2018

2018 - Could Archaea Be an Important Key to Unlocking the Mysteries of Progeria? - Poster Presentation

Ashley Platt

Could Archaea Be an Important Key to Unlocking the Mysteries of Progeria?

Ashley Platt
Salt Lake Community College, Utah USA

Abstract

Although the genome for the archaeon *Halogeometricum borinquense* has been fully sequenced, not a lot of time has been spent determining which sequences represent protein-coding genes. I mapped the genome from nucleotides 900,000-950,000 to determine the position of potential genes using an Open Reading Frame Finder (ORF) program and found 55 ORFs that potentially code for proteins. I then annotated 4 of these genes to deduce their function. To test my hypothesis that an identified ORF codes for an actual protein, I compared ORF sequences in *Halogeometricum borinquense* with known protein sequences in other organisms using a variety of Bioinformatic programs.

To verify whether or not the gene HBOR_RS04875 codes for a protein, I looked for homology with other proteins. I used a BLAST protein search to compare the amino acid sequence of the potential genes to sequences present in other organisms. In addition, finding significant homology allows me to begin to ascertain the potential function of a proposed gene by identifying homologous regions of known function in other organisms. Due to common ancestry, similar DNA sequences can be found in very different organisms.

I discovered HBOR_RS04875 has a domain named M48A_Zmpste24p_like. A protein with this domain is considered a type 1 CaaX endopeptidase from the Peptidase M48 subfamily A. This protein has a domain homologous with an enzyme in humans called ZMPSTE24. Mutations in ZMPSTE24 or in the protein that codes for its substrate (prelamin A) result in laminopathies, one of the most severe being Hutchinson Gilford progeria syndrome (HGPS). Because ZMPSTE24 interacts with nuclear lamina, and archaea do not contain nuclei or appear to contain intermediate filaments, I propose a rescue experiment to see if the archaeal gene can replace the eukaryotic analog in yeast. If so, this could lead to further understanding of the complex interaction between DNA and lamins.

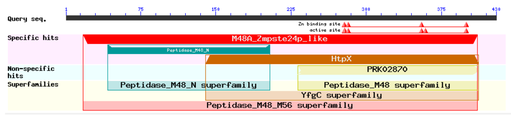
Methodologies

Multiple bioinformatic programs used

BLAST: Basic Local Alignment Search Tool

The information from BLAST indicates that HBOR_RS04875 has similar sequences to proteins from many other organisms.

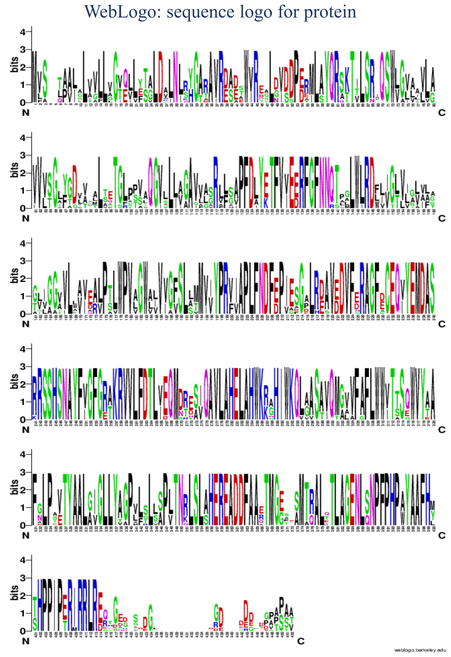
CDD: Conserved Domains



The information from CDD shows a conserved domain found in HBOR_RS04875 named M48A_Zmpste24p_like.

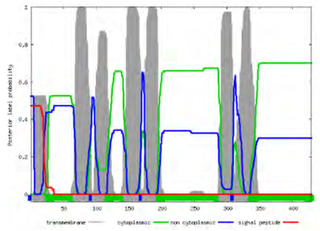
T-Coffee: Tree-based Consistency Objective Function For alignmEnt Evaluation

WebLogo: sequence logo for protein



The programs T-Coffee and WebLogo show sequence alignments of HBOR_RS04875 with those in other organisms, showing the significant amount of homology present.

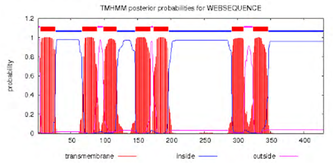
Phobius: Prediction of transmembrane helices and signal peptides in proteins



The information from Phobius shows the potential location of a signal sequence at the N-terminus end and the number of transmembrane helices in HBOR_RS04875.

TMHMM: Prediction of transmembrane helices in proteins

```
# WEBSEQUENCE Length: 429
# WEBSEQUENCE Number of predicted TMHs:  7
# WEBSEQUENCE Exp number of AAs in TMHs: 151.99894
# WEBSEQUENCE Exp number, first 60 AAs:  22.51021
# WEBSEQUENCE Total prob of N-in:        0.01869
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE     TMHMM2.0        outside     1     3
WEBSEQUENCE     TMHMM2.0        TMhelix     4    26
WEBSEQUENCE     TMHMM2.0        inside     27    66
WEBSEQUENCE     TMHMM2.0        TMhelix    67    89
WEBSEQUENCE     TMHMM2.0        outside    90    98
WEBSEQUENCE     TMHMM2.0        TMhelix    99   118
WEBSEQUENCE     TMHMM2.0        inside    119   146
WEBSEQUENCE     TMHMM2.0        TMhelix   147   169
WEBSEQUENCE     TMHMM2.0        outside   170   173
WEBSEQUENCE     TMHMM2.0        TMhelix   174   196
WEBSEQUENCE     TMHMM2.0        inside    197   291
WEBSEQUENCE     TMHMM2.0        TMhelix   292   309
WEBSEQUENCE     TMHMM2.0        outside   310   323
WEBSEQUENCE     TMHMM2.0        TMhelix   324   346
WEBSEQUENCE     TMHMM2.0        inside    347   429
```



The information from TMHMM shows HBOR_RS04875 crosses the plasma membrane 7 times.

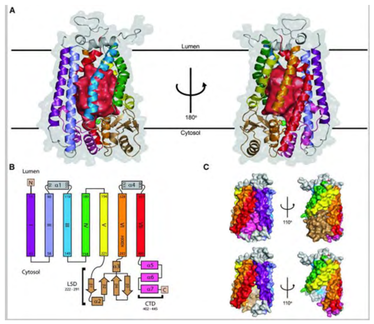
PSORTb- Prediction of subcellular localization

```
SeqID:
  Analysis Report:
    CMSVM_a          CytoplasmicMembrane      [No details]
    CWSVM_a          Unknown                  [No details]
    CytoSVM_a        Unknown                  [No details]
    ECSVM_a          Unknown                  [No details]
    ModHMM_a         CytoplasmicMembrane      [7 internal helices found]
    Motif_a          Unknown                  [No motifs found]
    Profile_a        Unknown                  [No matches to profiles found]
    SCL-BLAST_a      Unknown                  [No matches against database]
    SCL-BLASTe_a     Unknown                  [No matches against database]
    Signal_a         Unknown                  [No signal peptide detected]
  Localization Scores:
    Cytoplasmic              0.00
    CytoplasmicMembrane      9.99
    Cellwall                 0.00
    Extracellular            0.00
  Final Prediction:
    CytoplasmicMembrane      9.99
```

The information from PSORTb predicts HBOR_RS04875 is located in the plasma membrane of the cell.

Results

Significant homology suggests that the protein is a peptidase from the M48 superfamily with a ZMPSTE24-like domain region, classifying this protein as a type 1 CaaX endopeptidase containing a 7-pass transmembrane region in the plasma membrane of the archaeon.



Images showing the structure of an analog of HBOR_RS04875 (named Ste24p) found in the yeast *S. mikatae*. Without this enzyme cleaving the C-terminal CaaX region of prelamin A, mature lamin A cannot be synthesized.

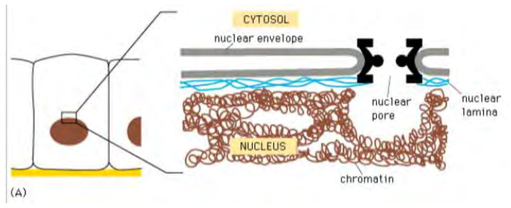


Image showing the location of nuclear lamina in a eukaryotic cell. Lamins make up the lining of the nucleus, and mutations in ZMPSTE24 that disrupt the structure of the nucleus can lead to laminopathies such as Progeria.

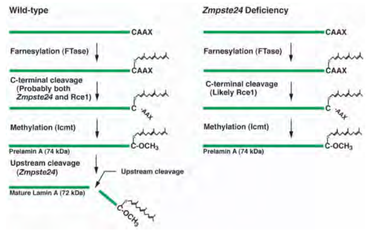


Diagram comparing the generation of mature lamin A in normal cells to cells that are not able to generate mature lamin A due to deficiency or mutation of the enzyme ZMPSTE24.

Hypothetical Experiment

Determine if the protein in *Halogeometricum borinquense* is able to replace the homologous endopeptidase found in eukaryotes:

1. Insert gene from *Halogeometricum borinquense* into a plasmid
2. Bacterial transformation- insert plasmid into bacteria
3. Lyse bacteria colonies that contain the gene and isolate the DNA sequence
4. Insert DNA sequence into yeast (eukaryotic cell) after homologous sequence in yeast has been inactivated using site-directed mutagenesis
5. Rescue experiment used to determine if inserted gene functions in new host
6. If rescue experiment successful, pull down assay used to determine if the substrate for the protein in *Halogeometricum borinquense* is similar to prelamin A

Conclusions

The bioinformatics programs used support the hypothesis that the protein coded by the gene designated at locus HBOR_RS04875 is a real protein due to homology found in many different species. Evidence given by CDD suggests the protein is a CaaX peptidase from the M48 superfamily with a domain region analogous to the region found in the human enzyme ZMPSTE24. Conservation of the gene sequence of ZMPSTE24 is important for the production of a protein that is able to cleave prelamin A so mature lamin A can be synthesized; lamin A is needed to provide support of the nuclear envelope in cells and prevent laminopathies such as Progeria. If the archaeal gene can rescue a knockout in yeast, the next step would be to test if the protein from *H. borinquense* has a substrate similar to prelamin A. If so, identifying this substrate could identify intermediate filament precursors in archaea. It also could shed light on the role of lamins in Progeria.

References

Pryor, E. E., Horanyi, P. S., Clark, K. M., Fedoriw, N., Connelly, S. M., Koszelak-Rosenblum, M., … Dumont, M. E. (2013). Structure of the Integral Membrane Protein CaaX Protease Ste24p. *Science (New York, N.Y.)*, *339*(6127), 1600–1604.

Sinha, J. K., Ghosh, S., & Raghunath, M. (2014). Progeria: A rare genetic premature ageing disorder. *The Indian Journal of Medical Research*, *139*(5), 667–674.

Young, S.G., Fong, L.G., Michaelis, S. (2005). Prelamin A, Zmpste24, misshapen cell nuclei, and progeria--new evidence suggesting that protein farnesylation could be important for disease pathogenesis. *J Lipid Res.*, 46(12), 2531-2558. PubMed PMID: 16207929.

Alberts, B., Bray, D., Johnson, A., Lewis, N., Raff, M., Roberts, K., Walter, P. *Essential Cell Biology*. (2014). Garland Pub.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z. Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25:3389-3402.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004). WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190.

Schneider, T.D., Stephens, R.M. (1990). Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* 18:6097-6100.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S.. Lo, R., Dao, P,. Sahinalp, S.C., Ester, M., Foster, L.J., Brinkman, F.S.L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13):1608-1615.

Somnhammer, E.L.L., von Heijne, G., Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *American Association for Artificial Intelligence*.

Duggan, K. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302 (1), 205-217.

Phobius: A combined transmembrane topology and signal peptide predictor. http://phobius.sbc.su.se/cgi-bin/predict.pl