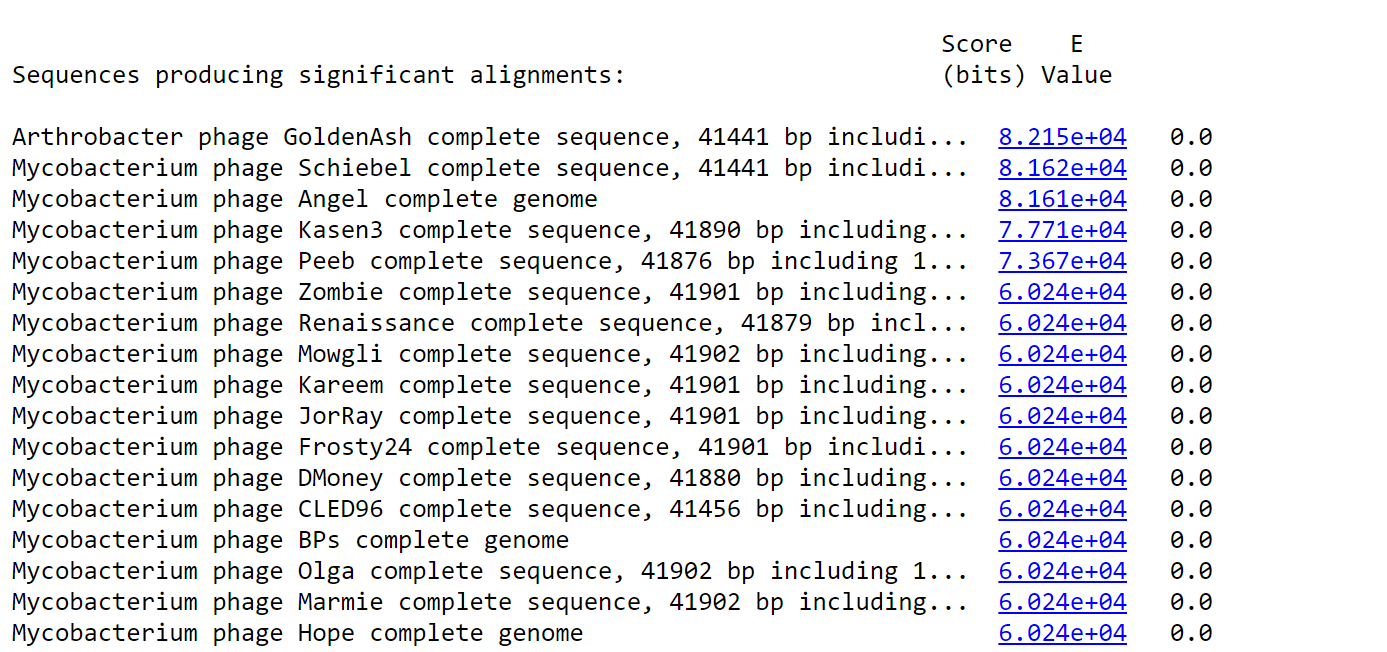
**BIOS 252 Science Research III**

**Annotation Template**

**Panagakis**

List most closely related phages to **GoldenAsh** by BLAST in phagesdb.org:

**Total Genome**



Schiebel-annotated 2006

Angel-2007

Kasen3-2016

Peeb-2018

Zombie-2012

Renaissance-2016

Mowgli-2013

Kareem-2016

JorRay-2019-Draft

Frosty24-2011

DMoney-2017

CLED96-2015

BPs-2006

Olga-2017

Marmie-2021-Draft

Hope-2008

Annotation for Gene # \_\_\_1\_\_\_\_\_ of Phage **GoldenAsh** (**Cluster G1**)

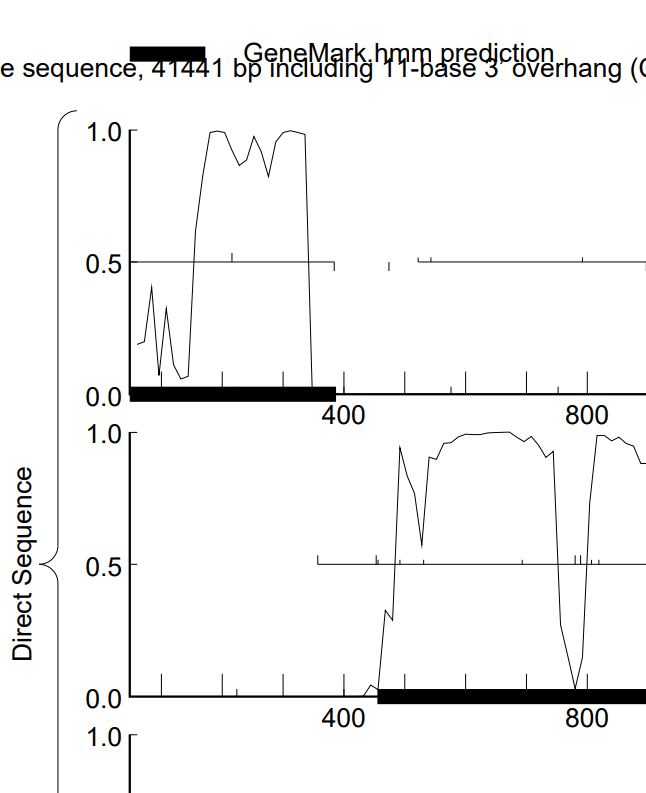
Glimmer predicted nucleotide start : 46

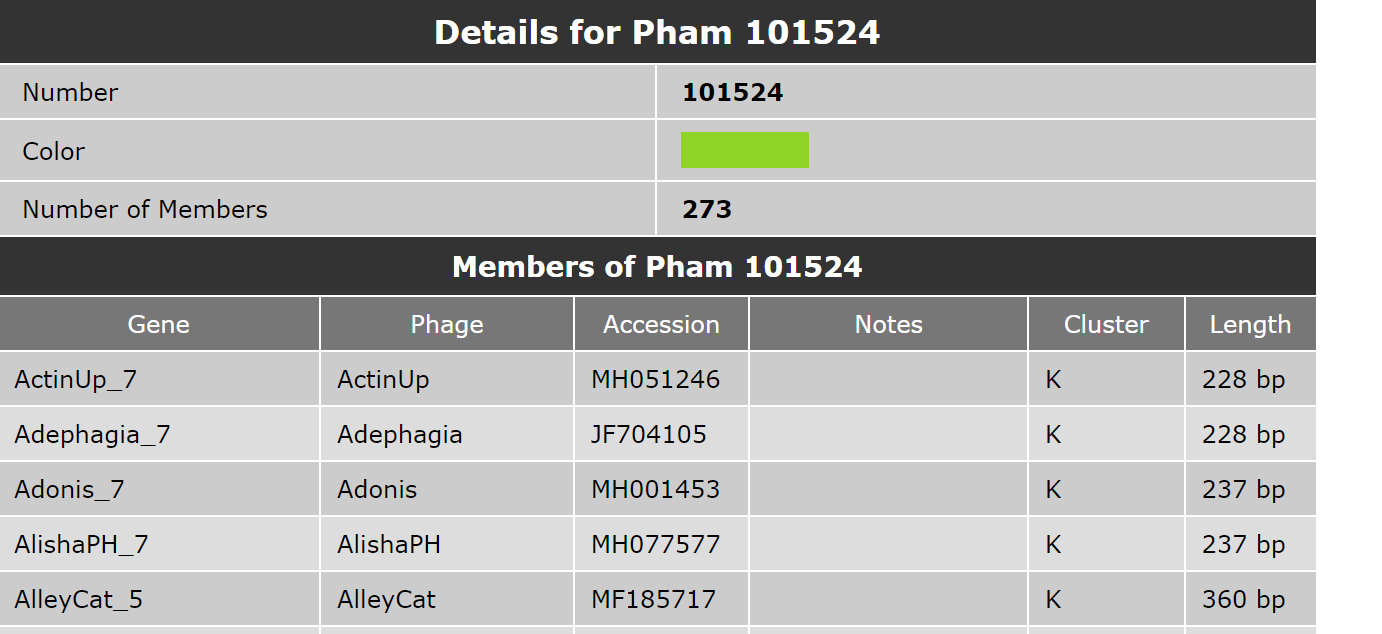
Glimmer predicted nucleotide stop: 387

**1. Is the ORF a gene?**

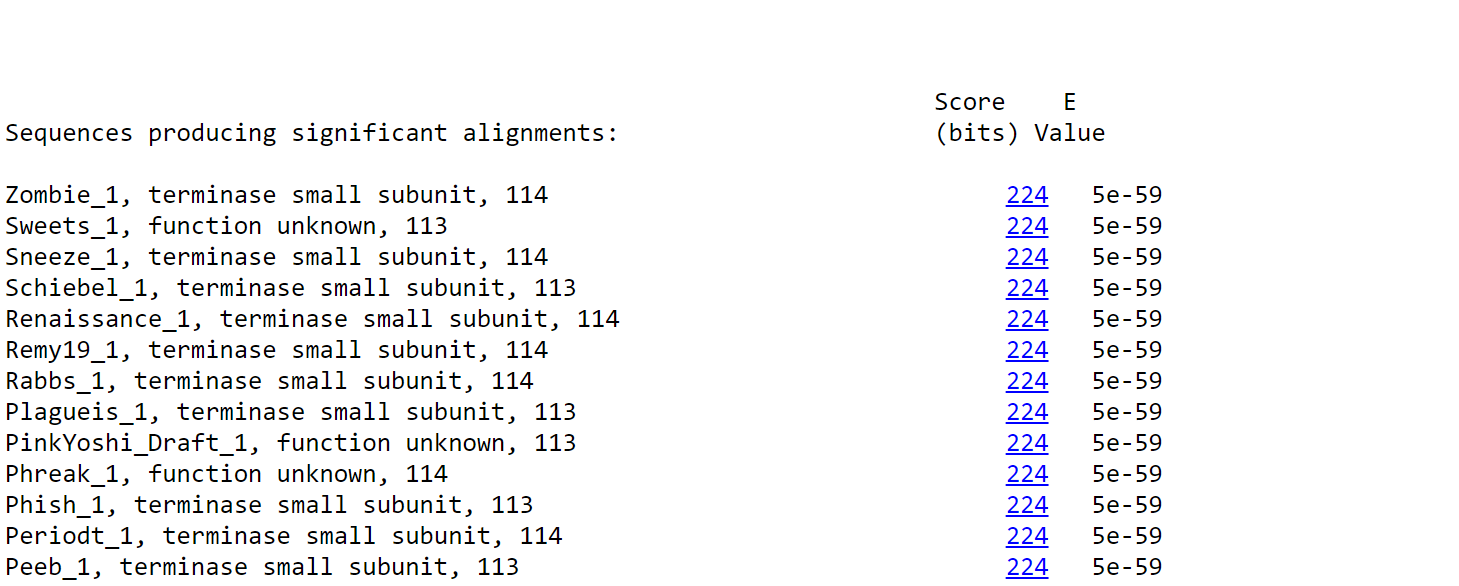
a. Do the Glimmer and GeneMark gene calls in the auto-annotation agree? \_\_No; GeneMark called it at 43.\_\_\_\_\_

b. Does the ORF have high coding potential based on the GeneMark output? \_\_Yes.\_\_\_\_\_



c. Do we see this ORF in phage of the same subcluster using a Phamerator gene map? \_\_Yes.\_\_\_\_\_ If so, what is the Pham number? 

d. Does a protein BLAST (from either phagesdb.org BLAST, or from NCBI BLAST) produce any matched alignments to proteins with well-known functions? \_Yes; terminase, small subunit.\_\_\_\_\_\_



Sweets-2014

Peeb-2018

Phish- annotated 2015

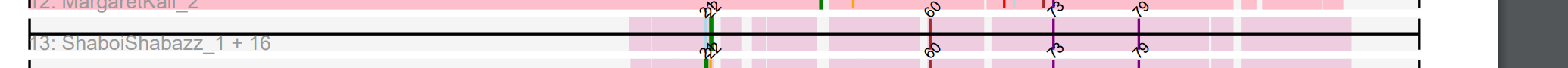
**Conclusion: Is the ORF a gene? Yes.**

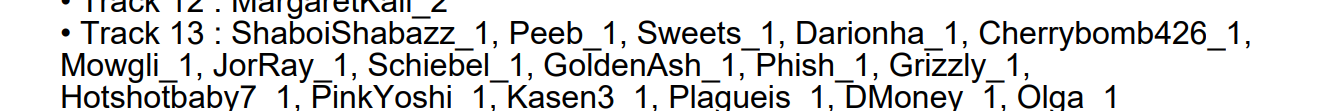
**Rationale for the decision: Called by both Glimmer and GeneMark; high coding potential; synteny; potential function.**

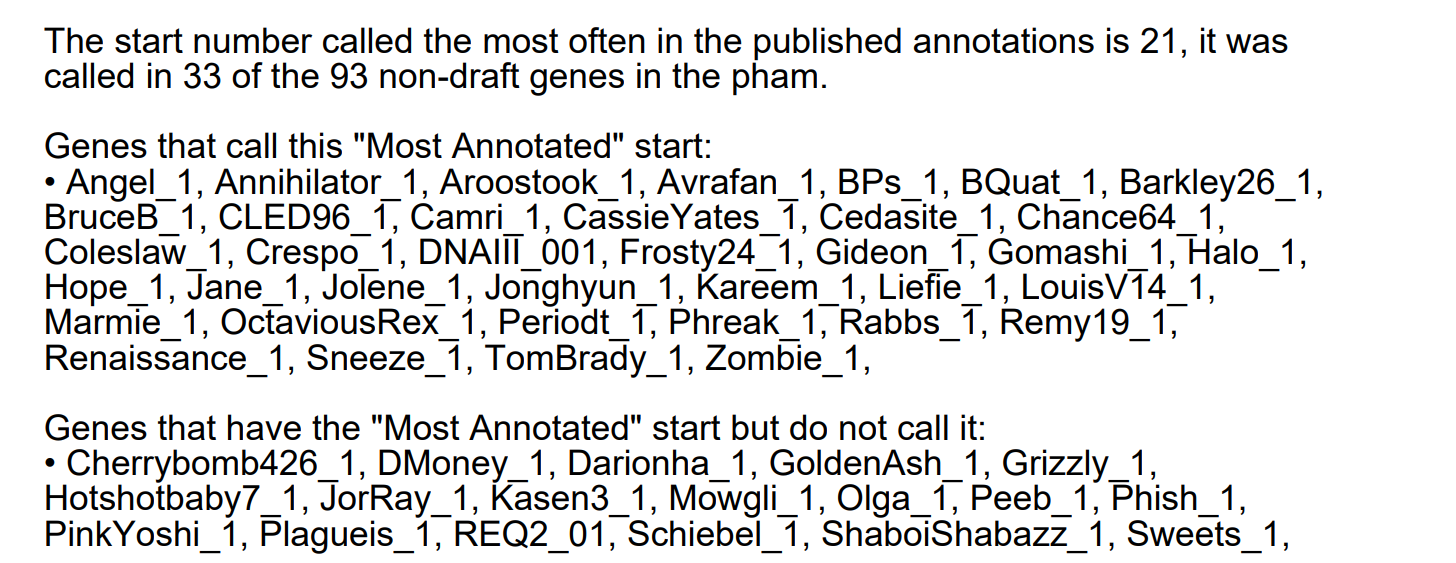
**2.** **If it is a gene, where does it start?**

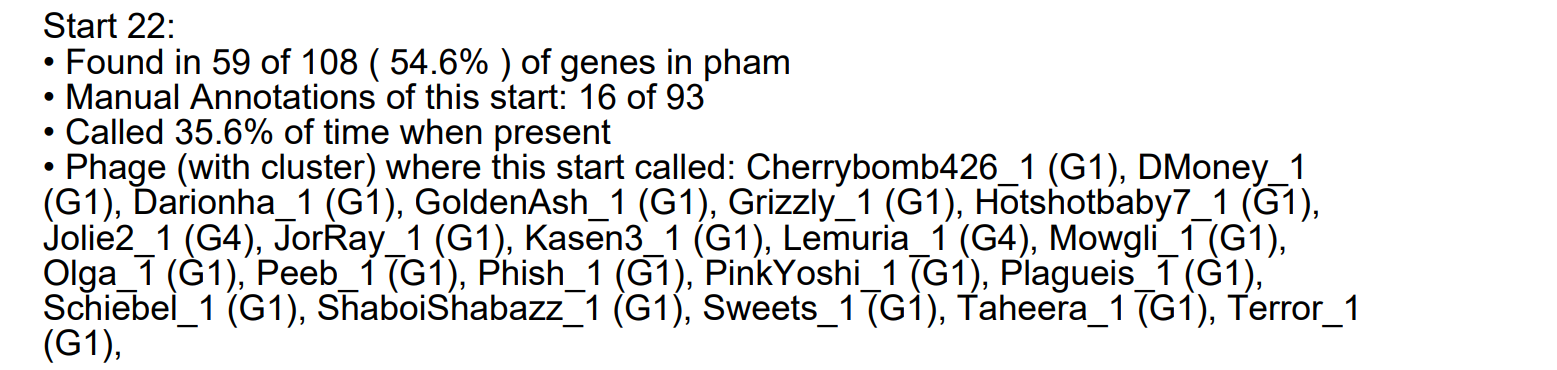
a. Is this start conserved among all members of the Pham in Starterator? \_\_\_No.\_\_\_\_

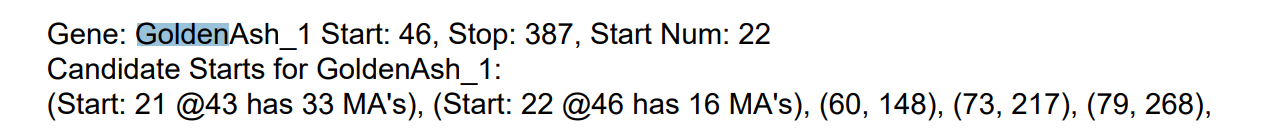
(phagesdb.org/Phage/Phams/enter Phamily number/get starterator report) to evaluate gene starts.





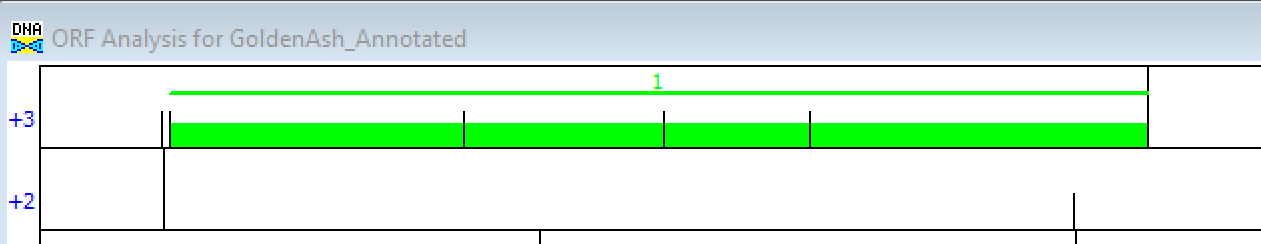


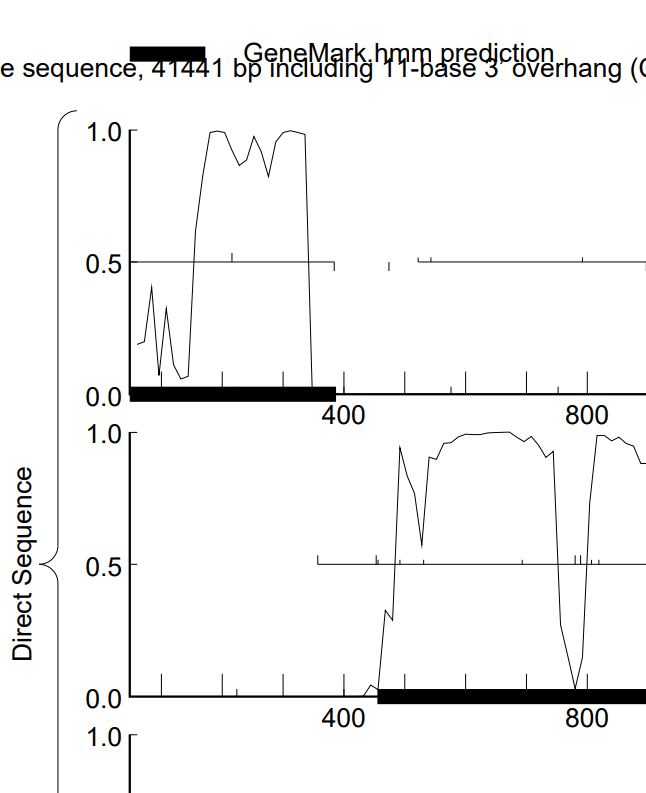




b. Does the currently predicted start site include all of the coding potential in the GeneMark output? Yes.

+3 Frame

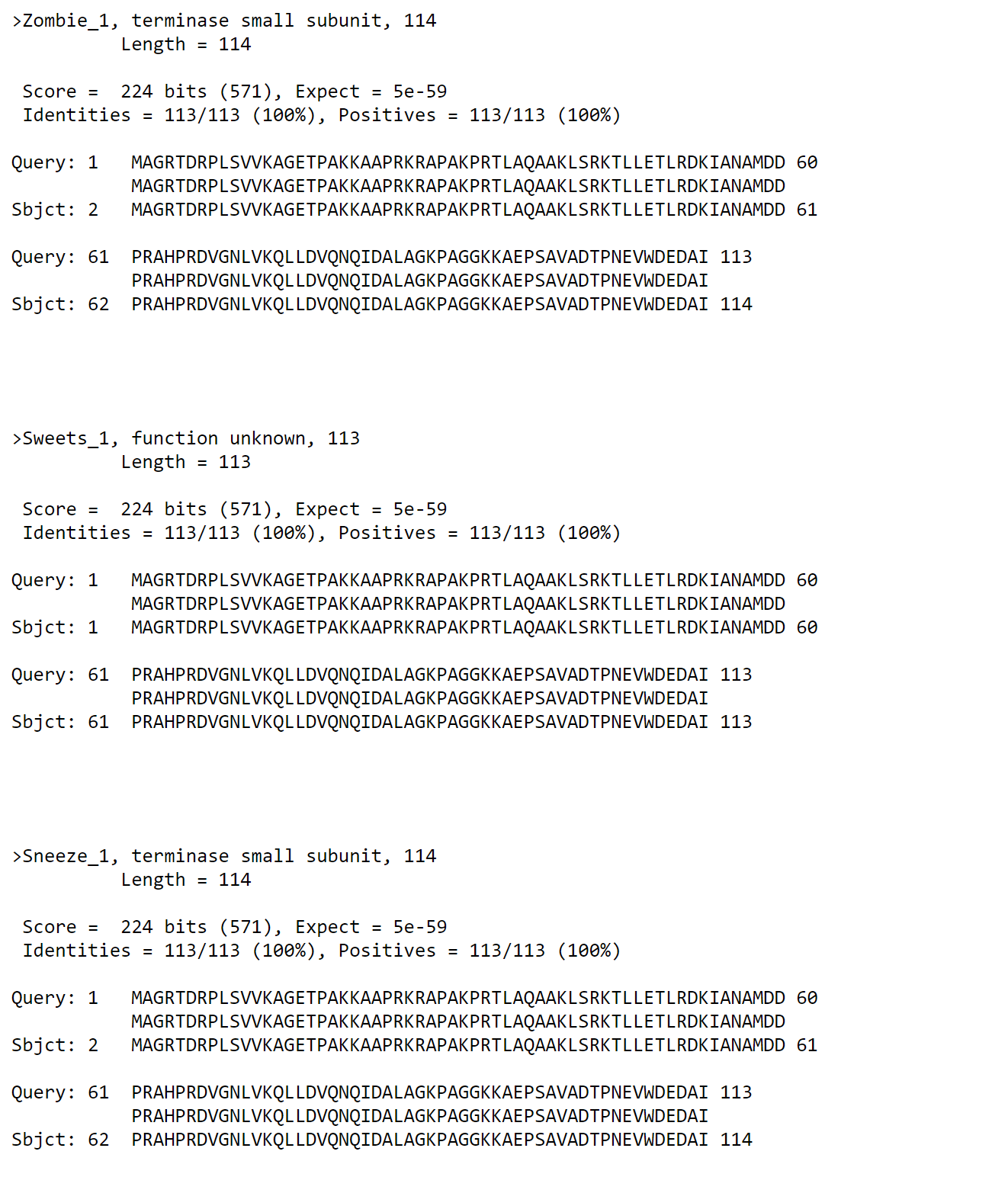




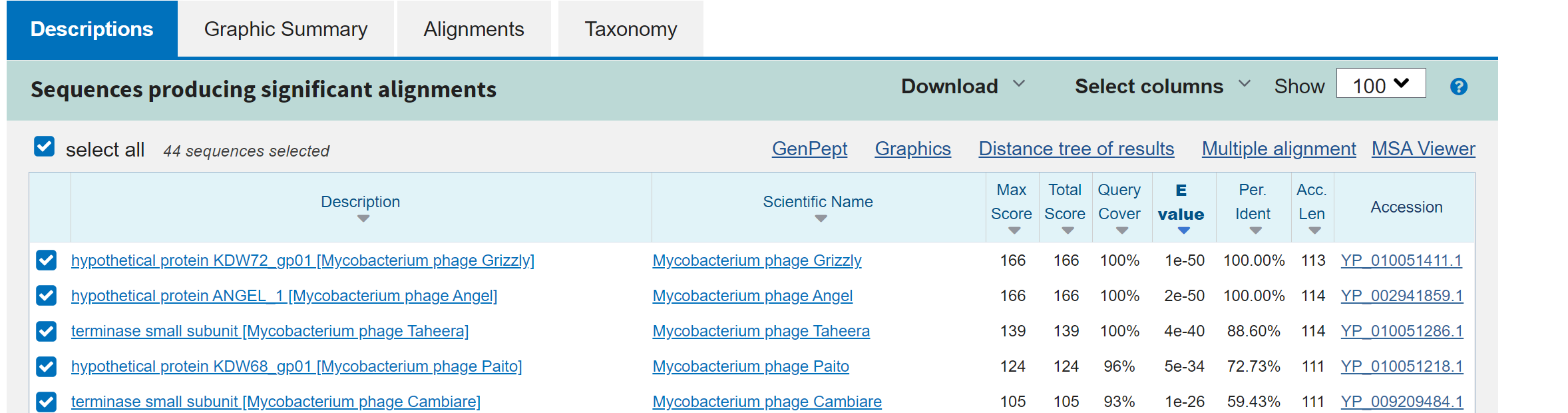
c. Did Glimmer and GeneMark agree on the start? **No: Glimmer-46; GeneMark-43.**

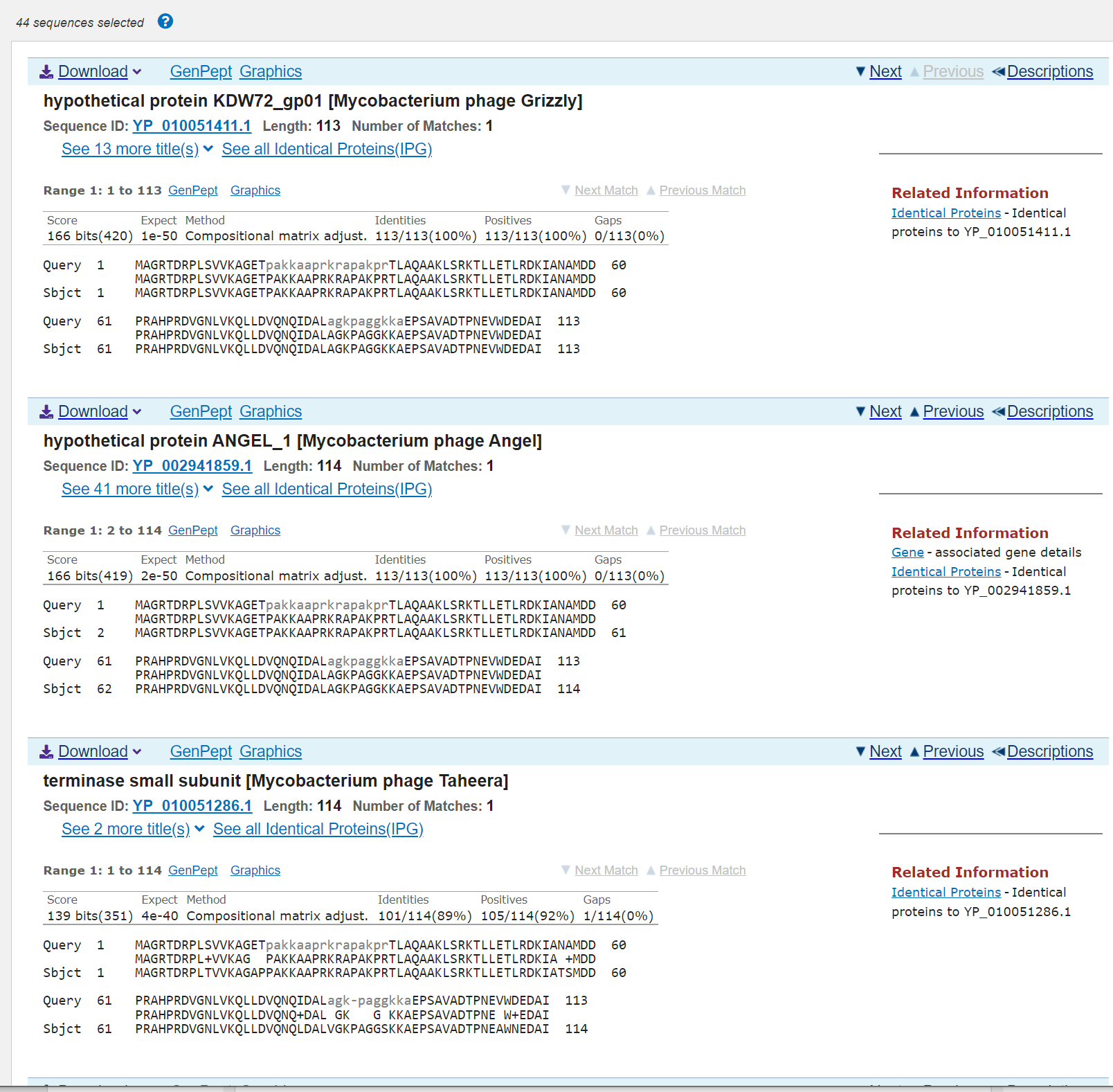
d. Is the predicted start codon the longest possible for the ORF without causing excessive overlaps? Check DNA Master DNA/Frames. **No.**

e. For genes with functional predictions from BLAST, does the start choice include the full-length protein? **Yes, some. But total 113 vs 114 amino acids with some phages like Zombie.**

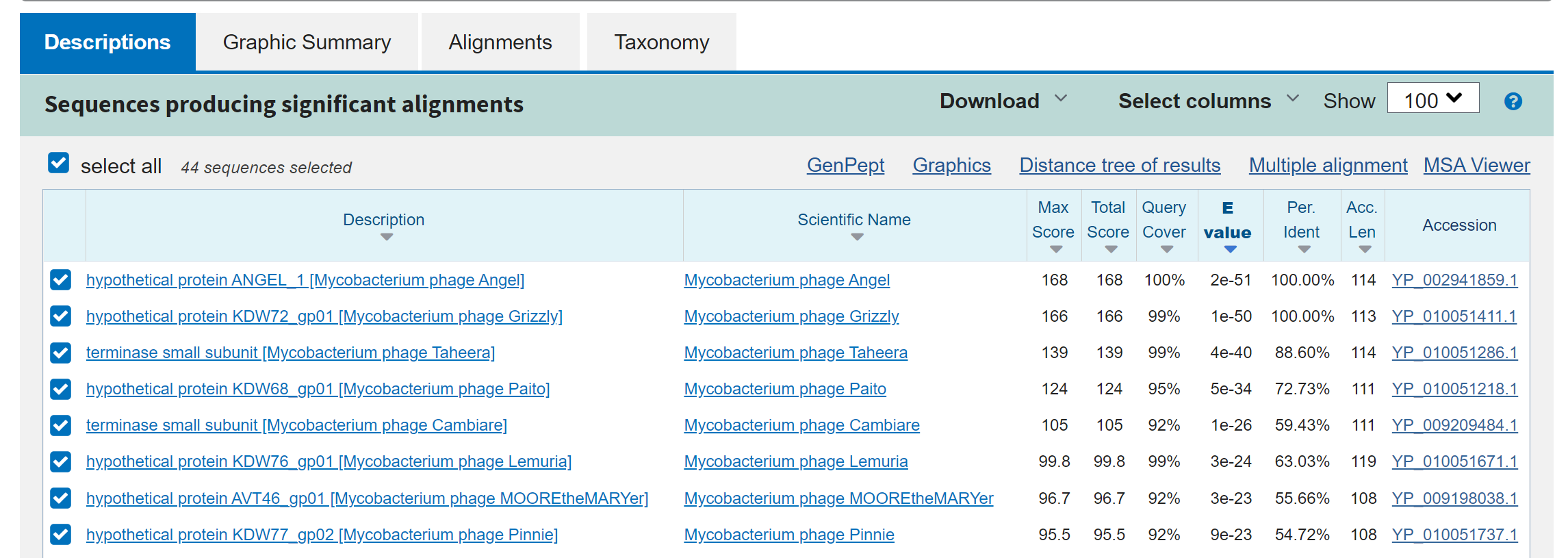


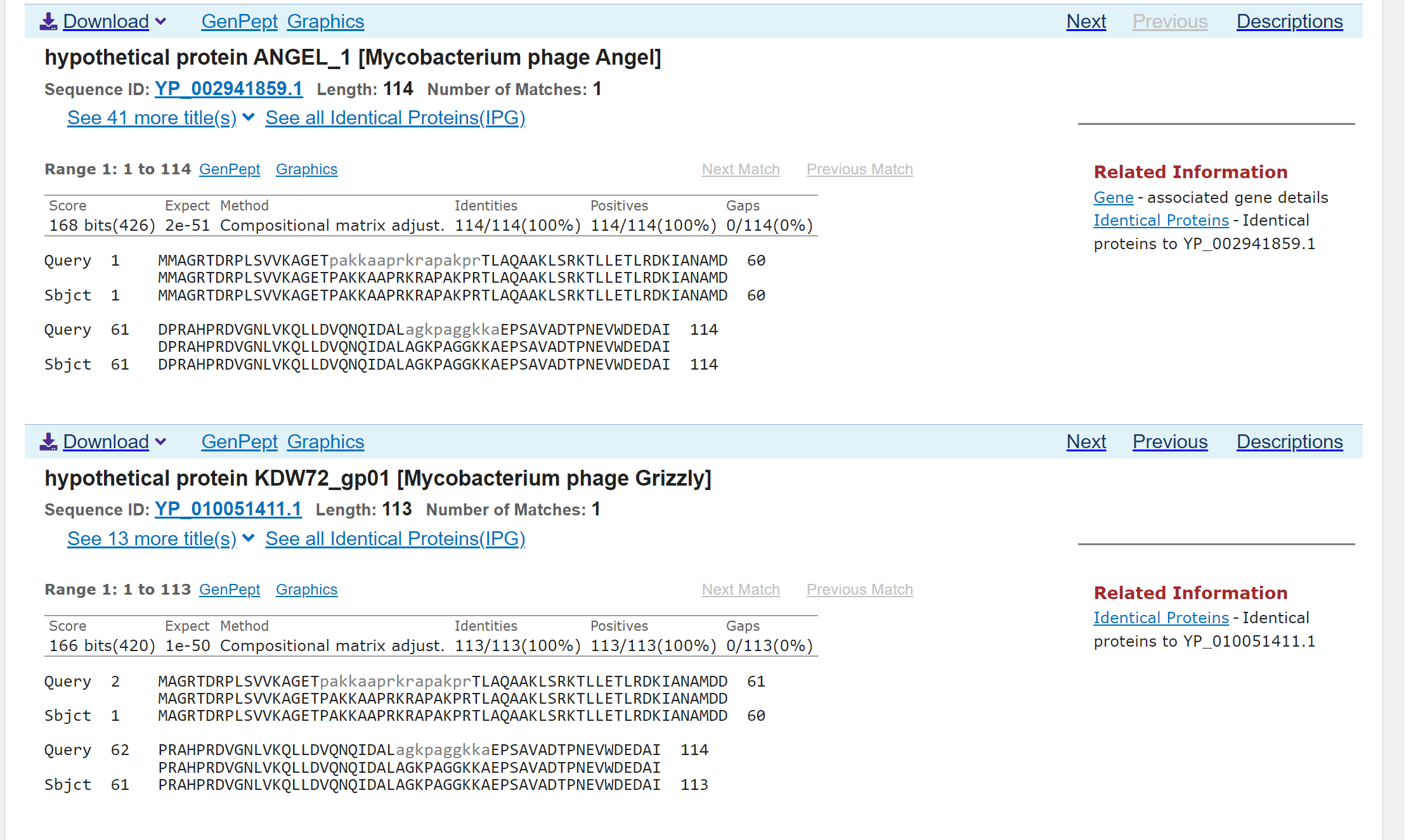
**BLASTP shorter protein:**

****

****

**BLASTP with longer protein starting at BP 43:**

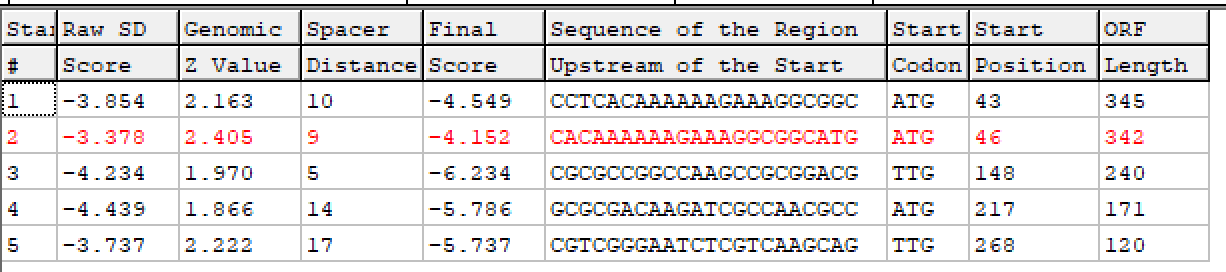




f. Does the start site match other starts for similar genes in NCBI BLAST and the phagesdb.org BLAST? \_\_**Yes.**\_\_\_\_\_

g. Does the predicted start have an associated ribosome binding site (RBS, i.e. Shine-Dalgarno sequence) with a high score or recognizable sequence? **\_\_\_\_Yes\_\_\_**

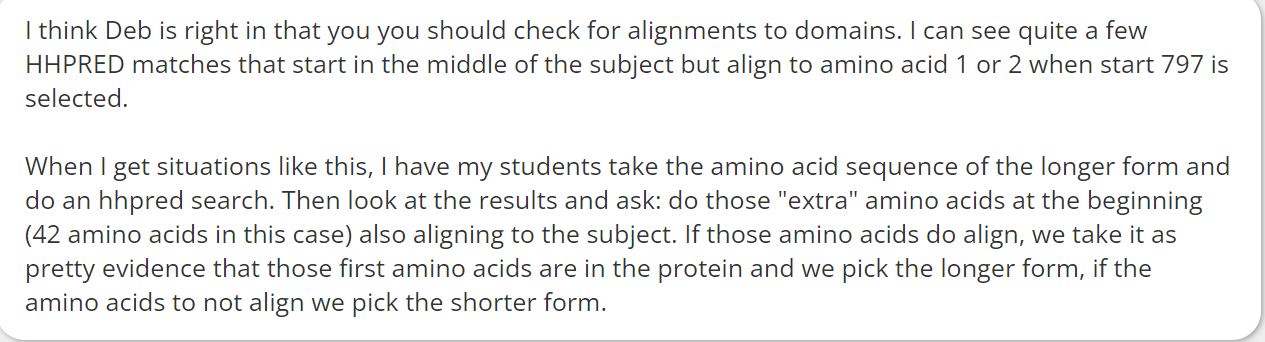
Use RBS Sequence Finder in DNA Master.



**Conclusion: At which nucleotide does the gene start? \_\_43\_\_\_\_\_**

**Rationale for the decision:**

Starterator-no consensus, 59 vs. 69 phages with this start/273 total phages, most annotated start is much shorter; contains full coding potential; Glimmer favored over GeneMark; not longest ORF by 3bp; matches functional products of other phages, but does not include 1AA with some phages; high RBS score; HHPred with longer protein doesn’t align any better, actually slightly LOWER probability match (see below).

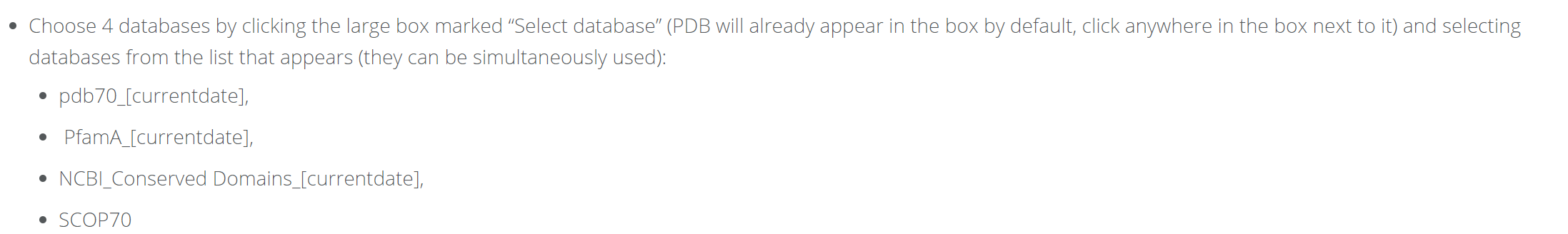


**3. What is the gene's function?**

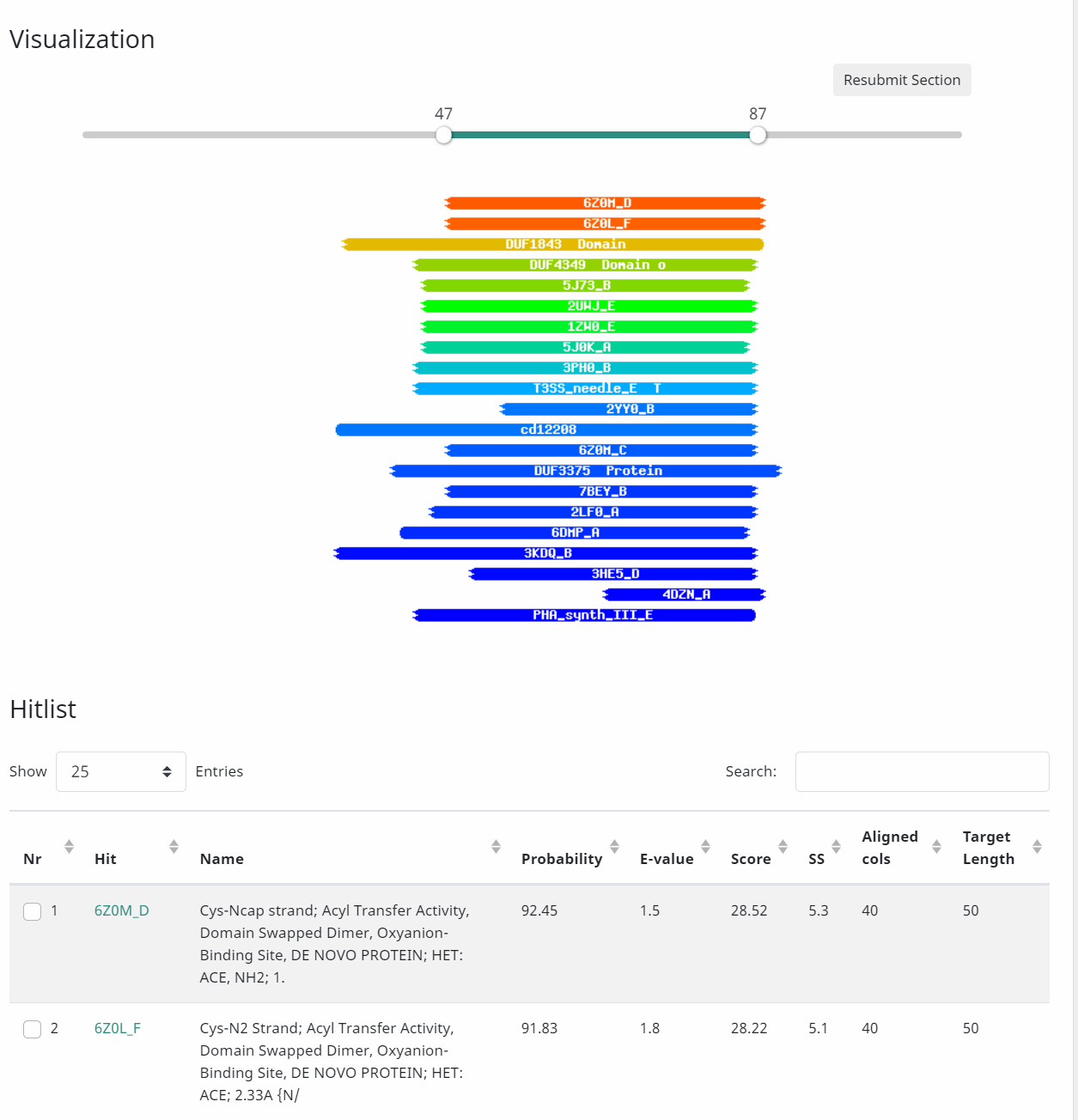
a. BLASTP: Does this protein align with a protein with a functional assignment in NCBI or phagesdb.org? \_\_\_Yes with some; terminase, small subunit (see above)/hypothetical.\_\_\_\_

b. HHPred: Does this protein align with a protein with a functional assignment in the PDB? \_\_Yes, has acyl transfer activity.\_\_\_\_\_

Use <https://toolkit.tuebingen.mpg.de/#/tools/hhpred> and paste protein sequence into the box.

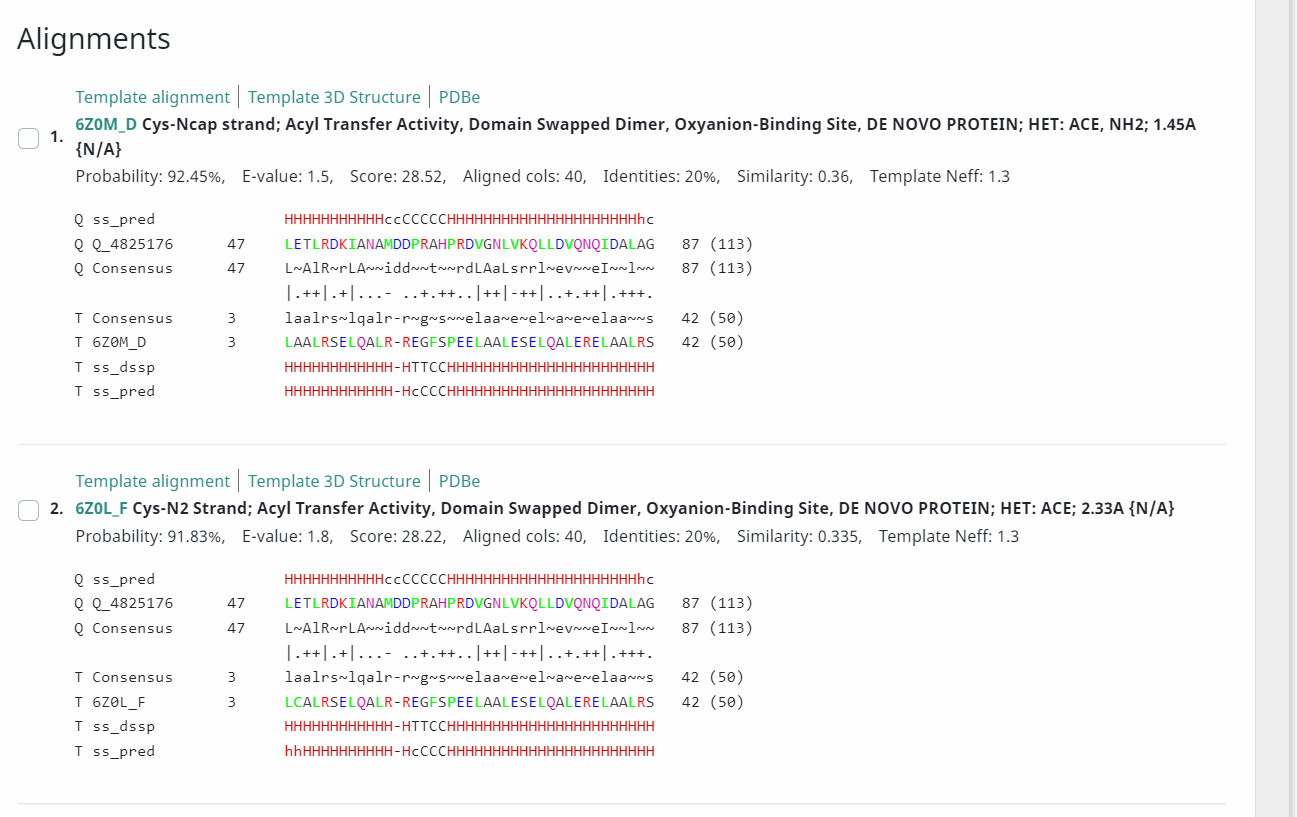


**Shorter protein:**



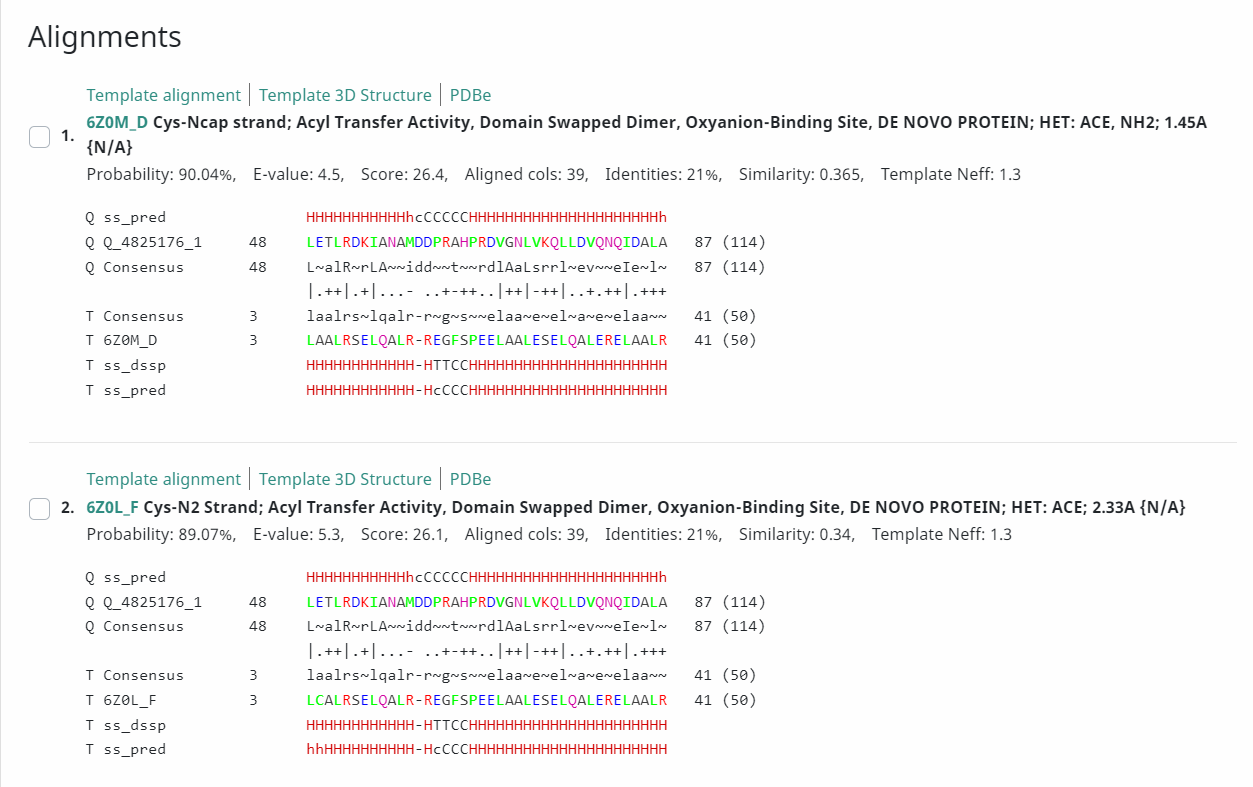
Tailed DNA bacteriophages assemble empty procapsids that are subsequently filled with the viral genome by means of a DNA packaging machine situated at a special fivefold vertex. The packaging machine consists of a “small terminase” and a “large terminase” component. One of the functions of the small terminase is to initiate packaging of the viral genome, whereas the large terminase is responsible for the ATP-powered translocation of DNA. The small terminase subunit has three domains, an N-terminal DNA-binding domain, a central oligomerization domain, and a C-terminal domain for interacting with the large terminase.

***The*small terminase subunit*is thought to form a nucleoprotein structure that helps to position the*terminase*large*subunit*at the packaging initiation site.***



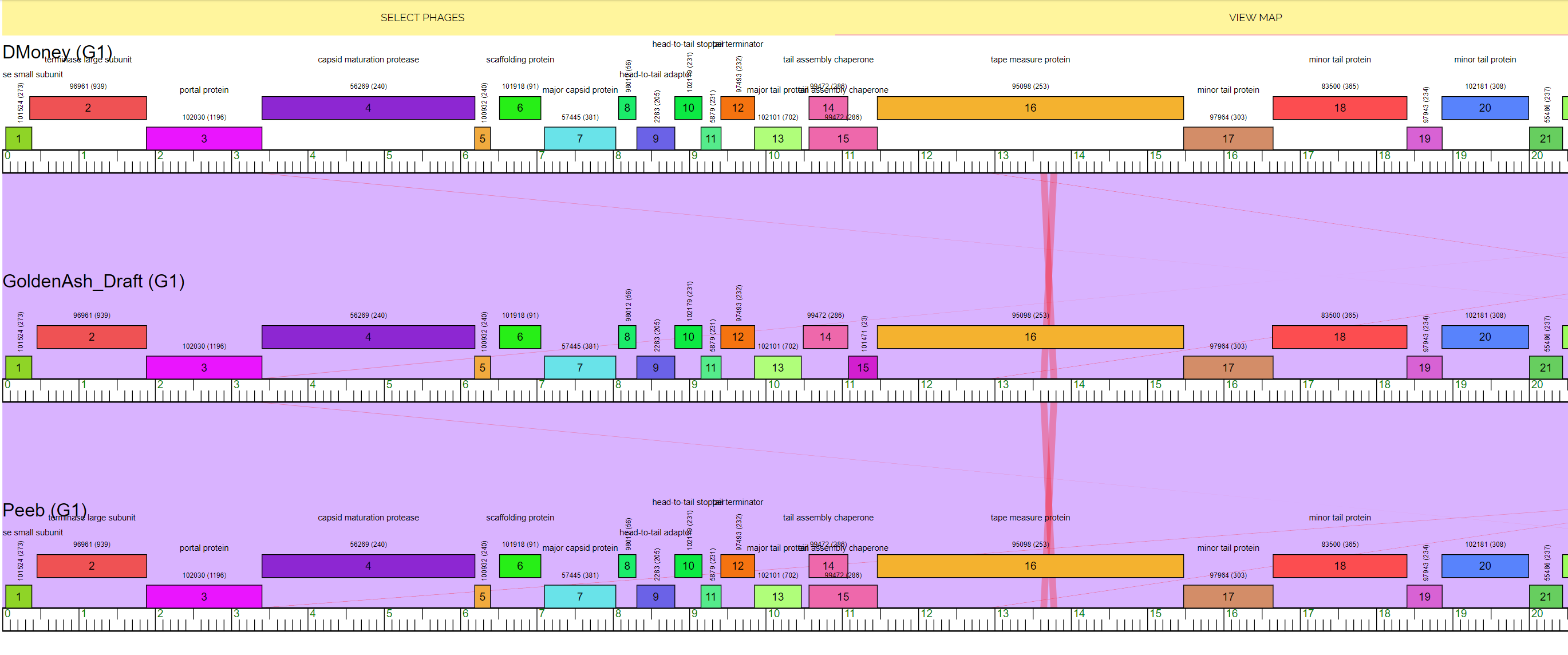
**Longer protein:**





c. Synteny: Is this gene located adjacent to other genes of known function, in a region of the genome that shows high conservation of gene order? \_\_\_Yes; terminase/terminase, large subunit.\_\_\_\_

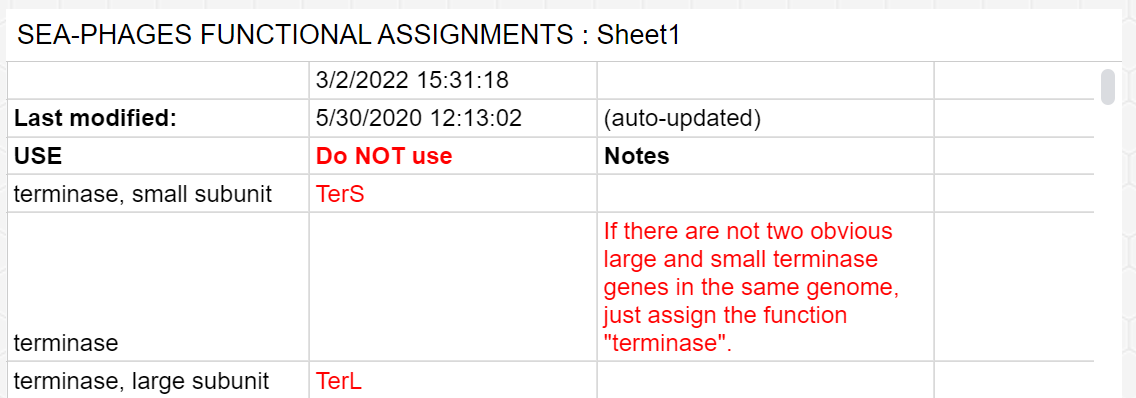
Use Phamerator of closely related phages.



d. Official function from the current SEA-PHAGES FUNCTIONAL ASSIGNMENTS list found at https://seaphagesbioinformatics.helpdocsonline.com/article-96.

**Conclusion: What is the gene function? Terminase, small subunit.**

**Rationale for the decision: HHPred hit for Acyl Transfer activity; evidence and alignment with other phages; gene 2 HHPred hit for Terminase, large subunit.**



**4. Annotation Notes.**

This is the short-hand code list in the Features/Notes section of DNA Master for each gene.

SSC: CP: SCS: ST: BLAST-Start: Gap: LO: RBS: F: SIF-BLAST: SIF-HHPred: SIF-Syn

The instructions how to record the notes are found at <https://seaphagesbioinformatics.helpdocsonline.com/officialdocumentation>

Generally, DNA packaging machines consist of three components ([3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B3), [7](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B7)). The first component is a dodecameric portal protein located at the special fivefold vertex of the capsid through which the DNA is threaded into the head. Crystal structures of portal proteins from phages φ29 ([8](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B8)), SPP1 ([9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B9)), and P22 ([10](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B10)) show that they form cone-shaped structures with a central cylindrical channel about 36 Å wide. Cryoelectron microscopy (cryo-EM) studies of phages φ29 ([8](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B8)), SPP1 ([11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B11)), T4 ([12](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B12)), P22 ([13](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B13), [14](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B14)), and ϵ15 ([15](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B15)) show that the wider end of the cone-shaped portal is inside the capsid, and the narrower end protrudes out of the capsid. The portal provides a site of attachment for the DNA packaging motor to the procapsid and for the tail to the filled capsid. To what extent the portal participates in DNA packaging is not clear, but it might act as a valve to stop the DNA from escaping the head during successive strokes of the packaging motor ([16](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B16)) and when completely packaged ([17](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B17)). It was also proposed that the portal might be involved in sensing when the head is fully packaged ([18](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B18), [19](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B19)).

The second component of the DNA packaging machine is the large “terminase” motor protein which has both an ATPase activity to provide energy for packaging and a nuclease activity for packaging initiation and termination ([20](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B20)). In most DNA phages, the newly replicated genome is a branched concatemer without any accessible free ends ([3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B3)). It needs to be cleaved in order to create a free end to initiate packaging. After packaging is complete, the genomic DNA is again cleaved to terminate packaging, and the remaining DNA is transferred to another empty procapsid. The crystal structure of T4 large terminase, gene product 17 (gp17), shows that its N-terminal domain has the conserved nucleotide-binding fold found in many ATPases and the C-terminal domain belongs to the RNase H/resolvase/integrase superfamily ([21](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B21), [22](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B22)). These two domains are in close contact with each other in the crystal structure, representing a “tensed” conformation ([21](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B21)). Cryo-EM studies of the T4 packaging motor show that gp17 forms a pentamer on the procapsid on the outside of the portal. However, the N- and C-terminal domains are spatially separated, forming a “relaxed” conformation. Sun et al. ([21](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B21)) proposed that gp17 alternates between the relaxed and tensed conformations while packaging the genomic DNA in a piston-like fashion.

The third component of the DNA packaging machine is a small oligomeric protein (“small terminase”) that is essential for initiating packaging ([3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B3)). The small terminase recognizes viral DNA and brings it to the large terminase for the initial cleavage. The DNA-binding aspects of the small terminase in phages λ (gpNu1) and SPP1 (G1P) have been well-characterized ([23](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B23), [24](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B24)). These proteins bind to specific sequences in their genomes (cos and pac sites, respectively) from where packaging is initiated. The T4 phage packages 1.02 genome lengths of DNA (approximately 171 kb) into each procapsid before a “headful” signal is sent to cleave the DNA and disengage the packaging motor ([25](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B25)). In contrast, there is no unique pac site in the T4 genome. The T4 small terminase, gp16, probably binds only weakly to DNA ([26](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B26)). It is not clear whether small terminases actively participate in the packaging process. Although small terminases stimulate the large terminases’ ATPase activity in the absence of the other packaging components, most in vitro packaging systems (T4, T3, and λ) do not require small terminases for packaging precut DNAs ([27](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B27)–[29](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B29)). Moreover, in a “defined” in vitro T4 packaging system, consisting of procapsids, gp17, DNA, and ATP, the addition of gp16 inhibits packaging ([30](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B30), [31](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271864/#B31)).

**SEA-PHAGES Forum Question:**

Hi we have a question concerning the start for a cluster G1 phage, GoldenAsh\_1:

* Glimmer calls the start at 46; GeneMark at 43. Hard to see exactly where the coding potential begins on the GeneMarkS readout.
* The Starterator report does not show consensus; the most called start is at bp 43, but only in 33/93 annotated genomes.
* BLASTP shows alignment from the beginning of the protein with phages containing 113 amino acids; some phages have 114 amino acids, however. Likely function: terminase, small subunit.
* HHPred has a slightly higher probability with the shorter protein starting at bp 43 (92% vs. 90%).

The evidence doesn’t show clear support for the longer ORF/protein, so we were going to stick with the original Glimmer call at bp 46. Any thoughts would be appreciated!