Summary of Research

Functionality and Evolutionary History of the Chaperonins in Thermophilic Archaea. A Bioinformatical Perspective

Principal Investigator: Samuel Karlin

Period covered by the report:
07/01/2002 – 06/30/2003

Issued to:
Stanford University
651 Serra Street, Room 260
Stanford, CA 94305-4125

Grant Number: NCC 2-5497

We used bioinformatics methods to study phylogenetic relations and differentiation patterns of the archaeal chaperonin 60 kDa heat-shock protein (HSP60) genes in support of the study of differential expression patterns of the three chaperonin genes encoded in *Sulfolobus shibatae*. The results of this work have been published in:

Kagawa H.K., Yaoi T., Brocchieri L., McMillan R.A., Alton T., and Trent J.D. (2003) The composition, structure, and stability of a group II chaperonin is temperature regulated in a hyperthermophilic archaeon. *Mol Microbiol*, 48, 143-156.

Methods. Fifty archaeal HSP60 sequences from DNA and protein databases were aligned using the multiple sequence alignment program ITERALIGN (Brocchieri and Karlin, 1998). From this alignment, we identify 397 positions that were used to produce a high-confidence alignment of all sequences. This alignment was used to produce phylogenetic trees by the neighbor-joining (NJ) procedure (Saitou and Nei, 1987) and by the maximum-likelihood (ML) method implemented in the computer program PUZZLE (Strimmer and von Haeseler 1996). In the neighbor-joining procedure pair-wise distances between sequences were estimated: (1) inverting SSPA (Significant Segment Pairwise Alignment) similarity values (Brocchieri and Karlin, 1998) as suggested by (Feng and Doolittle, 1997); (2) by the transformation of Ota and Nei (1994) based on the Γ-distribution of the position-dependent mutational rate. In the Γ-distribution for our sequence data, the parameter a was estimated as a=1.42 by the procedure implemented in PUZZLE. Trees were also produced by setting a = 0.5, 1.0, 1.5, or 2.0, to evaluate robustness of the results over different parameterizations. NJ trees were tested by perturbation analysis ("bootstrap") with 1000 independent re-sampling experiments.

Results. Our analysis suggests that the gamma genes of HSP60 originated by duplication of an alpha gene in the progenitor of the Sulfolobales. The alpha gene is one of two paralogs (alpha and beta) resulting from a duplication of the primordial HSP60 gene that occurred before the radiation of the crenarchaeal species. The alpha-beta and alpha-gamma duplication events are indicated by a symbol '+' in the crenarchaeal section of the tree. From the length of the different branches, we can infer that preceding the species radiation the alpha paralog differentiated much more than the beta paralog. In contrast, the gamma paralog differentiated only modestly after its duplication from alpha within the Sulfolobales lineage and before radiation of the *Sulfolobus* species. After speciation of *S. tokodaii* from the progenitor of *S. shibatae* and *S. solfataricus,* the gamma gene underwent independent differentiation within the two lineages. Long branches may

indicate positive selection or relaxation of functional constraints and can be associated with the acquisition of new functions (Nei and Kumar, 2000).

From our phylogenetic tree we can also infer that the crenarchaeal paralogs originated from duplication events, which are not correlated to those that gave rise to the paralogs observed in euryarchaeal species. These are the result of at least eight (possibly ten) other independent duplication events that occurred in the lineages of *Methanobacterium*, *Thermococcus*, *Archaeoglobus*, and *Thermoplasma* (with one duplication each), or *Haloferax* (with two duplications), and *Methanosarcina* (with at least two, possibly four duplications). We gain a new perspective on the relationship between archaeal HSP60 from the recently sequenced genomes of *Methanosarcina mazei* and *acetivorans*. Both species have three closely related archaeal-type HSP60 genes that arose from one, possibly two, duplication events before their split. *M. acetivorans* has two additional archaeal-type genes (*hsp-4* and *hsp-5*) that are similar to each other (97% identity) but different from all other archaeal HSP60 genes (approx. 25% identity) and the parent gene for *hsp-4* and *hsp-5* is unrecognizable. In addition to the archaeal group II HSP60 both *M. mazei* and *M. acetivorans* have the bacterial group I HSP60 (GroESL operon). The *Methanosarcina* GroELs have 55% identity with each other and 55-59% identity with bacterial GroELs. Bacterial-type HSPs have been described in Archaea and attributed to lateral transfer (Klump and Baumeister, 1998), but this is the first example of a bacterial-type chaperonin in Archaea and the functional implications of this observation remain to be elucidated.

Conclusions. The independent evolutionary histories of HSP60 genes suggest that in different lineages they may have evolved different functions. The gamma genes among the Sulfolobales, for example, diverged and evolved rapidly. In *S. shibatae* and *S. solfataricus* the gamma genes (as well as the alpha and beta genes) remained nearly identical, while the gamma gene in *S. tokodaii* diverged. Such rapid evolution suggests an accommodation to a different or more specialized function. In general, the HSP60 genes in different lineages appeared to have duplicated independently and the different paralogs show different patterns of differentiation. These observations suggest that HSP60s in Archaea (and perhaps Eukarya and some groups of Bacteria) have been subject to intense evolutionary processing, possibly due to their high levels of expression (Karlin and Brocchieri, 2000). The duplications and modifications in HSP60s suggest that they may be exploited for multiple functions in the cell and it may therefore be inappropriate to ascribe to them a single function for all species.

Our phylogenetic analyses suggest that if HSP60s participate in a single, generic function, then this must be due to parallel or convergent evolution and. If however, chaperonins have evolved different functions in different species, new functional hypotheses must be scrutinized.

Brocchieri, L., and Karlin, S. (1998) *J. Mol. Biol.* **276,** 249-264.

Feng, D.-F., and Doolittle, R.F. (1997) *J Mol Evol* **44,** 361-370.

Karlin, S., and Brocchieri, L. (2000) *Proc Natl Acad Sci USA* **97,** 11348-11353.

Klumpp, M., and Baumeister, W. (1998) *FEBS Letters* **430,** 73-77.

Nei, M., and Kumar, S. (2000) *Molecular evolution and phylogenetics.* Oxford UK: Oxford Univ. Press.

Ota, T., and Nei, M. (1994) *J Mol Evol* **38,** 642-643.

Saitou, N., and Nei, M. (1987) *Mol Biol Evol* **4,** 406-425.

Strimmer, K., and Haeseler, v.A. (1996) *Mol Biol Evol* **13,** 964-969.

Functionality and Evolutionary History of the Chaperonins in Thermophilic Archaea. A Bioinformatical Perspective

Principal Investigator: Samuel Karlin

Period covered by the report:
07/01/2002 – 06/30/2003

Issued to:
Stanford University
651 Serra Street, Room 260
Stanford, CA 94305-4125

Grant Number: NCC 2-5497

No patents or inventions were conceived under this Grant.