Nucleic acid stability in thermophilic prokaryotes: a review

Seema Trivedi1*, Satyawada Rama Rao2 and Hukam Singh Gehlot2

1Department of Zoology, J N Vyas University, Jodhpur (Raj.), India 2Department of Botany, J N Vyas University, Jodhpur (Raj.), India (*author for correspondence)

Received 09 February 2005; Accepted 11 May 2005

Abstract

In order to survive at temperatures of ≥60°C, thermophilic prokaryotes (Archaea and Eubacteria) have adopted different strategies. These strategies include high CG content in the coding sequences, nucleotide arrangement of purine-purine and pyrimidine-pyrimidine, methylation of nucleotides, histone/histone like proteins, reverse gyrase, cations, etc., also provide thermal stability to genome. Strategies adopted at the level of DNA are naturally, though not universally, reflected in RNA in thermophiles. Increased purine load (particularly of adenine), preferential codon usage, post-transcriptional modifications etc. provide thermal stability. All these factors may differ from taxa to taxa as no single or all the factors together can be universally attributed for providing thermal stability to nucleic acids.

Key Words: Archaea, thermophiles, stability, CG content, reverse gyrase

Introduction

Most of the organisms that are mesophiles thrive at optimum growth temperature (OGT) of 24-40°C. Organisms living at higher OGT of 50-70°C are thermophiles and at OGT of greater than 80°C are hyperthermophiles which include members of domains Archaea and Eubacteria. Archaea are probably the earliest living organisms that occupy diverse habitats and are accordingly identified as thermophiles, halophiles or psychrophiles etc. (Woese et al., 1990; Doolittle, 1995; Zlatanova, 1997; Makarova and Koonin, 2003). Since most of the Archaea and some Eubacteria (Kreil and Ouzounis, 2001; Bao et al., 2002) live under extreme conditions, certain characteristic features might have enabled them to survive in these environments. Some of these features are modifications in the metabolic pathways of synthesis of cofactors like heme, acetyl CoA, acyl CoA, and folic acid which are either greatly reduced or are eliminated in thermophiles because of constrains of high temperature. This trend is seen in Archaea...
growing at OGT around 60°C and enhances with the increase in OGT 83°-85°C and 95°-100°C (Kawashima et al., 2000). In this review we focus on features which have rendered thermostability to Archaeal and Eubacterial DNA and RNA with emphasis on the former. This is because only a few thermophilic Eubacteria have been studied in this regard. Further, the term thermophile in this review is being used of both thermophile and hyperthermophile prokaryotes unless specified as thermophilic an hyperthermophilic Archaea or Eubacteria respectively.

**DNA**

Protection of DNA would be essential, especially at high temperature, to ensure not only integrity of genetic information but also to carry out metabolic functions. Thermophilic Archaea and Eubacteria may have different nucleotide ratio, nucleotide preference, flexibility, modification of bases, and association with histone/histone like proteins as compared to the ones living at more moderate temperatures. These strategies may not necessarily work in isolation but may work in synergistic way.

One school of thought believes that higher CG content in genome particularly in the protein coding regions would be an essential feature to protect DNA from thermal extremes even at 83°-100°C OGT or beyond (Gorgan, 1998; Kreil and Ouzounis, 2001; Bao et al., 2002; Singer and Hickey, 2003). However, amongst prokaryotes living at high temperature, there are differences in CG content. It has been found that thermophiles have high CG content more than that of mesophiles but it is less than that of hyperthermophiles (Kreil and Ouzounis, 2001). On the contrary, other investigations suggest that CG content is not as high as expected in the thermophile genome except for in t-RNA and r-RNA sequences (Gorgan, 1998; Galtier et al., 1999; Kawashima et al., 2000; Nakashima et al., 2003). From the above reports it appears that though the CG content may be high in coding regions of the genome it may not be high in the intergenic regions and therefore possibly the total genome does not show consistent pattern of high CG content. At the same time there may be exceptions to the high CG contents in the coding region, which is evident from the report by McDonald et al., (1999) where they do not find any symmetric and consistent pattern of high CG content particularly in protein coding genes in thermophiles. Moreover, it is suggested that at high temperature, it is not the question of CG content and denaturation but thermodegradation that is important against which thermophiles need protection (Lobry and Chessel, 2003). Possibly the report on marked differences in relatively low CG content of Archaea living in cold environment and high CG content in hyperthermophiles tilts the balance in favor of high CG content being an important factor in protection of DNA (Saunders et al., 2003). However, there is no uniform pattern of high CG content in the thermophiles, it is apparent that it may be important but CG content alone cannot be responsible for providing thermal stability to DNA.

So if it is not the CG content (at least, not in all cases), there must be other features for thermal protection of the genetic material. One such feature is the effect of cations like potassium ions, and polyvalent cations in particular that neutralize the nucleic acid in order to maintain double stranded structure and provide thermal stability irrespective of the nucleotide ratio particularly in Archaea which grow at 60°C OGT or above (Gorgan, 1998; Kawashima et al., 2000). Probably differences in methylation of the nucleotides may also provide thermo protection. This is evident in report where it has been seen that in order to protect DNA from temperature induced mutations, thermophiles generally have N4-methylcytosine (m4C) instead of 5-methylcytosine (m5C) and N6-methyladenine (m6A) (Ehrlich et al., 1985). This study further point to the fact, that if the methylation patterns are different in thermophiles they may have a different type of DNA methylase to enable this difference. This is evident from the study that reports presence of DNA methylases only in prokaryotes (Bujnicki and Radlinsks, 1999); however, the evolutionary significance of this enzyme is still illusive. Since numbers of methylated nucleotides are not different in mesophiles and thermophiles (Gorgan, 1998) it is apparent that it is not the number of methylated nucleotides but the position that may be important for thermal stability.

In addition to the above factors, it is also possible that arrangement of nucleotides may render thermostability irrespective of CG content. DNA of hyperthermophiles has higher frequencies of purine-purine and pyrimidine-pyrimidine composition, which
renders least flexibility to DNA and probably provides thermostability as compared to other organisms. This kind of nucleotide arrangement is reported to increase with increasing OGT above 60° C (Kawashima et al., 2000) and may also help in supercoiling of DNA affecting thermostability of DNA (Nakashiona et al., 2003). Another feature that is common in most of the hyperthermophile Archaea and Eubacteria and not in mesophiles is the presence of reverse gyrase which provides relaxed to positive supercoil to DNA. Its presence probably helps in replication as well as in gene regulation at high temperature (Napoli et al., 2001; Bao et al., 2002). A nonspecific DNA-binding protein, Smj12, has been found to additionally help in positive supercoiling of DNA (Napoli et al., 2001). Contrary to this in *Picrophilus torridus* (Futterer et al., 2004) and *T. volcanium* (Kawashima et al., 2000) reverse gyrase gene has not been found. *T. volcanium* has gyrase and topoisomerase I and does not have chaperons like DnaK, DnaJ and GrpE. It also has a negative supercoil unlike other thermophilic Archaea (Kawashima et al., 2000). Absence of these proteins in *T. volcanium* is an exception to other thermophilic Archaea adopting strategies of alternate DNA folding. This could also be because of the fact that *T. volcanium* and *Picrophilus torridus* OGT 60°C, is the least amongst the Crenarchaeota. Therefore, it cannot be categorically stated that reverse gyrase and chaperone proteins are universally employed for thermal protection of DNA.

It is possible that transient binding/association of histone and non-histone proteins with DNA may additionally provide thermostability (Peak et al., 1995, also refer to review by Gorgan, 1998 and White, 2003). This gets support from studies on Archaeal histones HMf and HTz that wrap DNA in positive supercoil at high salt concentrations but induce negative supercoiling at low salt concentrations, whereas, MkaH induces negative supercoil under different salt concentrations. Such alternate packing may provide physiological advantage to Archaea even if it is salt concentration dependent rather than temperature dependent (Sandman et al., 1994; Musgrave et al., 2000). Probably this transient type of supercoil may vary at the time of gene activity/lack of it and may protect DNA at high temperature as only part of it would be exposed during gene activity. To alter the salt concentrations, special control of ion channels in the plasma membrane would be essential. Further studies would probably reveal the precise signal that would trigger this control. Irrespective of differences in ion concentrations, histones or histone like proteins (e.g. Crenarchaeota lack histone proteins but have Alba protein) contribute in folding and unfolding (Bailey et al., 2002) and thermal protection of DNA (White, 2003; Weng et al., 2004). In this regard, report suggest that six different histones in *M. jannaschii* may be involved with different sequences in transient manner, which would be helpful in gene regulation especially at high temperature where there is need for thermal protection of DNA (Bailey et al., 2002). Besides, these proteins, single strand binding proteins (SSB) have also been found to stabilize DNA at elevated temperature. The roles of DNA stabilizing proteins have been extensively reviewed by White (2003). The author further elaborates that though SSB is present in all the three domains, the thermal stable Archaeal SSB share partial homology with Eukaryotes and Eubacteria. However, SSB is not found in *Pyrobaculum aerophilum* and hence, the author suggests possibility of other proteins that are similar in function to SSB that may provide thermal protection to DNA in this organism. These reports clearly indicate that histones or histone like proteins are one of the essential factors that provide thermal protection to DNA irrespective of CG content of genome. Since these proteins are also present in mesophiles, the difference in the protective mechanism of these DNA associated proteins in thermophiles is not clear. These reports focus on thermophilic and mesophilic Archaea and Eubacteria and show differences in CG contents of hyperthermophiles and mesophiles but since no specific difference between Archaea and Eubacteria have been indicated, it can be speculated that there are no significant difference regarding this strategy amongst the two domains.

Despite the protection provided by the aforesaid factors, DNA in thermophiles is susceptible to damages. DNA damage in hyperthermophiles may be caused at higher rate as compared to other organisms due to hydrolytic deamination of cytosines and adenosines, hydrolytic depurination or oxidation of guanine, methylation of bases and strand break at high temperature (Gorgan, 2000; Yang et al., 2000). Therefore, it can be speculated that repair systems should be taking care of these damages. Several proteins employed for DNA damage repair in mesophiles are also found in thermophiles with some
exceptions. Homologues of Eukaryal/Eubacterial recombination repair, DNA polymerase and SSB proteins associated with repair activities are reported in thermophilic porkayotes (reviews by Gorgan, 1998, 2004; White, 2003). Crearchaeal SSB proteins are unique in the sense that they share high level of sequence similarity with Eukaryal SSB OB fold but the tail is similar to Eubacterial SSB. However, in *pyrobaculum aerophilum* genes for SSB have not been found. It is speculated that if they are present, they would be very different from SSB of other thermophilic Archaea (White, 2003). Possibly some unique proteins present in thermophiles and in particular in hyperthermophilic Archaea are employed for damage repair. In this regard unique repair-associated mysterious proteins (RAMPs) believed to repair DNA damage have been found in only thermophiles (Makarova et al., 2002) besides other repair enzymes which are also found in eukarya (Klenk et al., 1997). It is not known however whether this unique repair protein enables DNA repair only during stress conditions especially at elevated temperatures. It is believed that horizontal gene transfer may be responsible for similarities found in thermophilic and hyperthermophilic Eubacteria and Archaea regarding the RAMPs (Makarova et al., 2002). Amongst several putative proteins predicted to be involved in repair besides RAMP, two subunits of nucleotidyltransferase in Archaea raise special interest because of speculation of their being involved in thermal adaptation besides DNA repair (Makarova and Koonin, 2003). Besides these proteins, O6 alkylguanine-DNA alkyltransferase activity and not alkylphosphotriester-DNA alkyltransferase activity for DNA damage repair in the Archaea is reported. It may provide thermal protection to DNA by binding strongly and co-operatively to the double stranded DNA. This enzyme is more efficient in thermophiles than in mesophiles and the efficiency increases with rise in temperature (60°-100°C) (Skorvaga et al., 1998). Information gathered form studies done so far on DNA repair systems in thermophiles indicate absences of nucleotide excision repair (NER) and mismatch repair (MMR) system in thermophiles and in particular in Archaea. However, in some thermophiles, the repair pathways may be present that may not be complete like mesophiles (Fitz-Gibbon et al., 2002; Gorgan, 1998, 2000, 2004) with some exceptions (White, 2003). In hyperthermophilic archaeon *Pyrobaculum aerophilum*, putative U/G and T/G mismatch-specific glycosylase (Mth-MIG) has been found. Since it is not present universally, it is believed that Archaea not having MIG would be employing other methods to repair U/G and T/G mismatch (Yang et al., 2000). In thermophiles where nucleotide excision repair pathway is found, it highlights the differences between the two domains living in extreme conditions. This repair pathway (UVrABC) is complete in thermophilic Eubacteria. NER pathways in hyperthermophilic Archaea do not share any homology with Eubacteria but do so with Eukaryal pathway. However, thermophilic Archaea *Methanobacterium thermoautotrophicum* exhibits intermediate situation where it shares homology of the repair system both with Eubacteria and Eukarya (Gorgan, 2000; White 2003). In Crenarchaea (which lack UVrABC), orthologues of Eukaryal helicase and nuclease are implicated in NER (White, 2003; Gorgan, 2004). The latter author suggests that either hyperthermophilic Archaea have systems to avoid mutations or have efficient replication system which reduces requirement for other repair systems eg. Sulfolobus spp. In hyperthermophiles, B-family polymerase can detect uracil in DNA and stalls. How it repairs or recruits repair machinery is not known (Gorgan, 2004). Since complete NER and MMR systems are present in thermophilic Eubacteria, it appears that hyperthermophilic Archaea are different from thermophilic Eubacteria and have alternative strategies for excision repair of DNA.

In addition to these factors, photolyase gene activity amongst Archaea has been reported only in *Sulfolobus solfataricus*, *Methanobacterium thermoautotrophicum* (Gorgan, 2000) and *Picrophilus torridus* (Futterer et al., 2004). This gene does not show similarity with any other microbe (Skorvaga et al., 1998) except for *Picrophilus torridus* photolyase which shows similarity with bacterial homologues (Futterer et al., 2004). Since photolyase appears to be uncommon amongst Archaea, it cannot be considered as a possible universal means of DNA repair in thermophile and hyperthermophilic Archaea but may be universal in thermophilic Eubacteria. However, photolyase mediated translesion repair mechanism in thermophilic Archaea remains illusive. In hyperthermophilic Methanopyrus kandleri unique type IB topoisomerase with DNA repair activities has been found. Additionally, in hyperthermophilic Eubacteria
and Archaea (\textit{Thermotoga maritima} and \textit{Archaeoglobus fulgidus}) respectively, uracil–DNA glycosylases (UDG) differ significantly from the other Eubacterial and Eukaryal UDG enzymes at the amino acid sequence level. It is believed that the thermophiles make efficient usage of genetic material by having multiple activities in one enzyme (Sandigursky and Franklin 1999; Belova et al., 2001). High number of the exonuclease family and the accumulation of PIN-domain are reported in thermophilic Archaea and Eubacteria. This augmentation is believed to help in DNA and RNA repair, which is required more due to damages caused at elevated temperature. It is suggested that increase in PIN doamains in thermophile prokaryotes is to compensate and substitute lack of repairs systems in them. However, reasons for increase in number of exonuclease in mesophiles remain illusive (Arcus, et al., 2004).

Recombination repair systems are also reported to protect DNA at elevated temperatures as recA homologous to radA have been found to repair double stranded breaks in \textit{Pyrococcus} furiosus even at 95°C (Di Ruggiero, 1997, White, 2003; Gorgan, 1998, 2004). Additionally it is also possible that some of the repair mechanisms are not unique in thermophiles. This is evident from the report on endonuclease IV (an apurinic/apyrimidinic endonuclease and repair diesterase) from \textit{Thermotoga maritima} that shows significant similarity in structure and function with \textit{Escherichia coli} enzyme. However, significant difference is found in thermal stability where \textit{T. maritima} enzyme is stable at 88°C but \textit{E. coli} enzyme begins to denature at 70°C (Haas et al., 1999). Therefore, it appears that it is the thermal stability of repair systems that may be important at least in thermophilic Eubacteria in addition to unique repair systems (Gorgan, 1998). Many homologues of Archaeal repair systems in eubacteria and eukaryotes have been found (see Aravind et al., 1999 for details), it is still not known whether they function in the same manner in thermophilic Archaea. We agree with the proposal by Gorgan (1998) that possibly hyperthermophiles do not need efficient repair systems at the cost of vital cellular functions. The author also suggests that mutations due to replication or inefficient repair system may be beneficial to these organisms living under stressful conditions.

The common underlying feature(s) believed to provide thermostability to DNA in thermophilic Archaea and Eubacteria thus appear to be high CG content, difference in methylation of nucleotides, cations, histones/ histone like proteins, SSB and flexibility of DNA affected by nucleotide arrangement. However, differences in thermophilic Archaea and Eubacteria show that thermophilic Archaea have yet unknown system to avoid DNA damage specially by UV radiations, absence of UVrABC system and MMR that is not present in thermophilic Eubacteria and Archaea still keeps the research area open to look features that are essential for DNA repair.

**RNA**

In order to survive the high temperature conditions, thermophiles have adopted strategies at the level of transcription too. Investigations suggest that high CG content and external factors like high temperature may influence the usage of codons in RNA of thermophiles (Lobry and Chessel, 2003; Singer and Hickey, 2003).

Generally thermophiles have purine rich genome and it is speculated that purine load is preferred probably for efficient transcription of m-RNA and to avoid unnecessary double strand interaction. This is evident from study where purine loading is still preferred even in those thermophiles which have low C/G content which is not found in non-thermophiles (Lao and Forsdyke, 2000). This is supported by findings of purine richness in genomes of thermophiles (Lobry and Chessel, 2003) particularly in most of the m-RNA except for short m-RNA (Paz et al., 2004). Comparison between thermophiles and hyperthermophiles does not show significant differences in purine richness except on excluding \textit{Aeropyrum pernix}. This may be due to high CG ratio in this organism (Paz et al., 2004). However, in hyperthermophilic \textit{Pyrobaculum aerophilum} long tracts of purines have been found in coding and noncoding regions. Hence, it is suggested that purine richness could not be of significant adaptive value, as it is not preferably present in coding regions alone (Fitz-Gibbon et al., 2002). Contrary to this, it is suggested that purine rich RNA is preferred in thermophiles and particularly adenine is preferred as compared to guanine in t-RNA and r-RNA besides m-RNA. This preference is in order to avoid undesirable bonds with other RNA and specially to avoid any...
competition with t-RNA anticodon. Purine richness may be helpful in high transcription rate of many m-RNA from pyrimidine rich template and may be more stable to spontaneous hydrolysis at high temperature (Paz et al., 2004).

It is also evident that besides nucleotide content of a genome, the second most important factor is the optimal growth temperature, which decides the preference for codon usage. It has been found that thermophiles have a different preference for synonymous codon usage from that of mesophiles (Lobry and Chessel 2003) e.g. arginine (AGR and CGN) and isoleucine (ATH) etc (Lynn et al., 2002, Singer and Hickey, 2003). This must be primarily for not only ensuring stable pre-mRNA transcript and mRNA but also for translation at high temperatures. But additionally it appears that tRNA abundance may also play an important role in preferential codon usage. Possibly some elusive selective forces at higher temperatures promote this kind of nucleotide content in genome (Lynn et al., 2002, Singer and Hickey, 2003). Thermophiles are expected to avoid of 5’-UpA-3’ and 5’-CpA-3’ phosphodiester bonds in RNAs but this has not been observed in thermophilic prokaryotes. Therefore, it appears that selective pressures have not acted against these bonds even though they are susceptible to hydrolysis. It is possible that these thermophilic prokaryotes either do not need to avoid these bonds or there may be other mechanism that prevents hydrolysis or stabilises RNA (Lobry and Chessel, 2003).

Survival strategies extend to provide thermal stability to t-RNA and in particular to aminoacyl t-RNA which is labile at high temperatures but not at lower temperatures (Stepanov and Nyborg, 2002), and possibly several factors help in stabilizing RNA in thermophilic Eubacteria. These factors are high CG content, short length of RNA, increase in hydrogen bonds in helices, minimizing alternative folding, additional Watson-Crick base pairs at the base of stem loop and shortened connections between helices (Brown et al., 1993). These factors may be important in thermophilic Archaea also. However, it has been observed that higher content of dihydrouridin is present in thermostable t-RNA of other thermophiles but is underrepresented in Archaea. Therefore, at least in most thermophilic Archaea, dihydrouridin may not be important for t-RNA stability at high temperature. In fact, its absence is considered to provide thermostability in Archaea. It has also been reported that post-transcription modified nucleotides e.g. 2-thiolation of T (54) and 5-Methylcytidine contribute directly to the thermostability of tRNA species (Edmonds et al., 1991; Yokoyama et al., 1987; Lobry and Chessel, 2003). This type of post-transcriptional modification may be universally employed by thermophilic Archaea and Eubacteria. Besides this, archaeosine, a derivative of 7-deazaguanosine is exclusively present in the D-loop of Archaea t-RNA (Watanabe et al., 1997) amongst other unique derivatives (refer for details to Edmonds et al., 1991). In addition to these factors, high number of ribose methylated nucleotides and presence of Mg ions, help in stabilization of t-RNA (Edmonds et al., 1991; Kowalak et al., 1994). Like t-RNA, thermostability in r-RNA and other non-coding RNA (ncRNA), though not universally, is probably achieved by double stranded douplex, post-transcriptional modifications and higher CG content, which leads to increased hydrogen bonding (Bao et al., 2002; Klein et al., 2002; Nakashima et al., 2003; Lobry and Chessel, 2003). Reports on contribution of high CG content being important for thermal stability is contradicted by the finding that report no significant differences in the CG content in t-RNA of cold-adapted Archaea and thermophiles. Therefore, it appears that additional factors provide thermal stability to t-RNA at high temperature (Saunders et al., 2003). It is also possible that rigidity rendered to t-RNA in thermophiles provides protection against thermal damages and hence in some cases there are no differences in CG contents of thermophiles and psychrophiles. In the studies carried out so far, since only few thermophile Eubacterial genomes have been included, it is not possible at present to deduce whether all thermophilic Eubacteria have adaptations that are similar to Archaea.

**Conclusion**

From the forgoing information it is apparent that thermophilic Archaea and Eubacteria have adapted to high temperature conditions or possibly because of their unique features they have been able to survive in these conditions. The adaptations have happened in DNA, RNA which is naturally reflected in translation. Whether evolutionary forces induced changes in DNA that are reflected in expression or whether it was the
necessity of undergoing modification at the expression level that the DNA had to undergo changes still remains debatable. However, it is clear that factors like CG content, histones/histone like proteins, methylation, post-transcriptional modifications, cations etc. either alone or in combination help in nucleic acid stability even though CG content may or may not be significantly different in mesophiles and thermophiles. Differences in repair systems between the two domains also suggest that both NER and MMR systems are either absent or have not been detected in hyperthermophilic Archaea. We agree with Gorgan (1998) suggesting a possibility of evolutionarily and functionally divergent repair pathways in hyperthermophilic Archaea that do not show similarity with well known established systems. So far only few thermophilic Eubacteria have been chosen for comparison with thermophilic Archaea which have revealed few differences between the two domains regarding the strategies used for thermoprotection. The situation is compounded by the fact that the two thermophlic Eubacteria used for studies, are considered close to Archaea in lineage and appear to share many genes with Archaea possibly due to horizontal gene transfer. In future, understanding and discoveries in more thermophilic bacteria (if more exist) would probably help in comparing the strategies of the two domains. These differences and similarities may help in understanding the specific strategies that are common and necessary for surviving at high temperature.

References


