Complete Genome Sequence of the Hyperthermophilic Archaeon *Thermococcus kodakaraensis* KOD1 and Comparison with *Pyrococcus* Genomes

Toshiaki Fukui[1,2], Haruyuki Atomi[1,2], Tamotsu Kanai[1,2], Rie Matsumi[1],
Shinsuke Fujiwara[3], and Tadayuki Imanaka[1,2*]

Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering,[1] and Katsura Int'tech Center,[2] Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, and Department of Bioscience, Nanobiotechnology Research Center, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda 669-1337,[3] Japan.

Supplemental material

General features

Against the nonredundant (nr) database excluding the protein sequences from Thermococci, 45% of the proteins on the *T. kodakaraensis* genome are most similar to those of Euryarchaeota (Methanococci, 19%; Archaeoglobi, 11%, Methanopyri, 6%; Methanobacteria, 6%; Thermoplasmata, 2%; Halobacteria, 1%). A further 11% and 1% correspond most to crenarchaeal and nanoarchaeal counterparts, respectively, while 20% and 1% are closest to bacterial and eukaryal proteins. Among the bacteria-related proteins, 179 proteins exhibit the highest homology to proteins of hyperthermophilic bacteria *Thermotoga*, *Aquifex*, and *Thermoanaerobacter.*

T. kodakaraensis genome has three clusters of 29 bp-short repeats organized in tandem and regularly spaced with unique intervening sequences of almost constant length (35-48 bp) (long clusters of tandem repeats [LCTRs] or short regularly spaced repeats [SRSRs]) (Mojica et al. 2000; Zivanovic et al. 2002). These repeat clusters are located at 374071-373034 (16 repeats), 470496-469048 (22 repeats), and 833495-835847 (35 repeats) in the genome, and the consensus sequence of the repeats is GT(T/C)GCAATAAGACTCTA(A/G)GAGAATTGAAA. This consensus sequence exhibits high homology to R2 family repeats in *Pyrococcus* genomes (Zivanovic et al. 2002), while there is no cluster of R1 family repeats on the *T. kodakaraensis* genome. In addition to these short repeats, 1 pair of 8,127 bp-repeats with 97% identity is detected, but detailed analysis has revealed that they are a part of virus-related regions related to each other, as described in the main report. No other large repeat longer than 200 bp is present in the genome.

Comparison with *Pyrococcus* genomes.

There are 3-7 genes for putative transposases in Thermococcales, except for *P. furiosus* which contains more than 30 transposase genes. The majority of those in *P. furiosus* are 23 copies of the IS6 family transposase (COG3316), including two transposases at both termini of the 16-kbp composite transposon shared with *T. litoralis*. The transposases belonging to this family are not present in *P. horikoshii*, *P. abyssi*, and *T. kodakaraensis*. On the other hand, some transposases of the IS200/IS605 family in *T. kodakaraensis* show high homology to relatives found in the genus *Pyrococcus*. TK0298, TK0495, and TK0850 relate to PH0585, PH0630, and PF0760 with 76-85% identity at the amino acid level. One of the entire sets of the IS607 family, site-specific integrase (TK0655) and transposase (TK0654), share 92% and 86% identities with PF2023 and the neighboring PF2024, respectively. In addition, the primary structures and the organization of genes shared by TKV2 and TKV3 regions are also homologous to a corresponding region in the *P. horikoshii* genome (proPOF1) (Makino et al. 1999). These facts suggest that horizontal gene transfer of mobile elements naturally occurred among the order Thermococcales. However, the content of IS elements and virus-related elements in *T. kodakaraensis* are not so high when compared to, for example, those in *S. solfataricus* having as many as 170 IS elements.

Information processing

By genomic context analysis of thermophilic prokaryote genomes, Makarova et al. have predicted a thermophile-specific DNA repair system (Makarova et al. 2002). *T. kodakaraensis* possesses one cluster (TK0445-TK0464) corresponding to the predicted DNA repair system, including genes for RecB-like exonuclease (COG1468), HD-like nuclease (COG2254), two DNA helicases (COG1203), and four RMAP-superfamily proteins (COG1583, COG1688). The gene organization in this cluster is related to concatenation of two clusters in the *P. horikoshii* genome. However, unlike in *P. furiosus* and *P. horikoshii*, an orthologous gene for putative novel polymerase (COG1353) is absent in *T. kodakaraensis*.

Central metabolism

As well as most archaea, *T. kodakaraensis* does not possess genes for an oxidative branch in the pentose phosphate cycle, such as glucose-6-phosphate 1-dehydrogenase and 6-phosphogluconate dehydrogenase, indicating the absence of a complete cycle. The orthologs for split transketolase (TK0270-0269) and ribose-5-phosphate isomerase (TK1426) in the non-oxidative branch are presumed to be functional to provide the ribose moiety as an important building block for tryptophan, purines, and pyrimidines. However, almost no information is available for archaeal enzymes involved in heptose metabolism within this branch.

We have reported the first characterization and the pentagonal structure of archaeal ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), a key CO_2-fixation enzyme in Calvin-Benson (reductive pentose phosphate) cycle, from *T. kodakaraensis* (TK2290) (Kitano et al. 2001; Maeda et al. 1999). This archaeal Rubisco actually exhibited carboxylase activity with high specific activity and a highly carboxylase-specific nature against the competitive oxygenase activity, and the Rubisco protein is actually present within the cells (Ezaki et al. 1999). Nevertheless, the metabolic role of the archaeal Rubisco is still unknown, since the gene for ribulose-5-phosphate kinase, for provision of the substrate for Rubisco, ribulose-1,5-bisphosphate, is lacking on the genome. Although orthologs for the archaeal Rubisco classified as type III are present in the genera *Pyrococcus*, *Methanococcus*, *Methanosarcina*, and *Archaeoglobus*, the absence of ribulose-5-phosphate kinase is a common feature among all known archaea. Recently, a modified pathway has been proposed for generation of ribulose-1,5-bisphosphate from 5-phospho-1-pyrophosphate in methanogenic archaea (Finn and Tabita 2004). This pathway might also be functional in Thermococcales harboring the orthologs responsible for the reaction (TK0434 in *T. kodakaraensis*). Alternatively, it has been demonstrated that Rubisco-like protein without carboxylase activity in the heterotrophic bacterium *Bacillus subtilis* plays a role in a methionine salvage pathway (Ashida et al. 2003). This fact raises the possibility that the archaeal Rubisco may be involved in the methionine salvage or in the conversion of the keto-enol states of an unknown substrate.

Amino acid metabolisms

In *T. kodakaraensis*, extracellular archaeal serine protease (TK2168), three subtilisin-like proteases (TK0076 within TKV1 region, TK1675 (Kannan et al. 2001), and TK1689), and predicted thiol protease (TK1295) can be responsible for digestion of protein substrates. We also identified a unique homolog for serine protease inhibitor Serpin (TK1782) that is absent in *Pyrococcus*, although its physiological function is unclear. The resulting peptides are supposed to be imported within cells by ABC-type transporters of the dpp/opp family (TK1800-1804) and oligopeptide transporter (TK1779). *T. kodakaraensis* possesses more than 20 peptidases predicted to exhibit a variety of substrate and cleavage specificities toward the imported peptides. The amino acids can then be deaminated in a glutamate dehydrogenase-coupled manner, followed by oxidation to generate the corresponding CoA derivatives, as proposed in *P. furiosus* (Fig. 2C) (Adams et al. 2001). There are a number of aminotransferase genes for the deamination in *T. kodakaraensis*, including alanine aminotransferase (TK1094), aspartate aminotransferase (TK2268), aromatic aminotransferases (TK0260, TK0548), and multiple substrate aminotransferase (TK0186). The oxidation step involves at least four types of ferredoxin-dependent oxidoreductases with distinct substrate specificities (Schut et al. 2001), that are POR, 2-oxoisovalerate:ferredoxin oxidoreductases (TK1978-1981), KORs, and indolepyruvate:ferredoxin oxidoreductases (TK0136(α)-0135(β), TK1643(α), and TK2244(β)). The distribution and gene orders of these oxidoreductases are highly conserved among Thermococcales, except for the α- and β-subunits of the third KOR (TK0816 and TK0817, respectively) in *T. kodakaraensis* and *P. furiosus*. The CoA-derivatives are converted to the corresponding acids by the two acetyl-CoA synthetases with concomitant substrate-level phosphorylation to generate ATP. In *P. furiosus*, it has been reported that ACS I preferentially utilizes acetyl-CoA and branched chain acyl-CoAs, while ACS II is active towards aryl-CoAs in addition to branched chain acyl-CoAs (Mai and Adams 1996). Thermococcales further share three genes related to α-subunits of ACSs, and these genes may also participate in the degradation of acyl-CoAs with various side chains.

As an alternative assimilation pathway for amino acids, 2-oxoacids derived from amino acids have been proposed to be converted to corresponding aldehydes by ferredoxin-independent reactions of the oxidoreductases (Ma et al. 1997), where the resulting aldehydes can be further oxidized by the function of tungsten-containing aldehyde:ferredoxin oxidoreductase (TK1066). Alcohol dehydrogenases might be responsible for production of alcohols from aldehydes (Ma et al. 1997). In the *T. kodakaraensis* genome, there are two genes for Fe-dependent alcohol dehydrogenase (TK1008 and TK1569) and four genes for short-chain alcohol dehydrogenases (TK0673, TK0736, TK1512, and TK1677). Tungsten-containing formaldehyde:ferredoxin oxidoreductase in *P. furiosus*, capable of acting on formaldehyde, was thought to have a role in conversion of aliphatic C4 to C6 semialdehydes generated through catabolism of basic amino acids (Roy et al. 1999). Interestingly, the counterpart cannot be identified in *T. kodakaraensis*, whereas there are two genes for potential tungsten-containing oxidoreductases. One is obviously related to Wor4 in *P. furiosus*, while the other is unique to *T. kodakaraensis*. The functions of these probable oxidoreductases are still unknown. Although Wor4 in *P. furiosus* has been reported to be unable to oxidize a variety of aldehydes, hydroxy acids, nor sugar aldehydes, it has been described that its function might be in aldehyde conversion since the gene is adjacently organized with a gene encoding a protein of aldose/ketose reductase family (Roy and Adams 2002). The presence and the context of these genes are conserved among Thermococcales.

Nucleotide metabolisms

When purine biosynthesis genes are examined among the Thermococcales strains, *purO*, encoding archaeal IMP cyclohydrolase, can be identified only in *T. kodakaraensis* (TK0430). This enzyme catalyzes the second step in the two-step conversion of 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) to IMP, while the entire two steps are generally catalyzed by a bifunctional enzyme PurH in bacteria and eucarya. PurO was originally identified in *M. jannaschii* (Graupner et al. 2002), and the orthologs are present in methanogenic and halophilic euryarchaea besides *T. kodakaraensis*. The gene responsible for the first step of PurH activity, AICAR transformylase, has not been discovered in most archaea except for the normal folate-dependent enzyme in *Halobacterium* sp. NRC-1. It has been demonstrated that the transformylation step of AICAR in a few archaea was folate-independent and required formate as a C1 donor (White 1997), similar to the formate-dependent form PurT of phosphoribosylglycinamide transformylase (TK0207). A possible candidate for the novel AICAR transformylase in *P. abyssi* has been proposed to be PAB0547, a predicted carboxylate-amine ligase of the ATP grasp superfamily (Cohen et al. 2003). While the orthologous genes for PAB0547 are clustered with two subunits of GMP synthase in *Pyrococcus* spp., the counterpart in *T. kodakaraensis* TK0196 is organized with *purL* encoding a synthetase subunit of phosphoribosylformylglycinamidine synthase (TK0197). These facts may support the participation of this candidate in purine biosynthesis in Thermococcales. *T. kodakaraensis* and *P. furiosus* harbor *purK* encoding the ATP-binding subunit of phosphoribosylaminoimidazole carboxylase (TK0835) organized with the catalytic subunit gene *purE* (TK0836), but this gene is lacking in *P. abyssi* and *P. horikoshii*.

References

Adams, M.W., J.F. Holden, A.L. Menon, G.J. Schut, A.M. Grunden, C. Hou, A.M. Hutchins, F.E. Jenney, Jr., C. Kim, K. Ma, G. Pan, R. Roy, R. Sapra, S.V. Story, and M.F. Verhagen. 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **183:** 716-724.

Ashida, H., Y. Saito, C. Kojima, K. Kobayashi, N. Ogasawara, and A. Yokota. 2003. A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302:** 286-290.

Cohen, G.N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Querellou, R. Ripp, J.C. Thierry, J. Van der Oost, J. Weissenbach, Y. Zivanovic, and P. Forterre. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* **47:** 1495-1512.

Ezaki, S., N. Maeda, T. Kishimoto, H. Atomi, and T. Imanaka. 1999. Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. *J Biol Chem* **274:** 5078-5082.

Finn, M.W. and F.R. Tabita. 2004. Modified pathway to synthesize ribulose 1,5-bisphosphate in methanogenic archaea. *J Bacteriol* **186:** 6360-6366.

Graupner, M., H. Xu, and R.H. White. 2002. New class of IMP cyclohydrolases in *Methanococcus jannaschii*. *J Bacteriol* **184:** 1471-1473.

Kannan, Y., Y. Koga, Y. Inoue, M. Haruki, M. Takagi, T. Imanaka, M. Morikawa, and S. Kanaya. 2001. Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl Environ Microbiol* **67:** 2445-2452.

Kitano, K., N. Maeda, T. Fukui, H. Atomi, T. Imanaka, and K. Miki. 2001. Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. *Structure (Camb)* **9:** 473-481.

Ma, K., A. Hutchins, S.J. Sung, and M.W. Adams. 1997. Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc Natl Acad Sci U S A* **94:** 9608-9613.

Maeda, N., K. Kitano, T. Fukui, S. Ezaki, H. Atomi, K. Miki, and T. Imanaka. 1999. Ribulose bisphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. *J Mol Biol* **293:** 57-66.

Mai, X. and M.W. Adams. 1996. Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **178:** 5897-5903.

Makarova, K.S., L. Aravind, N.V. Grishin, I.B. Rogozin, and E.V. Koonin. 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* **30:** 482-496.

Makino, S., N. Amano, H. Koike, and M. Suzuki. 1999. Prophages inserted in archaebacterial genomes. *Proc Jpn Acad Ser B* **75:** 166-171.

Mojica, F.J., C. Diez-Villasenor, E. Soria, and G. Juez. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36:** 244-246.

Roy, R. and M.W. Adams. 2002. Characterization of a fourth tungsten-containing enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **184:** 6952-6956.

Roy, R., S. Mukund, G.J. Schut, D.M. Dunn, R. Weiss, and M.W. Adams. 1999. Purification and molecular characterization of the tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: the third of a putative five-member tungstoenzyme family. *J Bacteriol* **181:** 1171-1180.

Schut, G.J., A.L. Menon, and M.W. Adams. 2001. 2-Keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Methods Enzymol* **331:** 144-158.

White, R.H. 1997. Purine biosynthesis in the domain Archaea without folates or modified folates. *J Bacteriol* **179:** 3374-3377.

Zivanovic, Y., P. Lopez, H. Philippe, and P. Forterre. 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res* **30:** 1902-1910.