Nanoarchaeum equitans failed to maintain the balance between DNA stability and melting potential

Yoshio Kawai* and Yoshimi Maeda

Faculty of Life Sciences, Toyo University, Oura-gun, Gunma 374–0193, Japan
College of Science and Engineering, Iwaki Meisei University, Iwaki, Fukushima 970–8551, Japan

Key Words—GC-content; RNA gene; supercoil; thermal stability

Nanoarchaeum equitans is a tiny archaeon discovered in 2002 (Huber et al., 2002) and lives in an extremely high-temperature environment near boiling. They are known to coexist with the crenarchaeon Ignicoccus hospitalis. The reason why N. equitans cannot be cultivated in the absence of I. hospitalis is not yet clear, but the latter can be cultivated well. N. equitans genome DNA of 490,885 nucleotides in length has been completely sequenced (Waters et al., 2003) and some interesting features have been revealed. One of the current interesting topics is that N. equitans creates functional tRNAs from a combination of separate genes (Randau et al., 2005). On the other hand, recently published was the report on the genome of I. hospitalis (Podar et al., 2008) that allows a genomic analysis. Furthermore, fermentation studies (Jahn et al., 2008) and studies by using electron micrographic techniques (Junglas et al., 2008) unveiled some interactions involved in the association between N. equitans and I. hospitalis. It may be possible to get some clues which explain the parasitic/symbiotic partnership. In addition to the N. equitans/I. hospitalis relation, there are some interesting questions on N. equitans itself. One of them is that even though N. equitans lives in an extreme high-temperature environment of 90°C, its averaged genome G-C content is only 31.6% (Table 1). This value of G-C content corresponds to the DNA melting temperature of 82.3°C, which means that at least more than half the region of the genome DNA is always in a “melting” state if there is no specific mechanism to maintain the genome DNA in its double helix. Since additional data on Escherichia coli DNA in cells and T4 phage DNA in the head shells obtained with the heat leakage scanning calorimeter indicate that DNA melting is completed within a temperature range of about 10°C (Kawai, 1999), it may be stated that N. equitans genome DNA is, in fact, in a single-stranded state in its high-temperature environment near 90°C. Though the melting temperature, \( T_m \), of a double-stranded oligonucleotide is a function of sequence, the relation between \( T_m \) and G-C content (Marmur and Doty, 1962) gives a complete feature since, for long DNAs of heterogeneous sequence, the detailed sequence effects are almost completely averaged out (Cantor and Schimmel, 1980). It is, however, interesting to see more detailed or local features along the genome DNA because N. equitans genome DNA is covalently closed, circular DNA (cccDNA) that is topologically constrained. Unfortunately we cannot see the genome DNA state in N. equitans cells living in a high-temperature environment of 90°C. Therefore, it may be worthwhile to theoretically examine thermal stability along N. equitans genome DNA instead of direct observation of the genome. To discuss thermal stability, it

* Address reprint requests to: Dr. Yoshio Kawai, Faculty of Life Sciences, Toyo University, Itakura-machi, Oura-gun, Gunma 374–0193, Japan.
is necessary to calculate probability, $p_i$, in which the $i$-th base-pair is completely hydrogen-bonded at a given temperature. Thermal stability is usually calculated according to the Poland-Fixman-Freire algorithm (Fixman and Freire, 1977; Poland, 1974) based on the helix-coil transition, and so we followed this as well by adopting parameters $69.0 \degree C$ for A-T base-pairs and $110 \degree C$ for G-C base-pairs as the midpoint temperature for melting in $1.0 \times$ SSC buffer. Finally a thermal stability map can be obtained by plotting each temperature, $t_i$, against the $i$-th base position, where $t_i$ is a temperature (thermal stability) at which $i$-th base-pair is in the 50\% coiled state (Wada and Suyama, 1986). Although this statistical processing is originally assumed to be applied to DNA having free ends, results obtained by applying it to cccDNA might not, as a first order approximation, lead us to a baseless conclusion.

Figure 1 shows the theoretical thermal stability map of $N. equitans$ genome DNA corresponding to nucleotides 1–50,000. The feature of the thermal stability map of $N. equitans$ is extremely simple: less bumpy and almost flat, with only six sharp peaks. Because it is obvious that each sharp-peaked region has high thermal stability compared with its flanking regions,
what we have to do is to clarify whether there are any interesting relations between sharp-peaked regions and genome information. The answer is easily brought out by, for example, reference to GenBank. Each of five sharp-peaked regions, A, B, C, D and E encode one of the tRNA genes including a half gene: peak A corresponds to tRNA_{Met}^3'-half, and peak B, peak C, peak D and peak E correspond to tRNA_{Ala}^{5'}, tRNA_{Phe}, tRNA_{Met}^5'-half and tRNA_{Ala}, respectively. Here, we should pay attention to the degree of thermal stability of these regions. Each thermal stability of peak A through peak E exceeds 90°C while that of other regions is lower than 90°C with only one exceptional peak indicated by an arrow. Namely, Fig. 1 tells us the following two points: (1) as expected from the relation between $T_m$ and G-C content, almost all regions along *N. equitans* genome DNA are melting in an environment of 90°C; (2) however regions encoding tRNA genes are surely held in a double-helix, even if not completely. Actually, among total 69 regions (named 69Rs) of which peak top values of thermal stability exceed 90°C along the whole genome, 49 are tRNA genes, including half genes. Among the remaining 20 regions, 3 are rRNA genes. Although it is rather difficult to evaluate a definite number of nucleotides involved in the 17 regions other than 49 tRNA genes and 3 rRNA genes, we speculate that it is less than 4,000 base pairs in length. Therefore, RNA genes are at least 66% conducive to maintaining the genome DNA in its double-helix because the total number of nucleotides involved in tRNA genes including introns and rRNA genes is about 7,900. The rRNA genes would be expected to contribute more greatly to the thermal stability than tRNA genes since 16S and 23S rRNA genes are, if compared with tRNA genes, large enough to have high thermal stability arising from the stacking effect. Regions encoding rRNA genes must be almost completely held in the ordered double helix in *N. equitans* cells surviving in an environment of 90°C because the thermal stability of the 16S rRNA locus exceeds 100°C (Fig. 2) and the same is true for the 23S rRNA locus (Fig. 3). The thermal stability of the 5S rRNA locus is lower than that of the 16S rRNA or 23S rRNA, at about 98°C.

When we take into account *N. equitans* averaged genome DNA G-C content, *N. equitans* survives in the seriously severe high-temperature environment of 90°C because the $T_m$ value of its genome DNA is evaluated as 82.3°C by using its averaged genome DNA G-C content and the relation between $T_m$ and G-C content. This means, as mentioned above, almost all regions along the genome DNA are always in the “melting” state if *N. equitans* has no specific mechanism to maintain the genome DNA in a double helix. Some genera of archaea have histone or histone-like proteins and condense DNA into compact structures (Fähnert et al., 2001), or have reverse gyrase which will give and improve thermal stability to the DNA double helix (de la Tour, 1990). Although *N. equitans* encodes reverse gyrase (NEQ318 and NEQ434) and histone
(NEQ288 and NEQ348) genes and, further, gene products (NEQ288 and NEQ348) were copurified from *N. equitans* cells (Friedrich-Jahn et al., 2009), we have doubts whether the expression of these four genes is, in *N. equitans* cells, full enough to maintain the two DNA strands in a double helix. Briefly, the two strands of *N. equitans* genome DNA are bound with each other mainly by regions encoding RNA genes. The total nucleotide length of 69Rs (less than 12,000 nucleotides) relative to the genome DNA is roughly estimated to be not over 3%. Considering that *N. equitans* DNA is cccDNA and has no free end, many helices should accumulate near or at the regions encoding tRNA and rRNA genes to generate positive supercoil in the genome DNA originally having negative supercoil. As a result, transcription of the RNA genes will be prevented or at a very low level since RNA polymerase cannot recognize or contact those genes by being sterically inhibited. In other words, the translation process must be nearly blocked. This would result in a low level of protein production even though protein genes are transcribed. And if so, it is supposed that *N. equitans* can hardly survive without any support in its extremely severe environment. This inference is well compatible with the report suggesting that *N. equitans* derives its membrane lipids and also obtains amino acids from *I. hospitalis* (Jahn et al., 2008), and furthermore with the genomic insight indicating that *N. equitans* should be able to import and use metabolites and energy (ATP) from its host (Forterre et al., 2009). It may be possible to imagine that *N. equitans* depends on *I. hospitalis* for almost all factors involved in transcription and translation like phages depend on their host bacteria. The blocking of the translation process caused by accumulation of positive supercoil seems to be the primary reason why *N. equitans* cannot be cultivated in the absence of *I. hospitalis*. However, we do not know why *N. equitans* had chosen *I. hospitalis* for its host.

Not only *N. equitans* but also *Hyperthermus butylicus*, *Pyrobaculum aerophilum*, *Pyrobaculum islandicum*, *Pyrococcus frows*, *Pyrococcus horikoshii* and *Staphylothermus marinus* survive in about 10°C higher temperature environments than the calculated melting temperature of their genome DNA. Among them, the last six can be cultivated without any host. Therefore, we speculate that RNA polymerase can transcribe in their cells. First, we will try to explain the exclusion of *S. marinus* from the above six strains. One plausible explanation for this is that the thermal stability of their RNA genes is not exceedingly high, so regions encoding RNA genes are already partially melting, and potentially able to generate positive supercoils like other regions along the genome DNA at temperatures of their environment (100±2°C). Each $\Delta N$ value in Table 1 roughly reflects such situations, especially for *H. butylicus*, *P. aerophilum* and *P. islandicum*. Here we defined $\Delta N$ as $\Delta N=(GC_{RNA} - GC_{CG}) / GC_{CG}$, where $GC_{CG}$ is G-C content of genome DNA and $GC_{RNA}$ is the weighted average value of GC molar percents of each tRNA genes. A large $\Delta N$ value suggests high thermal stability of the tRNA gene. The accumulation of positive supercoil might spread over various regions along the genome DNA, not localize near or at particular local regions, and this would be an important difference from the case of *N. equitans*. This idea is also compatible with the suggestion that archaeal hyperthermophiles favor general stability and local melting of DNA (López-Garcia, 1999) and, insofar as can be judged from the thermal stability map (data not shown), seems to be valid for *P. frowsi* and *P. horikoshii* even though their DNAs have positive supercoiling introduced by reverse gyrase or histones to prevent excessive denaturation. Although they live in about 10°C higher temperature environments than *N. equitans*, thermal stabilities of their genome DNA are similar level to that of *N. equitans*. Though we should explain the case of *S. marinus* in a slightly different way since its genome G-C content and $T_{opt}$ values are not so different from those of *N. equitans* and $\Delta N$ value, 0.90, is rather large, we cannot explain more properly now. It may be that we will find a relation between genome size and $\Delta N$ value. Genome size and G-C distribution along genome DNA will surely affect a degree of supercoiling.

It should be noted that the “simple and less bumpy” feature typically shown in Fig. 1 is unique to *N. equitans* or at least to some archaea. For example, *I. hospitalis* displays a rather different and rugged feature on its thermal stability map (Fig. 4a), which seems to be common to many other organisms including mesophilic bacteria *E. coli* (Fig. 4b) or eukaryotic *Drosophila* (data not shown). What we learned from the combination of Fig. 1 and genetic information is that a sharp-peak region having high thermal stability on a thermal stability map strongly suggests that it may encode tRNA gene with strong possibility, which is no longer applicable to Fig. 4a or 4b. Figs. 1–3 clearly show the feature of the genome with tRNA and rRNA genes having high G-C content in the very low G-C content (<32
The balance between DNA stability and melting potential

The very low G-C content genome DNA does not frequently occur or seem preferred by organisms. The very low G-C content DNA leads to easy DNA melting and spontaneous removal of negative supercoils even in not-so-high-temperature environments. The loss of negative supercoils would not be energetically advantageous to control and maintain steady transcription and replication by extra generation and accumulation of positive supercoils. The reason why \textit{N. equitans} has a small genome and thus fewer genes than other organisms, and the biochemical or physiological reason why its tRNA and rRNA genes have high G-C content are unclear, \textit{N. equitans} might, in the process of evolution, have failed to keep the balance between DNA stability and melting potential, resulting in a loss of an appropriate DNA geometry for function.

References


