

Annotation of the *Nanoarchaeum equitans* Kin-4M Genome at Locus Tags NEQ348 and NEQ349

Maxwell Sun, Elizabeth Lindahl and Peter Hentschke

The Harley School, Brighton, NY and the Western New York Genetics in Research and Health Care Partnership

Abstract

Two genes from the microorganism *Nanoarchaeum equitans* (NEQ348 and NEQ349) were annotated using the collaborative genome annotation website GENI-ACT. GENI-ACT did not provide a gene product prediction. The genes were assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. After analysis, the research concluded in a proposition for a gene product annotation for these genes.

Introduction

The species *Nanoarchaeum equitans* is the only species of the Nanoarchaeota that has been cultured. Organisms of this species are found associated with host bacteria of the species *Ignicoccus hospitalis*. And, physical contact between the two is crucial for *N. equitans* (Giannone et al., 2011). The two hyperthermophiles are the only known example of association between two Archaea (Podar et al., 2008). This relationship can not be confidently identified as symbiosis, commensalisms, or parasitism, as it has characteristics of all three. In addition, *N. equitans*, which are tiny cocci 350 to 500 nm in diameter, has a surprisingly small genome. In fact, it lacks most of the known genes for biosynthesis of cofactors, amino acids, lipids, and nucleotides. Furthermore, the species derives its membrane lipids from *Ignicoccus hospitalis*, one reason for the heavy dependence it has on its host (Jahn et al., 2007).

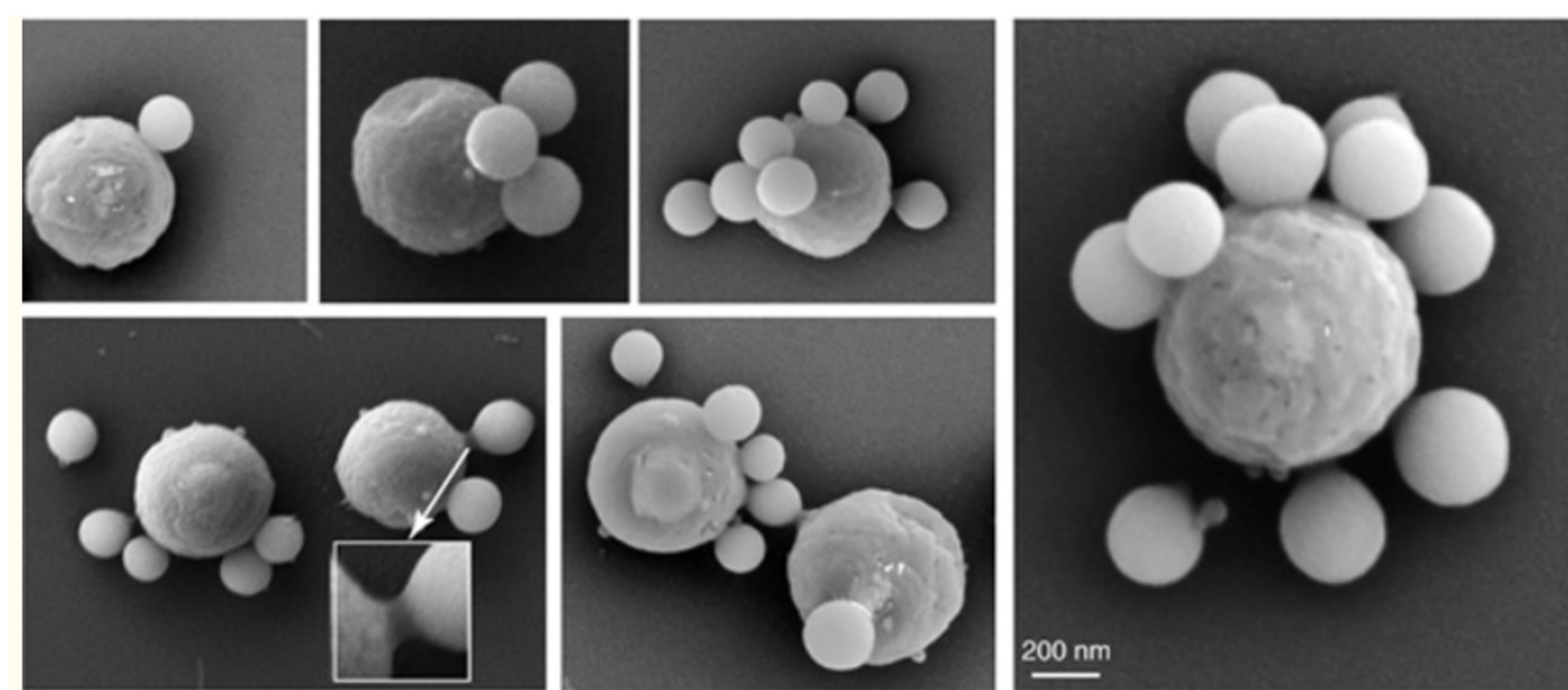


Figure 1: *N. equitans* cells colonizing on the larger host cell of *I. hospitalis*. The former fuses to the host cell (Giannone et al., 2015).

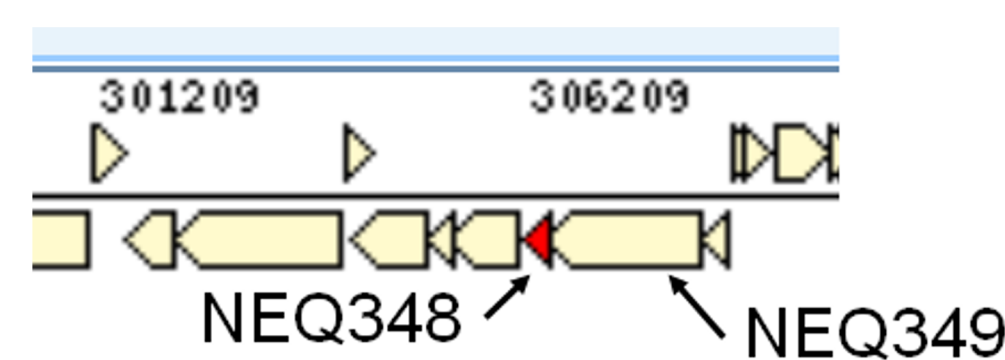


Figure 2. The locus tags and relative position of the genes under investigation in this research.

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to annotate genes of the *Nanoarchaeum equitans* Kin4-M genome. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

NEQ348:

The gene with locus tag NEQ348 is found in the DNA coordinates complement(306082..30630). It is a sequence of 228 nucleotides resulting in a 75 amino acid polypeptide. The top nr BLAST hit identified the gene product as an archaeal histone. The T-Coffee alignment was very well conserved, an indicator that the nr hits are similar to the desired gene product. The cellular location of the product was predicted to be in the cytoplasm. TMHMM found no transmembrane helices and SIGNALP did not detect a signal peptide. PSORTb and Phobius confirmed these results.

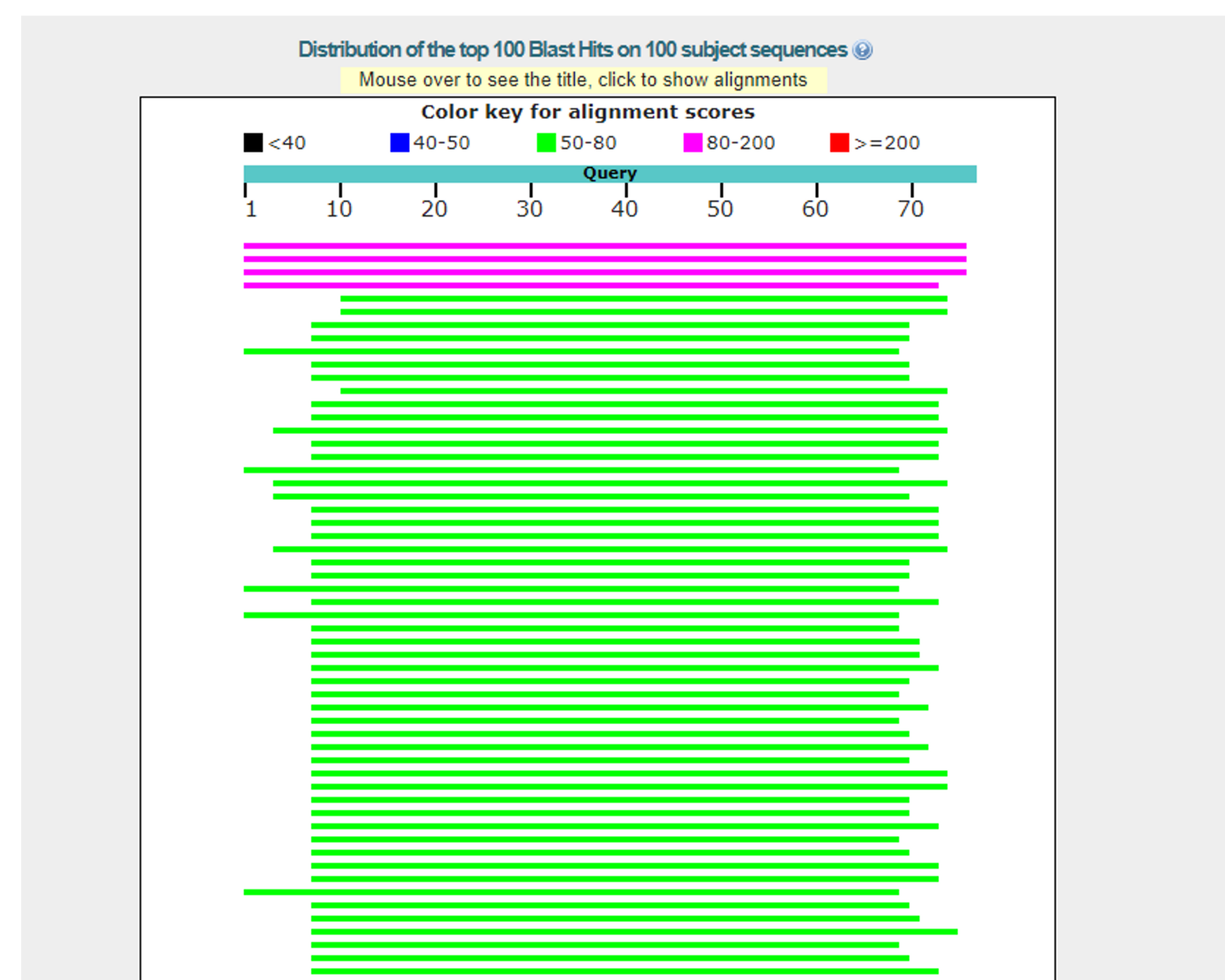


Figure 3. BLAST hits for NEQ348 taken over the nr database. There were a few medium-high scores for the top alignments shown in magenta but the vast majority of the matching alignments in green were scores of 50-80.



Figure 4. The T-Coffee multiple sequence alignment for NEQ348 showed that it was highly conserved except for the gaps in the beginning and end of the sequences.

NEQ349:

The gene with locus tag NEQ349 is located on the complement strand of the DNA and is found in the DNA coordinates 306335 through 308137. The gene consists of 1803 nucleotides and codes for a protein of 600 amino acids. After investigation with BLAST, CDD, TIGRFAM, Pfam, and PDB, it was found that all of the documented top hits that came up were of the Lon Protease family of ATP-dependent proteins or at least heavily related. One BLAST hit from the nr database was, in fact, an ATP-dependent LonB protease, very similar to the proposed product by IMG/M. The PDB hit was especially exciting.

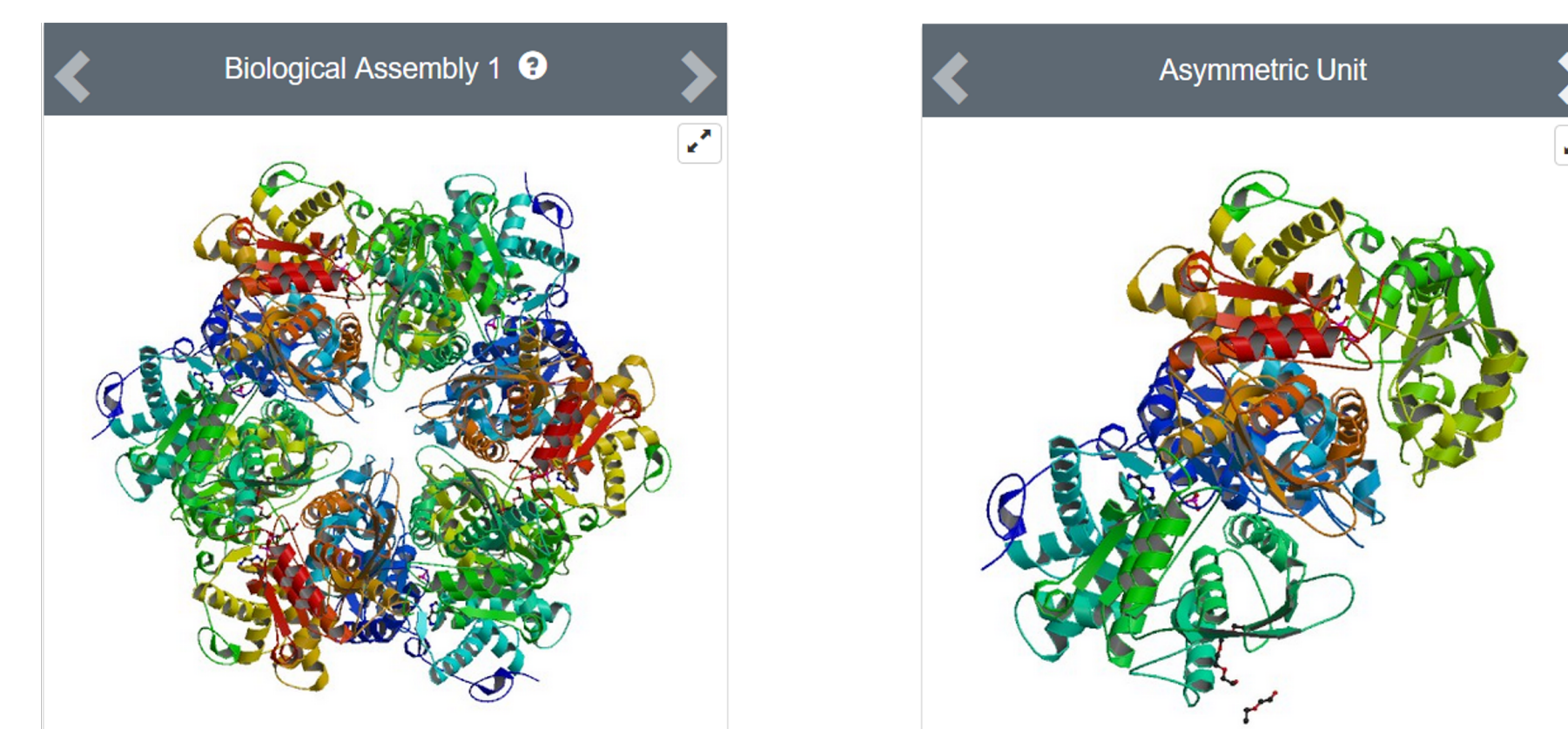


Figure 4. Crystal structure of Lon protease from *Thermococcus onnurineus* NA1. Notice that three of the asymmetric units make up the full biological assembly. This was the top PDB hit with score 572.007 bits.

In addition, it was predicted by the Psortb software that the protein was cytoplasmic with a strong localization score of 9.96 for the cytoplasm. And, two transmembrane helices were predicted by TMHMM and Phobius, yet this number is not high enough for the protein to be predicted a membrane protein. Lastly, the protein was predicted to have no signal peptide by SignalP.

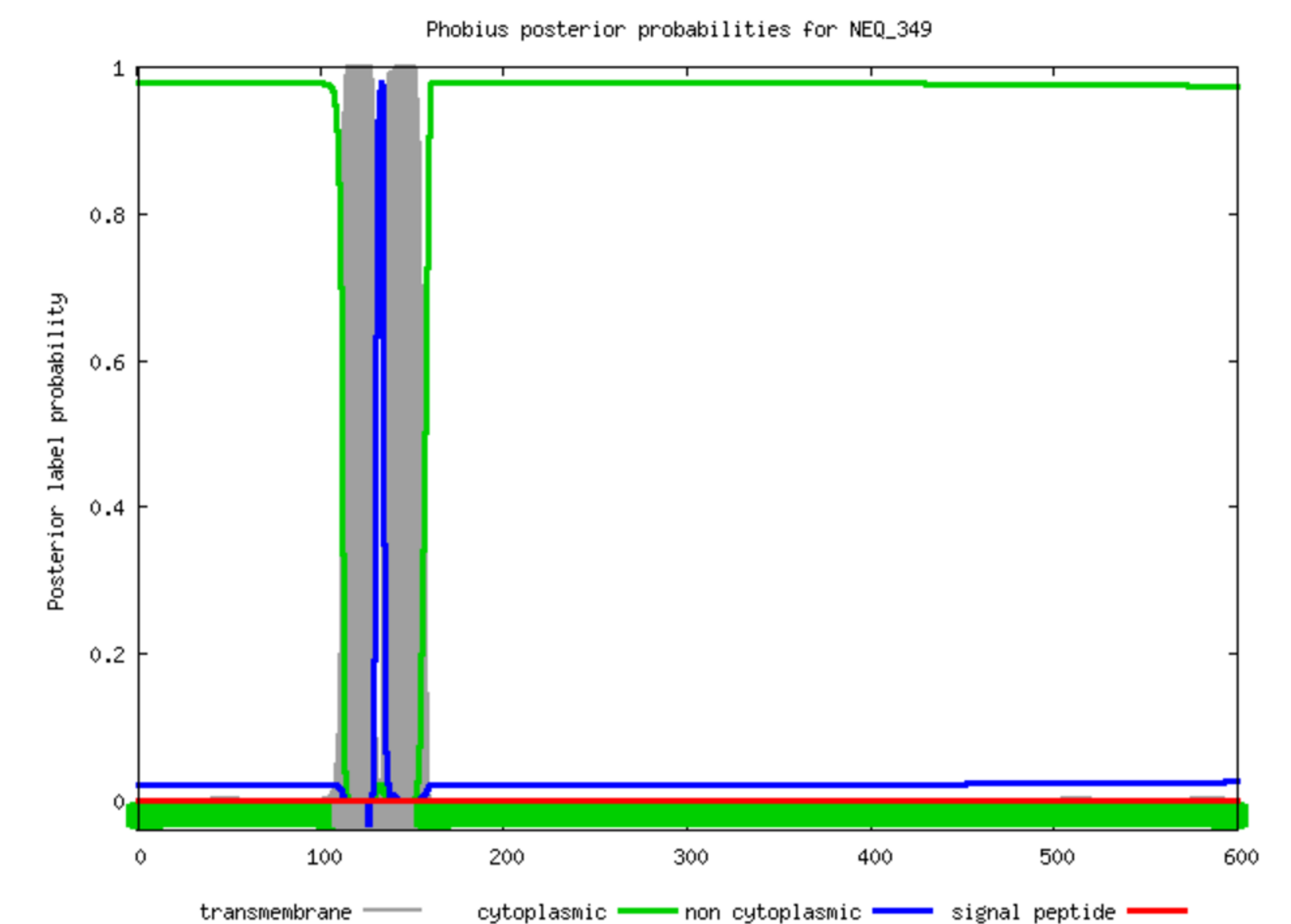


Figure 5. Phobius results. There are predicted to be two very-close-together TMHs and no signal peptide. The protein is not a membrane protein, so these TMHs must be in some special position.

Conclusion

Based on the analysis done in the research, the gene products proposed for these genes, NEQ348 and NEQ349, are shown below.

Locus Tag	Proposed Annotation
NEQ348	archaeal histone
NEQ349	ATP-dependent Lon protease

References

- Giannone et al., 2011. Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans* - *Ignicoccus hospitalis* relationship. PLoS One: 6(8):e22942. doi:10.1371/journal.pone.0022942
- Jahn et al., 2007. *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. J Bacteriol. 2007;190(5):1743–1750. doi:10.1128/JB.01731-07
- Podar et al., 2008. A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*. Genome Biol. 2008;9(11):R158. doi:10.1186/gb-2008-9-11-r158Sims et al. (2009).

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209