

Commentary

Open Access

***Thermus thermophilus* genome analysis: benefits and implications**

Efthimia E Lioliou, Anastasia A Pantazaki and Dimitrios A Kyriakidis*

Address: Laboratory of Biochemistry, Department of Chemistry, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Email: Efthimia E Lioliou - elioliou@chem.auth.gr; Anastasia A Pantazaki - natasa@chem.auth.gr; Dimitrios A Kyriakidis* - kyr@chem.auth.gr

* Corresponding author

Published: 10 May 2004

Received: 03 May 2004

Microbial Cell Factories 2004, 3:5

Accepted: 10 May 2004

This article is available from: <http://www.microbialcellfactories.com/content/3/1/5>

© 2004 Lioliou et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

The genome sequence analysis of *Thermus thermophilus* HB27, a microorganism with high biotechnological potential, has recently been published. In that report, the chromosomal and the megaplasmid sequence were compared to those of other organisms and discussed on the basis of their physiological and metabolic features. Out of the 2,218 putative genes identified through the large genome sequencing project, a significant number has potential interest for biotechnology. The present communication will discuss the accumulating information on molecules participating in fundamental biological processes or having potential biotechnological importance.

Introduction

The *T. thermophilus* HB27 genome sequence, published on April 4 2004 [1], has already made an impact on the research community and raised intriguing questions. In the last decade a great deal of effort has been put on the biochemistry and physiology of the prokaryotes and especially the thermophilic bacteria. Since the first publication of a complete microbial genome in 1995, more than 100 microbial genomes have been completely sequenced and another 300 microbial genome sequencing projects are estimated to be in progress worldwide. Every genome that has been sequenced to date has provided new insight into the biological processes, activities and potential of these species that had not been evident before. Gene transfer, environmental applications and virulence mechanisms are only some of the many processes for which significant insight has been gained through these projects.

The genome of *T. thermophilus*, consists of a 1,894,877 base pair chromosome and a 232,605 base pair megaplasmid, designated pTT27 [1]. The completion of the *T. thermophilus* HB27 genome sequencing project resulted in the discovery of a number of new genes with potential interest for biotechnological applications [1]. Since in ther-

mophilic bacteria, fundamental cell mechanisms and processes such as protein thermostability, replication, transcription, translation, secretion, cell-signaling etc are not well understood yet, the *T. thermophilus* genome analysis will greatly improve our understanding towards that direction.

Recent reviews discuss the enzymes from *T. thermophilus* with biotechnological interest [2] and the biochemical and molecular features of thermozyms [3-6]. All these thermozyms display higher stability and activity than their counterparts currently used in the biotechnological industry. These enzymes are not only more thermostable than their mesophilic homologues, but are also more resistant to chemical agents, properties that make them extremely attractive for industrial processes [3]. Recent structural comparisons among mesophiles (or mesozymes) and thermozyms have validated numerous protein stabilizing mechanisms including hydrophobic interactions, packing efficiency, salt bridges, hydrogen bonding, reduction of conformational strain, loop stabilization, resistance to covalent destruction [6] and binding to RNA [7]. The thermal adaptation of protein synthesis in *T. thermophilus* was attributed to a key enzyme, a thiolase

responsible for a post-transcriptional modification of the thermophilic bacterial tRNAs [8].

Discussion

The 2,218 identified putative genes from *T. thermophilus* [1] were compared to those of the closest relative sequenced so far, the mesophilic bacterium *Deinococcus radiodurans*. Both organisms share a similar set of proteins, although their genomes lack extensive synteny. Substantial similarity to annotated data base entries allowed the authors to assign putative functions to 1,482 protein-coding genes. Of the remaining 736 open reading frames, about 488 had no substantial similarity to entries in public databases, and were therefore designated hypothetical proteins [1]. These were found more frequently on the plasmid (39%) than on the chromosome (20%). The G+C content in the *T. thermophilus* genome was 69.4% on average [1]. Regions with substantial lower G+C content represented ribosomal RNA clusters and at least three more gene clusters that were flanked by mobile elements. A cluster containing 15 genes with atypical codon usage showed similarities to sugar transferases, epimerases and dehydrogenases involved in lipopolysaccharide O-antigen biosynthesis.

The authors posed the question of how the thermophilic cells respond to thermal challenge at the molecular level. Surprisingly, the Ung family of proteins, coding for uracil-DNA-glycosylases, was not found in *T. thermophilus*. It is known that this gene family act as a general line of defense against cytosine deamination and its members are nearly ubiquitous in the bacterial and eukaryotic domains of the phylogenetic tree. However, base excision repair genes are represented in the *T. thermophilus* genome, while in the absence of a LexA homologue, the corresponding repair genes may be expressed constitutively.

In bacterial genomes, functionally related genes often facilitate the coordination of operons. The finding of the well studied operons in the thermophilic organisms, now in the light of the total genome analysis of *T. thermophilus*, will form the basis for further investigation into genome evolution and gene regulation. Recently, an unusual gene structure has been described in *T. thermophilus*, in which the ribosomal protein L34 coding sequence (*rpmH*) was found to be entirely overlapped by the unusually large RNase P protein subunit sequence (*rnpA*) [9]. These genes, which were part of the same operon in *E. coli*, are located near the origin of replication [9]. This co-localization of genes in a wide range of bacterial genomes implies an important linkage in their regulation of gene expression. The gene disruption method for deleting genes of interest [10] combined with the *T. thermophilus* genome sequence project will provide a unique opportunity for taking a genomics-based approach toward identifying genes

encoding different biosynthetic enzymes or regulatory proteins.

Examination of all sequenced genomes reveals that almost 40% of each genome remains as hypothetical proteins. This demands that high-throughput methodologies be developed for the efficient analysis of these large data sets, including high-throughput proteomics, gene expression and protein-protein interaction studies [11]. Microarrays are rapidly becoming standard laboratory tools for investigating gene expression under different conditions, as well as for looking at the presence and absence of genes in different strains or species that are related to a reference genome.

The full genome sequence of *T. thermophilus* [1] will revolutionize both the character and the steps on discoveries of new biotechnological enzymes or products. Stable enzymes are useful not only for industrial applications, but also for structural biology research [1,2]. *T. thermophilus* cells can now be used as a host for the selection of stable enzymes. Within this context, vector systems have been developed for extreme thermophiles and the production of stabilized mutants (3-isopropylmalate dehydrogenase) has been accomplished [10]. Moreover, *T. thermophilus* can synthesize vitamin B12, carotens, polyhydroxyalkanoate polymers in different media, suggesting that thermophilic bacteria could be useful in producing compounds of much current interest [1,2,12]. *T. thermophilus* chaperonins, a kind of heat shock proteins (Hsp), demonstrate ability to facilitate folding of several enzymes [2]. In the past, *T. thermophilus* was characterized as a treasure house for the discovery of new naturally occurring polyamines or other secondary metabolites and now most of their biosynthetic genes are predicted [1]. Ornithine decarboxylase, the key enzyme for polyamine biosynthesis, and its antizyme were purified from *T. thermophilus* [7,13]. The observation that both these regulatory proteins are bound to nucleic acids indicates that this binding facilitates the stabilization of the enzymic activities of the thermophilic organisms. Fewer two-component signal transduction systems are predicted to exist in *T. thermophilus* compared to mesophiles [1], but a more specific alignment of the sequences of interest may reveal the presence of others. The products of *atoS* and *atoC* genes, for example, have been previously predicted to constitute a two-component signal transduction system in *E. coli*, regulating expression of the *atoDAEB* operon-encoded enzymes that control short-chain fatty acid catabolism upon acetoacetate induction and antizyme has been reported to be the *atoC* gene product [14,15]. The two proteins of the fore mentioned system which have been biochemically characterized in *E. coli* [16] as well as the corresponding regulated operon were found to have homologues in the *T. thermophilus* genome. The absence

of chemotaxis cascades or methyl-accepting chemotaxis proteins and flagella biosynthetic genes indicates that the motility of this organism is restricted. However, type IV pili-forming genes were identified. Also, numerous proteases, lipases, pullulanases, α - and β -glucosidases, galacturonases, esterases, phosphatases, alcohol dehydrogenases and DNA/RNA processing enzymes were identified in the genome sequence as well.

The availability of the full genome sequence of *T. thermophilus* will immeasurably advance the re-interpretation of bacterial evolution using criteria that include the ultra structure of the cell envelope that identifies this genus as an intermediate evolutionary step between the Gram-positive (monoderms) and the Gram-negative (diderms) bacteria [17]. It has recently been reported that *T. thermophilus* actually has a periplasmic space, which is functionally similar to that of proteobacteria [18].

Transfer of genetic material between different individual bacterial cells (i.e. horizontal gene transfer) is mediated by plasmids, bacteriophages or conjugative elements. The latter have been named gene islands, or depending on their function, pathogenicity, symbiosis or degradation islands [19]. The possible presence of gene islands in the genome or megaplasmid of *T. thermophilus* will enlighten the evolutionary mechanisms that took place in order to shape the genetic content of this microorganism.

Conclusions

We have attempted to channel into the advantages coming from the genome sequence of the extreme thermophile *T. thermophilus*. It is evident that additional information and a lot of work are necessary. Our understanding of the function of many thermostable enzymes or proteins will improve by the analysis of the *T. thermophilus* genome. Nevertheless, many uncertainties must be clarified before it can be considered worthwhile to put more research endeavor in these "hot" areas. Further advances on large-scale preparation of thermophilic bacteria with high transformation efficiency, in new bioreactors working at high temperatures, could lead to significant improvements in thermophilic enzymes and other cell factory products.

References

- Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, Jacobi C, Starkuviene V, Schlenczek S, Dencker S, Huber R, Klenk HP, Kramer W, Merkl R, Gottschalk G, Fritz HJ: **The genome sequence of the extreme thermophile *Thermus thermophilus***. *Nat Biotechnol* 2004, **22**:547-53.
- Pantazaki AA, Pritsa AA, Kyriakidis DA: **Biotechnologically relevant enzymes from *Thermus thermophilus***. *Appl Microbiol Biotechnol* 2002, **58**:1-12.
- Lasa I, Berenguer J: **Thermophilic enzymes and their biotechnological potential**. *Microbiologia* 1993, **9**:77-89.
- Leuschner C, Antranikian G: **Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms**. *World J Microbiol Biotechnol* 1995, **11**:95-114.
- Zeikus JG, Vieille C, Savchenko A: **Thermozymes: biotechnology and structure-function relationships**. *Extremophiles* 1998, **2**:179-183.
- Niehaus F, Bertoldo C, Kahler M, Antranikian G: **Extremophiles as a source of novel enzymes for industrial application**. *Appl Microbiol Biotechnol* 1999, **51**:711-729.
- Pantazaki AA, Liakopoulou-Kyriakides M, Kyriakidis DA: **Ornithine decarboxylase in *Thermus thermophilus*: An RNA-associated enzyme**. *Amino Acids* 1997, **13**:299-309.
- Shigi N, Suzuki T, Tamakoshi M, Oshima T, Watanabe K: **Conserved bases in the T Ψ C loop of tRNA are determinants for thermophile-specific 2-thiouridylation at position 54**. *J Biol Chem* 2002, **277**:39128-35.
- Ellis JC, Brown JW: **Genes within genes within bacteria**. *TRENDS Biochem Sci* 2003, **10**:521-523.
- Tamakoshi M, Yaoi T, Oshima T, Yamagishi A: **An efficient gene replacement and deletion system for an extreme thermophile, *Thermus thermophilus***. *FEMS Microbiol Lett* 1999, **173**:431-437.
- Nelson KE: **The future of microbial genomics**. *Environ Microbiol* 2003, **5**:1223-1225.
- Pantazaki AA, Tambaka MG, Langlois V, Guerin P, Kyriakidis DA: **Polyhydroxyalkanoate (PHA) biosynthesis in *Thermus thermophilus*: Purification and biochemical properties of PHA synthase**. *Mol Cell Biochem* 2003, **254**:173-183.
- Pantazaki AA, Anagnostopoulos CG, Lioliou EE, Kyriakidis DA: **Characterization of ornithine decarboxylase and regulation by its antizyme in *Thermus thermophilus***. *Mol Cell Biochem* 1999, **195**:55-64.
- Canellakis ES, Paterakis AA, Huang S-C, Panagiotidis CA, Kyriakidis DA: **Identification, cloning, and nucleotide sequencing of the ornithine decarboxylase antizyme gene of *Escherichia coli***. *Proc Natl Acad Sci USA* 1993, **90**:7129-7133.
- Jenkins LS, Nunn WD: **Regulation of the ato operon by the atoC gene in *Escherichia coli***. *J Bacteriol* 1987, **169**:2096-2102.
- Lioliou EE: **Polyamine biosynthesis regulation and metabolic control of a two-component signal transduction system**. PhD Thesis Aristotle University of Thessaloniki, Chemistry Department; 2003.
- Gupta RS: **The natural evolutionary relationships among prokaryotes**. *Crit Rev Microbiol* 2000, **26**:111-31.
- Castan P, Zafra O, Moreno R, de Pedro MA, Vallis C, Cava F, Caro E, Schwarz H, Berenguer J: **The periplasmic space in *Thermus thermophilus*: evidence from a regulation-defective S-layer mutant overexpressing an alkaline phosphatase**. *Extremophiles* 2002, **6**:225-232.
- van der Meer JR, Sentchilo V: **Genomic islands and the evolution of catabolic pathways in bacteria**. *Curr Opin Biotechnol* 2003, **14**:248-254.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

