A. Aptekmann[1], A.Nadra[1]

Archean core promoter, information content and its relation with optimal growth temperature.

We studied the relation between optimal growth temperature (OGT) and information content (IC), in the core promoter region of all the archeal genomes published to date, by calculating the information content of the motiff that represents the TATA binding site (TBS). We have tested several different approaches to predict transcription start sites (TSS) in a given genome we then used motiff prediction software in the flanking regions to the TSS, we constructed a database, compiling already available information from published sources, that contains characteristic growth conditions for each strain. Our work hipotesis is that protein-dna interfase in thermophiles should be different from that of mesophiles, in particular we propose and test a positive correlation between information content of binding sites and OGT in archeas.
We show that the information content increases with increasing optimal growth temperature, and this effect cannot be explained solely by an increased CG composition.
Selective pressure towards binding sites with higher binding affinity to the protein could be the reason for this correlation.
The established Rseq = Rfreq from molecular information theory doesnt take into account the effect of temperature as a selective pressure acting to skew the posible binding sites, and creating another cause for an increment in Rseq that doesnt apply to Rfreq.
Since entropy effects increase with temperature, Shannon entropy effects might as well.

***Keywords* — Information Content, Thermophiles, TATA binding protein, Basal Promoter.**

I. Purpose

Life is limited by physical and chemical extremes that define the "habitable space" within which it operates.
Aside from its requirement for liquid water, no definite limits have been established for life under any extreme[1].
We know about some aspects of the adaptations to extreme environments, in particular there has been a lot of research regarding the adaptations to temperature in order to maintain protein stability[2] and some possible mechanisms for stabilizing genomic DNA in archeas have also been proposed [3], but trough what means the fine regulation that is normally exerted trough protein-DNA interactions is maintained, still is relatively unknown and might give insights into the determinants that mediate the inter-molecular recognition process and how the extreme environments favor certain responses. Even tough our interest is in the general process of protein-DNA interaction we choose to limit our study to the realm of the archeas since they comprehend the majority of the hyper thermophiles.
One possible approach to understanding the adaptation process is looking for trends in the sequence changes of the organisms, this line of thinking lead to many proposed ideas, as for example the relation between GC content and the OGT[4].
We propose a novel relation between the information content of a protein binding site and the temperature at witch the organism lives.
We choosed to work with archeas due to the wide range of temperatures they endure.
Protein-DNA interactions are central to cell activity regulation including transcription initiation, one of the more studied cases available for archeas , being the TATA box binding protein (TBP).
TBP is involved in promoter recognition, the first step of transcription initiation. TBP is universally conserved and essential in archaea and eukaryotes. In archaea, TBPs have to be stable and to function in species that cover an extremely wide range of optimal growth temperatures (OGTs), from below 0ºC to more than 100ºC.
Thus, the archaeal TBP family is ideally suited to study the evolutionary adaptation of proteins to an extremely wide range of temperatures [5].
Organisms that do thrive in extreme environments might have in some way been affected by the selective pressure imposed by this conditions, for the particular case of the TBP we expect TBP and TATA box to co evolve responding to a number of factors, adaptation to temperature, pressure, salinity, and other extreme biophysical conditions.

II. Conclusion

Our preliminary results confirm our hypothesis, there is an apparent correlation between information content and optimal growth temperature.

References

[1] Harrison, JP et al, "The limits for life under multiple extremes ", Trends Microbiology ,2013 Apr;21(4):204-12. Epub 2013 Feb 27.
[2] Feller G, Gerday C. "Psychrophilic enzymes: Molecular basis of cold adaptation." Cellular and Molecular Life Sciences. 1997;53(10):830-41.
[3] Pereira, SL. "Archaeal nucleosomes" Proc. Natl. Acad. Sci. USA Vol. 94, pp. 12633-12637, November 1997 , Microbiology
[4] Mann, S. and Chen, YP. Bacterial genomic GC composition-eliciting environmental adaptation , Genomics, vol 95,1,2010.
[5] Kopitz, A ; Soppa, J. Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy, 2009 , page 799

This work was funded by CONICET.
[1]Department of Biological Chemistry, University of Buenos Aires , Argentina. E-mail: arielaptekmann@qb.fcen.uba.ar