequence and structure analysis of these proteins suggest that there were two major independent innovations of ATP-dependent DNA packaging systems in the viral universe. The first of these utilizes a HerA/FtsK superfamily ATPase and is seen in prokaryotic viruses with inner lipid membranes, large eukaryotic nucleo-cytoplasmic DNA viruses (including poxviruses) and a group of eukaryotic mobile DNA transposons. We show that ATPases of the φ29-like packaging system are also divergent versions of the HerA/FtsK superfamily that functions in viruses without an inner membrane. The second system, the terminase-portal system, is dominant in prokaryotic tailed viruses and typically functions with linear chromosomes. The large subunit of this system contains a distinct ATPase domain and a C-terminal nuclease domain of the RNAse H fold. We discuss the classification of these ATPases within the P-loop NTPases, genomic demography and positioning of their genes in the viral chromosome. We show that diverse portal proteins utilized by these systems share a common evolutionary origin and might have frequently displaced each other in evolution. Examination of conserved gene neighborhoods indicates repeated acquisition of Helix-turn-Helix domain-containing terminase small subunits and a third accessory component, the MuF protein. Adenoviruses appear to have evolved a third packaging ATPase, unique to their lineage. Relationship between one major type of packaging ATPases and cellular chromosome pumps like FtsK suggests an ancient common origin for viral packaging and cellular chromosome partitioning systems.

Proper segregation of chromosomes and their partitioning into daughter cells or capsids is a common problem faced by cellular and viral replicons. While diverse solutions to this problem have evolved in different viruses, they may all be categorized under two broad mechanistic themes [1, 2]. Most RNA viruses and several small DNA viruses do not appear to require an active energy-dependent process for packaging their genomes, and the process simply proceeds via coating of nucleic acids by capsid subunits. Coating is usually initiated by packaging signals in the form of sequence or structural features in the nucleic acid, resulting in condensation of the capsid proteins on the nucleic acid scaffold [3, 4]. In the second theme, an active ATP-dependent process loads the genome of larger double stranded (ds) DNA viruses and single stranded (ss) DNA viruses of the Inovirus family into empty capsids [2, 5, 6].

Extreme sequence divergence of viral proteins has hampered understanding of relationships between components of chromosome segregation and packaging systems of different viruses. However, recent availability of a wealth of crystal structures and complete sequences of numerous viral genomes allows us to address this problem using a variety of sequence and structure analysis techniques and comparative genomics. Some recent developments in this regard include structural studies on viral coat proteins revealing that the principal capsid or coat protein of several characterized viruses contains a distinctive β-strand fold with a β-jelly-roll topology [7, 8]. Remarkably, this structural conservation of capsid proteins transcends the diversity of viruses, which might be otherwise unrelated in terms of their genomic nucleic acid or replication and packaging mechanisms. This raised the intriguing possibility that principal capsid proteins of a notable subset of viruses might have descended from a common ancestor [8].

Similarly, sensitive sequence comparisons showed that packaging ATPase motors of diverse large eukaryotic and prokaryotic DNA viruses belong to the HerA-FtsK superfamily, which includes the DNA pumps involved in prokaryotic cellular chromosome segregation and related DNA pumps of several conjugative plasmids and transposons [5]. Packaging ATPases of the HerA/FtsK superfamily are encountered in the recently unified Nucleo-Cytoplasmic Large DNA Virus (NCLDV) assemblage and in several dsDNA phages like PRD1 and the Inovirus family [5, 6]. The other major functionally characterized ATP-dependent DNA-packaging system is the terminase-portal protein system first noticed in caudoviruses (tailed prokaryotic viruses) and herpesviruses [2, 9]. In its most basic form the system consists of the two-subunit terminase complex and a multimeric portal protein (PP) providing a conduit for nucleic acid entry into capsids. The terminase large subunit (TLS) has both ATPase activity that powers DNA translocation and nuclease activity, which cleaves the replicating

DNA into genome-sized fragments [10–13]. In addition to these systems, there are smaller families of packaging ATPases in φ29-like bacteriophages and the adenoviruses whose evolutionary affinities were previously unclear [14, 15].

In this article we build upon these previous studies to provide a synthetic overview of the protein components of various ATP-dependent phage DNA-packaging systems. We establish the evolutionary affinities of several poorly understood components and also describe new potential components. Relationships and structural features of packaging proteins presented here also throw light on various functional aspects of DNA packaging, with general implications for the origins of chromosome segregation.

Results and Discussion

The Demography of Packaging ATPases in Large DNA Viruses

Amongst large eukaryotic DNA viruses, all NCLDVs, namely poxviruses, iridoviruses, African Swine Fever Virus, phycodnaviruses and the mimivirus, share a packaging ATPase of the HerA/FtsK superfamily [5,16]. The herpesviruses contain a terminase-portal packaging system similar to the bacteriophages [9]. Additionally, a HerA/FtsK-type ATPase is also encoded by a novel DNA transposon that is widespread in *Trichomonas*, ciliate and nematode genomes [5]. The transposon also has some relationship with adenoviruses in terms of its DNA polymerase and processing protease, suggesting that it might assemble into virus like particles aided by this ATPase [5, 17]. The predicted packaging ATPase of the adenoviruses has thus far not been seen in any other viral lineage [15, 18]. Packaging ATPases, if any, of certain large DNA viruses like baculoviruses and the shrimp white spot syndrome virus are unknown, but they are unlikely to define large new lineages of packaging enzymes.

Our systematic survey of phage packaging system components in completed genomes of 288 prokaryotic DNA viruses showed that they encompass a comparable diversity in terms of their ATPases. Up to a genome size of about 20 kb there is a steady increase in the fraction of phages encoding packaging ATPases (fig. 1a). Beyond this size, 95% of phages encode a packaging ATPase. The majority of small phages lacking packaging ATPases are microviruses, which initiate their packaging through a passive interaction with a small genomically encoded polypeptide [4]. This suggests that 20 kb is the approximate size threshold above which packaging appears to require an active energy-dependent process. The most common packaging ATPase in currently available phages is the terminase-type ATPase (seen in ~70% of the phages), whereas ~15% of phages utilize a version of the HerA/FtsK ATPase superfamily (see supplementary material: SM). The presence of a terminase-type ATPase

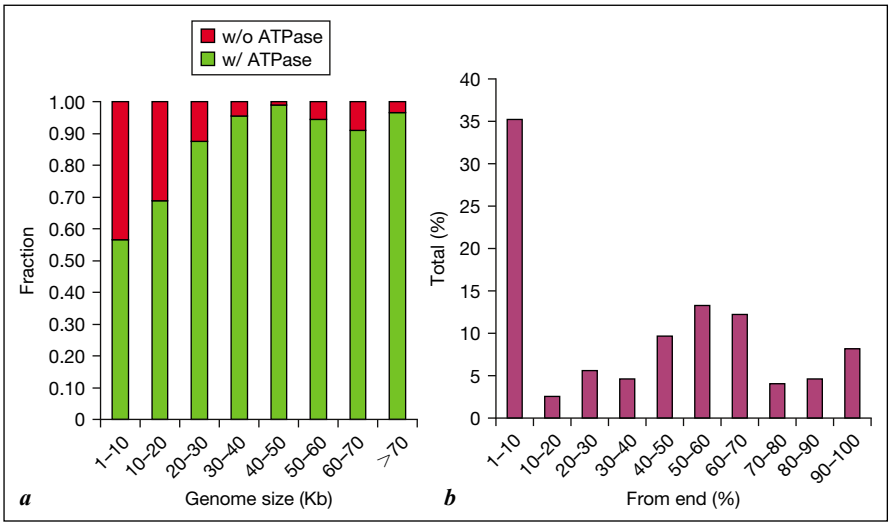


Fig. 1. Packaging ATPase presence/absence and positional distribution in viral genomes. ***a*** Presence/absence of a packaging ATPase in completely-sequenced viral genomes is depicted as a stacked column graph. Percentages of genomes containing a packaging ATPase within a certain genome size range are green columns while percentages lacking an ATPase are red. ***b*** Genome position frequency distribution of packaging ATPases from completely-sequenced viral genomes with linear chromosomes is shown as bars graph. Statistically significant preference for placement in the middle or termini of viral genomes is observed (χ^2: $p < 10^{-5}$).

is strongly correlated with the tailed capsid morphology typical of caudoviruses, the most common type of bacteriophage. The HerA/FtsK family appears to exclusively occur in phages with internal lipid membranes, such as tectiviruses, corticoviruses and *Sulfolobus* turreted virus, irrespective of their outer protein coat morphology (SM) [5, 6, 19]. These phages also often contain terminal inverted repeats in their genomes. Furthermore, 85% of viruses with terminase-type ATPases have linear chromosomes, while 68% of those with HerA/FtsK type ATPases have circular chromosomes (SM). This suggests that while each system can handle either chromosome type, there might be a preferred type for each of them.

A study of the positional distribution of genes for packaging ATPases in phages with linear genomes revealed that in 70% of the cases they are either located at an end or close to the center of the genome (fig. 1b). This unusual distribution is highly significant ($p < 10^{-5}$ by Chi-test) and appears to be related to the time of transcription of the packaging ATPase in the virus life

cycle. Placement of these genes towards the chromosome termini or in the middle may allow late transcription, thereby making the packaging apparatus available only at the last phase of the viral cycle. This bias in chromosomal position of the gene for packaging ATPases provides a contextual means of predicting potential packaging ATPases of uncharacterized viruses. We observed two archaeal globuloviruses (*Thermoproteus tenax* spherical virus 1: TTSV and *Pyrobaculum* spherical virus: PSV) with genome sizes greater than 20 kb lacking any known packaging ATPases. However, both viruses encode an uncharacterized P-loop NTPase at the termini of their genomes (TTSV: ORF1 and PSV: ORF582). Sequence searches with these proteins showed no close relation to any other ATPases involved in replication such as helicases or clamp loaders; supporting its possible role as a packaging ATPase.

Multiple Origins for Different Packaging ATPases within the P-loop NTPase Fold

All known and predicted packaging ATPases of DNA viruses belong to the P-loop NTPase fold, one of the most prevalent protein folds in both cellular and viral genomes [2, 5, 15]. Members of the P-loop NTPase fold are unified by the conserved nucleotide binding (Walker A) and Mg^{2+} binding motifs (Walker B) and belong to one of two major divisions; the KG division which includes P-loop kinases and GTPases, and the ASCE (additional strand conserved E (glutamate)) division [20, 21]. The latter division is characterized by an additional conserved acidic residue (typically a glutamate occurring immediately after the conserved Walker B aspartate) and a conserved polar residue (Sensor 1) occurring at the end of the 4th core strand of the domain [21, 22]. Examination of all characterized packaging ATPase domains, namely the TLS N-terminal domain, the HerA/FtsK ATPase domain, the φ29-like phage ATPase domain, and the putative adenoviral packaging ATPase domain revealed hallmark features of the ASCE division, indicating derivation from within this radiation of the P-loop fold [5, 15]. This observation is consistent with the fact that the majority of highly active ATPases mediating energy dependent processes in biological systems belong to the ASCE division [20–22].

However, relationships between different viral packaging ATPases and affinities to other major classes of ATPases of the ASCE group have remained largely unclear. Previous systematic analysis of the HerA/FtsK superfamily revealed that viral packaging ATPases of this superfamily do not form an exclusive virus-specific clade but are successive out-groups of the crown-group formed by cellular HerA and FtsK families [5]. The basal-most clade was comprised of packaging ATPases of filamentous inoviruses with ssDNA genomes while those from remaining diverse groups of lipid membrane-containing dsDNA viruses of prokaryotes and eukaryotes formed a large assemblage, an

immediate sister group of the cellular and plasmid members of this superfamily [5]. Preliminary sequence searches with φ29-like ATPases recovered only cognate ATPases of other related viruses and secondary structure prediction revealed that φ29-like ATPases contained an α-β unit C-terminal to strand-2 as seen in the FtsK, RecA, helicase and PilT assemblage within the ASCE division. Furthermore, we noted that φ29-like ATPases bore a conserved arginine at the base of strand-4, equivalent to identically positioned arginine fingers in HerA/FtsK ATPases. φ29-like ATPases also possessed a conserved asparagine at the end of the sensor-1 strand, equivalent to the glutamine seen in the HerA/FtsK superfamily (SM). These observations together with the statistically significant recovery of φ29-ATPases by sensitive profiles of the HerA/FtsK superfamily indicate that the former are a distinct branch of the latter superfamily. However, there were no specific features that unified φ29-like ATPases with HerA/FtsK-type packaging ATPases of other dsDNA viruses, suggesting that they are a rapidly diverging independent lineage within the HerA/FtsK superfamily.

TLSs are almost always two domain proteins with an N-terminal ASCE-type P-loop ATPase domain and a C-terminal nuclease domain with a RuvC-like version of the RNaseH fold [23, 24]. The secondary structure of the terminase ATPase domain revealed the presence of at least one additional strand after strand-2 (SM, fig. 2), placing them in a monophyletic assemblage of the ASCE division along with the HerA/FtsK, PilT, RecA and helicase superfamilies [5]. However, they lack the C-terminal β-hairpin or any other specific features characteristic of most members of the above assemblage [5] (SM). The TLS ATPase domain is distinguished from other related ATPases by the presence of a poorly conserved but universally present insert after the second β-α unit which includes strand-2. They also contain a characteristic arginine at the third position in the Walker A motif. While it could potentially act as an arginine finger in the terminase multimer, such a function remains uncertain as the arginine is absent in a few active terminases, like that of phage T1 (SM). Thus, it appears that TLS ATPase domains comprise a separate lineage within the above monophyletic assembly of the ASCE division.

The predicted adenoviral packaging ATPase (IVA2) consistently retrieved ABC ATPases as best hits. They specifically share with the ABC ATPases two polar residues at the end of the sensor-1 strand, one of which is a highly conserved histidine. Secondary structure predictions also suggest that they contain an insert with β-strands after helix-1 which might be equivalent to the corresponding insert found in all ABC ATPases [25]. Hence, the adenoviral IVA2 proteins were potentially derived from the ABC superfamily. These putative ATPases additionally contain a distinct C-terminal extension predicted to form an α+β domain with two conserved aromatic positions and several polar

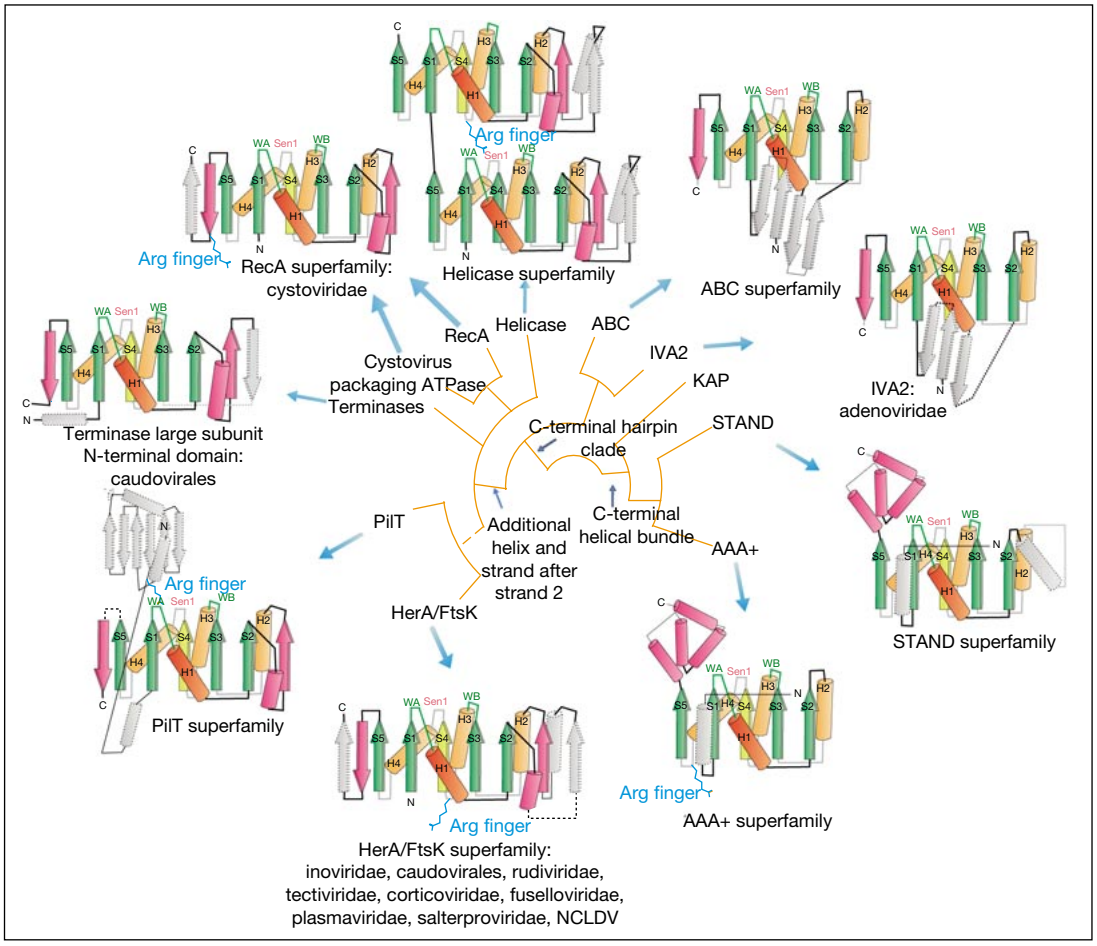


Fig. 2. Topology diagrams depicting ASCE division of P-loop NTPases and accompanying cladogram depicting higher-order relationships. Viral lineages with packaging ATPases from a specific superfamily are listed following a colon below the superfamily name. Strands and helices forming the core of the ASCE P-loop NTPase domain are numbered and colored. Strands are in green with the central strand S4 in yellow and helices in orange. Synapomorphies shared across different lineages are colored pink, elements not conserved across lineages are colored gray and outlined in broken lines. Lines connecting different lineages represent higher-order relationships constructed by comparison of shared structural and/or sequence similarities. Broken lines represent relationships with more uncertainty. Abbreviations: WA, Walker A; WB, Walker B and Sen1, sensor-1.

residues (SM). It might play a role in recognizing packaging-initiation signals in genomic regions.

Thus, it appears that DNA packaging has been derived on at least three independent occasions within the ASCE division of P-loop NTPases, with two of them being exclusively comprised of packaging ATPases (HerA/FtsK and terminase), and the third from a superfamily of ATPases that were ancestrally associated with DNA-related functions (ABC ATPases). The predicted packaging ATPases of archaeal globuloviruses can currently only be identified as members of the ASCE division with no close relationships to any of the other three classes, and might represent a fourth independent innovation. Interestingly, the only characterized packaging ATPases of dsRNA viruses, those of cystoviruses (e.g. φ12), represent another independent recruitment for packaging function from within the RecA superfamily [26, 27] (fig. 2).

Ancilliary Components of DNA Packaging Systems

Functional studies to date have not uncovered any conserved system of interacting proteins that function along with viral members of the HerA/FtsK superfamily. Previous studies have shown their cooperation with diverse nucleases in resolving target DNA during active pumping by these ATPases [5]. Hence, it is likely that these ATPases cooperate during packaging with different resolvases, including the frequently present RuvC-like resolvase, in NCLDVs and prokaryotic dsDNA viruses [19, 28, 29]. The φ-29 lineage of the HerA/FtsK superfamily appears to utilize a distinct portal protein (PP) containing a globular domain of the Src Homology 3 (SH3) fold [30]. This domain forms a multimeric ring similar to those formed by other nucleic acid binding members of this fold such as the RNA-binding Sm domain [30–32]. Adenoviruses possess a unique ancillary protein, not observed elsewhere in the viral universe, which probably functions in conjunction with the IVA2 protein to interact with genomic packaging sequences [33, 34]. Secondary structure predictions of this protein indicate a lineage-specific α-helical fold. Terminase systems show considerable diversity with different types of PPs and ancillary components like terminase small subunits (TSS). Given this diversity, we sought to investigate their origins and identify new interacting components using genomic context information.

Diversity of the Terminase-Dependent Packaging Systems: A Common Origin for Portal Proteins of All Tailed Bacteriophages

In addition to the TLS whose two domains supply ATP-dependent motor and nuclease activity, packaging systems of all characterized caudoviruses also require a PP. PPs form homo-multimers providing a conduit for DNA into the viral prohead [30, 31, 35]. In contrast to the common origin of TLSs, PPs of

these viruses were believed to belong to distinct families, typified by versions found in phage T4, T5, λ, A118 and Mu [36]. To investigate evolutionary affinities of PPs we initiated systematic transitive sequence profile searches from all known versions of PPs. As a result of these searches, we were able to recover PPs from a variety of phages or their equivalents such as the head-tail connector protein (gp8) of phage T7; consequently unifying all known PPs of tailed bacteriophages. These searches showed that every TLS-containing phage also encoded one predicted PP suggesting a strict functional association (SM). Unification of PPs of diverse phage families also implied descent from a common ancestor, just as their terminase counterparts. However, they have subsequently undergone rather drastic sequence divergence. Secondary structure prediction of the conserved core shared by PPs indicates a six-stranded region embedded between two predominantly α-helical elements. The most prominent sequence conservation is in the β-strand rich region and includes a Gxs (where 'x' is any amino acid and 's' a small residue) prior to the first conserved strand (SM). Sequence similarity-based clustering and examination of conserved shared motifs in the alignment helped us to discern eight distinct families (SM), which further grouped together into four higher order clades (T1/T5/λ-like clade, the T4/SPP1/φg1e-like clade, the phage μ-like clade and the phage T3/T7-like clade).

The presence of a conserved β-strand-rich region in PPs is reminiscent of the SH3 fold β-barrel in the φ29-type PP. The conserved Gxs motif in the former superfamily is also reminiscent of a similar motif seen in the corresponding position of SH3-like barrels [37]. Hence, despite lack of significant sequence similarity, it is not impossible that a similar β-barrel might be present in PPs of terminase-dependent systems. Likewise, herpesviral PPs, while displaying no significant sequence similarity to those of bacteriophages, also contain a core β-strand rich region suggesting the presence of a similar structure (data not shown). We propose that this β-strand rich region might form a comparable β-barrel domain, which multimerizes to give rise to the funnel shaped portal.

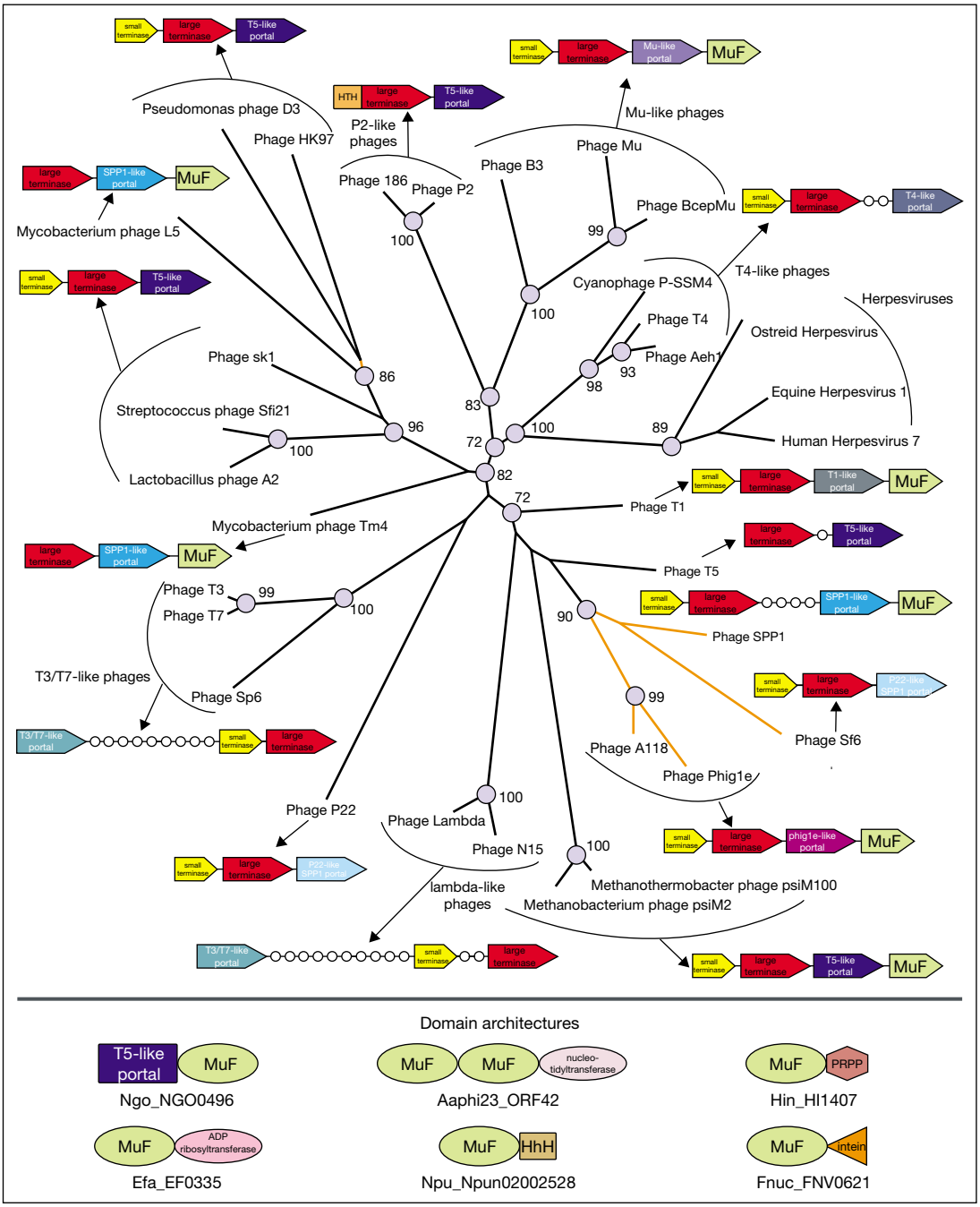
Contextual Information and Inference of Novel Components of the Terminase Portal Systems

Conserved gene neighborhoods (operons) and gene fusions have proven to be a powerful method for predicting previously unknown functional associations and protein-protein interactions in prokaryotes and their viruses [38, 39]. In order to identify other functional links to the terminase-portal system, we systematically explored all gene neighborhoods of terminase-portal pairs in bacteriophages. In terms of gene neighborhood, the most commonly found association is between the TLS and the PP, which typically occur as neighboring genes in several viral genomes (some exceptions include λ, T3/T7 and T5)

(SM, fig. 3). PP genes are rarely fused to other genes suggesting that multimerization and strict interactions with TLS are likely to select against fusion proteins. One notable fusion of the PP is with a lysozyme (e.g. *Burkholderia* prophage, gi: 78061894; fig. 3), which might correlate with the incorporation of lysozymes in viral capsids for their role in host entry.

Terminase small subunits (TSS) have been characterized in phages such as T4, T7, λ and SPP1, but corresponding small subunits have not been found in many other tailed bacteriophages [40–42]. Examination of gene neighborhoods suggested a strong association between the genes for the TSS and the TLS (fig. 3). The crystal structure of the λ small subunit shows a specialized derivative of the winged Helix-Turn-Helix (HTH) domain – the MerR-like HTH, which lacks the first of three characteristic helices of classical HTHs [43]. This suggests that the primary role of the TSS is binding DNA. Accordingly, we combined the contextual information of gene neighborhood and sequence profile searches to characterize the other TSSs and identify previously undetected versions. Our searches identified TSSs in 151 of the 206 phages containing terminase-portal systems. While all these small subunits contain the HTH fold, they included versions distinct from the MerR-type HTH seen in λ-like TSSs. In total, we identified seven distinct families of TSS and also few sporadic unclassified HTH domains. Of these, the largest families were SPP1-type TSS and D3-like TSS. The SPP1-like family was shown to contain a simple trihelical HTH module of the FIS type, while the remaining families did not belong to any previously characterized type of HTH domain and likely represent phage-specific divergent versions of the fold (SM). In a subset of phages, including P2, the SPP1-like TSS is fused to the TLS, supporting the strong functional association between the two subunits through physical interaction (SM). The above observations suggest that unlike the TLS, the TSS has been derived from the HTH fold on multiple occasions, and convergently evolved similar functional associations with the TLS.

The next major family of proteins, often encoded in the same conserved gene neighborhoods as other components of the terminase system, is the so-called MuF family. This family, typified by phage SPP1 gp7 protein, is a component of the phage prohead. In bacteriophages infecting Gram-positive bacteria, the MuF protein is known to associate with the PP and is believed to be led into the prohead by the latter [44, 45]. In our sequence profile searches, we detected MuF proteins in representatives of all major tailed prokaryotic virus families and their prophage derivatives (including one in archaeon *Methanococcus*: MJ0329). Nevertheless, several phages in each of these families lacked MuF, suggesting that it might not be an essential component of terminase-portal systems (fig. 3). The MuF gene is almost always immediately downstream of the PP gene and is associated with genes for several distinct



families of portal proteins in different phages like T1, T5, Mu, λ and SPP1 (fig. 3). In one instance it is fused to a T5-like portal gene (*Neisseria* prophage, gi: 59800934), reinforcing the strong functional association between these two components.

MuF contains a characteristic C-terminal region with conserved cysteines, histidines and acidic residues suggesting it might form a distinct metal-chelating domain, which might be involved in MuF-mediated DNA binding activity. MuF proteins show a number of fusions to other domains in several (pro)phages. These include fusions to the DNA-binding Helix-hairpin-Helix (HhH) domain (*Nostoc* prophage, gi: 23130420) and several catalytic domains such as ADP ribosyltransferase (*Enterococcus* prophage, gi: 29374974), pol-β-fold nucleotidyltransferase (phage Aaφ23, gi: 31408074), PRPP amidotransferase (*Haemophilus* prophage, gi: 16273315) and multiple intein-type HINT peptidase domains (*Fusobacterium*, gi: 34763916; *Bifidobacterium*, gi: 23335596). ADP ribosyltransferases have been observed in a variety of phages, including T4 and eukaryotic NCLDVs, like PBCV and mimivirus [19]. T4 ADP ribosyltransferases ModA, ModB and Alt are packaged into phage heads, and are involved in modifying a range of host proteins [46]. Hence, the MuF might help in loading ADP-ribosyltransferase and other catalytic activities in the phage head for modification of host or viral proteins. HINT peptidases fused to MuF are related to the BUBL1 peptidase of ciliates which is involved in cleaving tandemly-fused ubiquitin repeats and ADP ribosyltransferase domains [47]. Consequently, MuF associated HINT peptidases might be similarly involved in phage head maturation. In this context, it should be noted that the portal-terminase system genes including MuF are often combined with another conserved gene neighborhood, which contains proteases involved in capsid maturation belonging to ClpP or herpesvirus assemblin-like folds [49].

Fig. 3. Phylogenetic tree of TLS depicting gene displacement among portal protein families. Phylogenetic trees were built using the least-square method with subsequent local rearrangement to obtain the maximum likelihood tree (see SM for details). Reliability of the tree topology was assessed using the RELL bootstrap method of MOLPHY, with 10,000 replications (SM). Branches where gene displacement has occurred as discussed in the text are colored orange for emphasis. Gene neighborhoods corresponding to TLSs are adjacent to branch ends, genes are shown as boxed arrows. TLS genes are colored in red, TSS colored in yellow, MuF colored in green, and PPs are colored according to family type. Nodes with bootstrap support >70% are linked by circles and labeled by bootstrap value. Domain architectures are also given below the tree, with organism abbreviations and gene names (separated by an underscore) written below. Abbreviations: HhH, helix-hairpin-helix; PRPP, PRPP amidotransferase. Please see SM for phage abbreviations.

In Situ Gene Displacement in Terminase Portal Gene Neighborhoods

Diversification of PPs into several distinct subgroups and recruitment of several distinct types of HTH domains as TSS raised the question of whether there was a correlation between distinct families of these proteins and the phylogeny of TLSs. Only TLSs show sufficient sequence conservation to reconstruct a suitably resolved phylogenetic tree through conventional methods (fig. 3). Hence, we used this tree as a reference to study the distribution of other components of the terminase-portal system and structures of their gene neighborhoods. This distribution showed the following features: (1) MuF proteins show a sporadic distribution with related phages often differing in its presence or absence. (2) Phages with related TLS might often differ in the type of PP or TSS they are associated with. For example, phage SPP1 has an SPP1-like PP (PP2 family) while the related phage Sf6 contains a P22-like version. Likewise, related TLSs of phages P2 and B3 differ in terms of associated TSS and PP and presence or absence of MuF (fig. 3).

These observations suggest that terminase-portal gene neighborhoods are prone to: (1) frequent gene loss and acquisition, evidenced by sporadic distribution of MuF and (2) in situ displacement of functionally equivalent proteins by evolutionarily unrelated or distantly related counterparts. This situation is parallel to previously observed gene neighborhoods of phage single strand annealing proteins and capsid maturation proteases [48, 49]. Presence of relatively strict gene orders (TSS followed by TLS, PP and MuF) suggests strong constraints with respect to their synthesis and interactions. General rarity or absence of gene fusions among TSS, TLS and PP suggest that their interactions are strongly coupled without much scope for additional associations. Based on gene order and nature of domain fusions, we speculate that TSS is synthesized first and associates with viral DNA. It subsequently recruits the TLS which processes DNA and recruits the PP through which DNA is loaded into the prohead. The PP in turn appears to recruit MuF, which might help position DNA into proheads and recruit other catalytic activities for capsid maturation.

Evolutionary Considerations and General Conclusions

The systematic survey of diverse active viral DNA-packaging systems suggests that their motors have been derived from two major superfamilies of ASCE ATPases (HerA/FtsK and TLS N-terminal domain). The remaining packaging motors are also derived from the ASCE division, but are very limited in their spread and appear to lack an extended evolutionary history. Taken together with the monophyly of capsid proteins of several DNA and RNA viruses, this suggests an early origin for the two major ATP-dependent DNA-packaging systems in the context of ancient pre-existing capsid-like envelopes [19].

Interestingly, both ancient superfamilies of packaging ATPases function in conjunction with DNAses that process or manipulate the products of genome replication. While TLSs contain the C-terminal RNaseH fold nuclease domain, the HerA/FtsK superfamily functions with a range of distinct nucleases in cellular and viral systems, such as XerC/XerD, NurA, RCR, pT181/Rep, Sir2 and possibly RuvC-like resolvases (in several NCLDVs) [5, 28, 29]. The RNaseH-fold domain in TLS is most closely related in terms of its conserved active site to RuvC resolvases and nuclease domains of several transposases (such as TnpA, Mariner, Hermes, Rag1/Transib and retroviral integrases) [24, 50, 51]. Thus, ATPases of both packaging systems probably associated with an ancestral DNA manipulating nuclease of the RNaseH fold, which appears to have diversified into nuclease, integrase or resolvase families of viral and cellular replicons. HerA/FtsK ATPases form ring-structures and lack domain fusions with their nuclease partners. This appears to have allowed more frequent evolutionary displacements of their nuclease partners by functionally equivalent nucleases [5]. In contrast, there is no evidence for TLSs forming comparable arginine finger-stabilized rings, and fusion with their nuclease partner appears to have been retained throughout their evolution. In general, functional associations between nucleases and packaging ATPases suggest that from inception packaging systems were closely associated with post-replication genome segregation. Increasing size of DNA-based genomes probably provided the selection pressure for emergence of such systems [19].

Interestingly, diversification of several other superfamilies in the ASCE division of P-loop NTPases might be linked to emergence and expansion of DNA-based replication systems. These include DNA helicases of AAA+, recombinases of RecA, and higher order chromosome condensation proteins of ABC superfamilies. Hence, the two major DNA packaging systems probably arose as part of this diversification of ASCE NTPases concomitant with diversification of DNA-based replicons that occurred well before the emergence of the Last Universal Common Ancestor (LUCA) of cellular life [5]. The nature of the envelope of early replicons, lipid membranes or purely protein capsids, appears to have played a principal role in emergence of the two independent packaging motors. In this context, it is notable that cellular systems (bacteria and archaea) use packaging ATPases related to those of viruses with lipid inner membranes. Thus, precursors of cellular compartments could have emerged from systems similar to lipid containing viral capsids [19]. Thus, it appears that the precursors of the principal packaging ATPases of viral systems and cellular chromosome-pumping ATPases, like FtsK appear to have emerged during the pre-LUCA radiation of the ASCE clade, and followed independent history ever since. While both major packaging systems remained largely mutually exclusive in viruses, on rare occasions we do find potential hybrid systems. The φ29 system uses a

HerA/FtsK ATPase but depends on a PP analogous to caudoviruses. Like the latter, it lacks an inner membrane, and has a unique hexameric RNA component (prohead RNA or pRNA) [52]. It remains unclear if this pRNA is a remnant of a more ancient system or merely a lineage-specific innovation of φ29-like phages. Similarly, evolution of adenoviruses might have involved displacement of the HerA/FtsK ATPase of the above-mentioned Tlr-like DNA transposons by a neomorphic packaging system. The adenovirus IVA protein, as well as RecA-like packaging ATPases limited to cystoviruses, could have emerged via rapid divergence from either viral or cellular precursors. Our unification of PPs suggests that the terminase-dependent system deployed PPs from the earliest stages of their existence. The observation that most of these viruses also contain a version of the HTH domain (TSS) suggests that there might have been a third component that recruited the motor to DNA even in ancestral versions of this system.

We hope this overview might help in further experimental investigations on functional interactions in these systems.

Acknowledgements

The authors gratefully acknowledge the Intramural Research Program of the National Institutes of Health, USA for funding their research.

Supplementary Material

The supplementary material can be accessed from ftp://ftp.ncbi.nih.gov/pub/aravind/portal/.

Note Added in Proof

While this manuscript was being processed for publication, the crystal structure of the TLS was solved [53] and was shown to belong to the ASCE division as predicted. The presence of a helix strand unit after strand-2 reaffirmed its position with respect to other ATPases (fig. 2). The C-terminal region, however, is much diverged from other ATPases. The crystal structure also confirms the presence of an arginine finger in Walker A as predicted.

References

1 Wagner KE, Hewlett MJ: Basic Virology, ed 2. Blackwell Publishers, Oxford, 2003.
2 Catalano CE: Viral Genome Packaging: Genetics, Structure, and Mechanism, Kluwer Academic/Plenum publisher, New York, 2005.

3 Rao AL: Genome packaging by spherical plant RNA viruses. Annu Rev Phytopathol 2006;44: 61–87.
4 Bernal RA, Hafenstein S, Esmeralda R, Fane BA, Rossmann MG: The phiX174 protein J mediates DNA packaging and viral attachment to host cells. J Mol Biol 2004;337:1109–1122.
5 Iyer LM, Makarova KS, Koonin EV, Aravind L: Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 2004;32:5260–5279.
6 Stromsten NJ, Bamford DH, Bamford JKH: In vitro DNA packaging of PRD1: a common mechanism for internal-membrane viruses. J Mol Biol 2005;348:617–629.
7 Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, et al: The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc Natl Acad Sci USA 2002;99: 14758–14763.
8 Hendrix RW: Evolution: the long evolutionary reach of viruses. Curr Biol 1999;9:914–917.
9 Newcomb WW, Juhas RM, Thomsen DR, Homa FL, Burch AD, et al: The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. J Virol 2001;75:10923–10932.
10 Catalano CE: The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. Cell Mol Life Sci 2000;57:128–148.
11 Black LW: DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism. Bioessays 1995;17:1025–1030.
12 Rentas FJ, Rao VB: Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. J Mol Biol 2003;334:37–52.
13 Goetzinger KR, Rao VB: Defining the ATPase center of bacteriophage T4 DNA packaging machine: requirement for a catalytic glutamate residue in the large terminase protein gp17. J Mol Biol 2003;331:139–154.
14 Ibarra B, Valpuesta JM, Carrascosa JL: Purification and functional characterization of p16, the ATPase of the bacteriophage Phi29 packaging machinery. Nucleic Acids Res 2001;29:4264–4273.
15 Koonin EV, Senkevich TG, Chernos VI: Gene A32 product of vaccinia virus may be an ATPase involved in viral DNA packaging as indicated by sequence comparisons with other putative viral ATPases. Virus Genes 1993;7:89–94.
16 Iyer LM, Aravind L, Koonin EV: Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 2001;75:11720–11734.
17 Wuitschick JD, Gershan JA, Lochowicz AJ, Li S, Karrer KM: A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. Nucleic Acids Res 2002;30:2524–2537.
18 Zhang W, Imperiale MJ: Requirement of the adenovirus IVa2 protein for virus assembly. J Virol 2003;77:3586–3594.
19 Iyer LM, Balaji S, Koonin EV, Aravind L: Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 2006;117:156–184.
20 Leipe DD, Wolf YI, Koonin EV, Aravind L: Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002;317:41–72.
21 Iyer LM, Leipe DD, Koonin EV, Aravind L: Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 2004;146:11–31.
22 Neuwald AF, Aravind L, Spouge JL, Koonin EV: AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 1999;9: 27–43.
23 Kanamaru S, Kondabagil K, Rossmann MG, Rao VB: The functional domains of bacteriophage t4 terminase. J Biol Chem 2004;279:40795–40801.
24 Ponchon L, Boulanger P, Labesse G, Letellier L: The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily: a bioinformatics analysis validated by a functional study on bacteriophage T5. J Biol Chem 2006;281:5829–5836.
25 Holland IB, Blight MA: ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J Mol Biol 1999;293:381–399.
26 Lisal J, Kainov DE, Bamford DH, Thomas GJ Jr, Tuma R: Enzymatic mechanism of RNA translocation in double-stranded RNA bacteriophages. J Biol Chem 2004;279:1343–1350.

27 Kainov DE, Pirttimaa M, Tuma R, Butcher SJ, Thomas GJ Jr, et al: RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. J Biol Chem 2003;278:48084–48091.
28 Garcia AD, Aravind L, Koonin EV, Moss B: Bacterial-type DNA holliday junction resolvases in eukaryotic viruses. Proc Natl Acad Sci USA 2000;97:8926–8931.
29 Aravind L, Makarova KS, Koonin EV: SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res 2000;28:3417–3432.
30 Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, et al: Structure of the bacteriophage phi29 DNA packaging motor. Nature 2000;408:745–750.
31 Guasch A, Pous J, Parraga A, Valpuesta JM, Carrascosa JL, Coll M: Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage phi29 connector particle. J Mol Biol 1998;281: 219–225.
32 Mura C, Phillips M, Kozhukhovsky A, Eisenberg D: Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci USA 2003;100:4539–4544.
33 Perez-Romero P, Gustin KE, Imperiale MJ: Dependence of the encapsidation function of the adenovirus L1 52/55-kilodalton protein on its ability to bind the packaging sequence. J Virol 2006;80: 1965–1971.
34 Gustin KE, Lutz P, Imperiale MJ: Interaction of the adenovirus L1 52/55-kilodalton protein with the IVa2 gene product during infection. J Virol 1996;70:6463–6467.
35 Bazinet C, Benbasat J, King J, Carazo JM, Carrascosa JL: Purification and organization of the gene 1 portal protein required for phage P22 DNA packaging. Biochemistry 1988;27:1849–1856.
36 Mitchell MS, Matsuzaki S, Imai S, Rao VB: Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. Nucleic Acids Res 2002;30:4009–4021.
37 Anantharaman V, Aravind L: Novel conserved domains in proteins with predicted roles in eukaryotic cell-cycle regulation, decapping and RNA stability. BMC Genomics 2004;5:45.
38 Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV: Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 2001;11: 356–372.
39 Huynen M, Snel B, Lathe W 3rd, Bork P: Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 2000;10:1204–1210.
40 Gual A, Alonso JC: Characterization of the small subunit of the terminase enzyme of the *Bacillus subtilis* bacteriophage SPP1. Virology 1998;242:279–287.
41 Lin H, Simon MN, Black LW: Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. J Biol Chem 1997;272:3495–3501.
42 Bain DL, Berton N, Ortega M, Baran J, Yang Q, Catalano CE: Biophysical characterization of the DNA binding domain of gpNu1, a viral DNA packaging protein. J Biol Chem 2001;276: 20175–20181.
43 de Beer T, Fang J, Ortega M, Yang Q, Maes L, et al: Insights into specific DNA recognition during the assembly of a viral genome packaging machine. Mol Cell 2002;9:981–991.
44 Stiege AC, Isidro A, Droge A, Tavares P: Specific targeting of a DNA-binding protein to the SPP1 procapsid by interaction with the portal oligomer. Mol Microbiol 2003;49:1201–1212.
45 Droge A, Santos MA, Stiege AC, Alonso JC, Lurz R, et al: Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J Mol Biol 2000;296:117–132.
46 Depping R, Lohaus C, Meyer HE, Ruger W: The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. Biochem Biophys Res Commun 2005;335: 1217–1223.
47 Dassa B, Yanai I, Pietrokovski S: New type of polyubiquitin-like genes with intein-like autoprocessing domains. Trends Genet 2004;20:538–542.
48 Iyer LM, Koonin EV, Aravind L: Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics 2002;3:8.
49 Liu J, Mushegian A: Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. J Bacteriol 2004;186:4369–4375.

50 Rice PA, Baker TA: Comparative architecture of transposase and integrase complexes. Nat Struct Biol 2001;8:302–307.
51 Kapitonov VV, Jurka J: RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 2005;3:e181.
52 Xiao F, Moll WD, Guo S, Guo P: Binding of pRNA to the N-terminal 14 amino acids of connector protein of bacteriophage phi29. Nucleic Acids Res 2005;33:2640–2649.
53 Sun S, Kondabagil K, Gentz PM, Rossmann MG, Rao VB: The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. Mol Cell 2007;25:943–949.

L. Aravind
National Center for Biotechnology Information,
National Library of Medicine,
National Institutes of Health,
Bethesda, MD 20894, USA
Tel (301) 594-2445, Fax (301) 480-9241, E-Mail aravind@ncbi.nlm.nih.gov