J. Biomedical Science and Engineering, 2013, 6, 19-24 **JBiSE**
http://dx.doi.org/10.4236/jbise.2013.61003 Published Online January 2013 (http://www.scirp.org/journal/jbise/)

Hydrogen bonds are related to the thermal stability of 16S rRNA

Hiroshi Nakashima, Ai Fukuoka, Yuka Saitou

Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, School of Health Sciences, Kanazawa University, Kanazawa, Japan
Email: naka@kenroku.kanazawa-u.ac.jp

Received 29 October 2012; revised 30 November 2012; accepted 5 December 2012

ABSTRACT

The number of base pairs in the 16S rRNA secondary structures of 51 bacterial sequences was counted, and the number of hydrogen bonds was estimated. The number of hydrogen bonds was highly correlated with the optimal growth temperature (OGT) rather than with the G + C content. Paired and unpaired nucleotides in mesophiles were compared to those in thermophiles. OGT exhibited a relationship with paired nucleotides but not with unpaired nucleotides. The total number of paired as well as unpaired nucleotides in mesophiles was very similar to that in thermophiles. However, the components in base pairs in mesophiles significantly differed from those in thermophiles. As compared with mesophiles, the number of G·C base pairs in thermophiles was high whereas that of A·U base pairs was low. In this study, we showed that hydrogen bonds are important for stabilizing 16S rRNAs at high temperatures.

Keywords: Optimal Growth Temperature; 16S Ribosomal RNA; G + C Content; Hydrogen Bonds; Base Pairs; Nucleotide Compositions

1. INTRODUCTION

Bacteria can live in a wide temperature range from the freezing point of water to its boiling point. This indicates that the environment where water exists in the liquid state can be inhabited by bacteria. At their living temperature, macromolecules such as protein, DNA and RNA are stable and can perform their biological functions. DNA and RNA consist of nucleotides, sugars and phosphates. Thymine and deoxyribose in DNA are replaced by uracil and ribose in RNA. DNA is double stranded and RNA is single stranded. The method of DNA stabilization at high temperatures is different from that of RNA. The dinucleotide composition of DNA is related to the optimal growth temperature (OGT) [1,2], and mononucleotide composition *i.e.*, G + C content of RNA is proportional to their OGT [2-6]. The uracil content of 16S rRNA has a significant inverse correlation with the OGT [7]. Hyperthermophiles have higher RNA G + C content. The G·C base pair has 3 hydrogen bonds and A·U base pair has 2 hydrogen bonds. Therefore, hydrogen bonds seem to play an important role for RNA thermal stability, however, the relationship between the number of hydrogen bonds and OGT has not reported yet.

Ribosomes are the machinery necessary to produce proteins based on the mRNA, which is a blueprint of genetic information. There are 3 types of bacterial ribosomal RNAs—5S, 16S, and 23S named according to their molecular weights. 16S rRNA is the most conservative of the 3 rRNAs, and is used to identify bacterial species on the basis of the phylogenetic tree. It is believed that their secondary structure, determined by base pairing, is more conservative than the nucleotide sequence. Three-dimensional structure of 16S rRNA of *Thermus thermophilus* was resolved by X-ray crystallographic studies [8,9]. The Gutell group predicted base pairs in 16S rRNA of bacteria, which are available through the web [10]. Using these data, base pairs in the 16S rRNA structures were counted and the number of hydrogen bonds was estimated. In addition to these studies, we reexamined the relationship between the G + C content of 16S rRNA and OGT.

2. MATERIALS AND METHODS

The sequences and base pairs of 16S rRNAs were retrieved from the comparative RNA web site (http://www.rna.icmb.utexas.edu/) [10]. Base pair information of bacterial 16S rRNAs was available, but archaeal data were not available; therefore, we analyzed only bacterial data in this study. Fifty sequences were randomly selected from various species to cover a wide range of OGT (**Table 1**). The data for *T. thermophilus* was obtained from the literature [9]. The dataset included 13 sequences from thermophiles, 35 sequences from

Table 1. List of species used in this study.

Species	OGT	16S rRNA			
	(˚C)	Length	G + C (%)	Base pairs (%)	Hydrogen bonds (%)
Acidaminococcus fermentans	37	1506	54.3	55.5	64.7
Agrobacterium rhizogenes	26	1434	55.3	57.2	66.1
Anaerobranca horikoshii	60	1473	54.3	56.2	66.1
Anaerocellum thermophilum	75	1534	58.6	56.0	66.5
Aneurinibacillus thermoaerophilus	55	1485	58.0	57.4	70.2
Aquifex aeolicus	85	1587	65.0	60.0	77.3
Aquifex pyrophilus	85	1584	64.9	60.0	77.3
Arthrobacter polychromogenes	30	1472	56.6	58.2	66.0
Bacillus halodurans	30	1552	54.5	57.6	67.2
Bacillus subtilis	30	1535	55.1	57.9	67.7
Bacillus thuringiensis	30	1520	53.5	57.2	66.5
Brevundimonas bacteroides	30	1421	55.0	56.4	66.0
Carnobacterium alterfunditum	20	1552	53.8	57.6	67.6
Caryophanon latum	30	1491	53.9	57.2	66.1
Clavibacter michiganensis	30	1490	56.1	58.3	66.8
Clostridium perfringens	37	1454	52.5	57.8	66.4
Clostridium symbiosum	37	1479	53.2	55.1	63.1
Clostridium thermocellum	55	1509	55.3	57.1	66.5
Cryobacterium psychrophilum	10	1501	55.1	55.7	63.7
Enterococcus faecalis	37	1466	53.9	56.0	64.9
Erysipelothrix rhusiopathiae	37	1487	50.8	57.2	64.9
Eubacterium biforme	37	1488	52.0	56.4	65.3
Geobacillus stearothermophilus	55	1510	59.1	57.6	70.4
Geobacillus thermodenitrificans	65	1497	58.6	56.9	68.7
Gordonia hydrophobica	28	1470	57.8	58.1	66.4
Heliobacterium modesticaldum	50	1471	58.0	55.8	64.8
Hydrogenobaculum acidophilum	65	1436	54.6	55.7	65.4
Lactobacillus casei	30	1527	52.8	56.5	66.1
Lactobacillus catenaformis	37	1515	58.8	53.6	65.4
Lactococcus lactis	30	1489	51.4	56.3	64.0
Mycobacterium aichiense	37	1458	57.8	58.0	64.4
Mycobacterium bovis	37	1535	57.9	58.8	65.9
Mycoplasma pneumoniae	37	1465	45.7	57.0	61.5
Propionibacterium acnes	37	1522	57.1	57.3	61.8
Rhodospirillum rubrum	25	1422	55.9	55.4	65.0
Rhodovibrio salinarum	37	1439	59.0	56.5	67.6
Rubrobacter xylanophilus	60	1509	61.6	56.7	70.4
Ruminococcus hansenii	37	1453	53.1	54.0	61.9
Saccharococcus thermophilus	65	1538	57.7	57.3	69.0
Spiroplasma diabroticae	30	1432	49.4	54.2	60.1
Sporosarcina psychrophila	20	1458	55.1	55.1	64.7
Streptococcus agalactiae	30	1451	52.0	55.8	63.5
Streptococcus pseudoporcinus	37	1472	52.8	57.1	65.6
Streptococcus thermophilus	45	1542	52.6	56.9	64.9
Streptomyces avermitilis	28	1530	58.0	57.8	68.0
Streptomyces tendae	28	1530	59.1	57.7	69.0
Thermoanaerobacterium aotearoense	65	1478	54.8	57.0	67.9
Thermus thermophilus	85	1521	64.0	57.9	71.3
Vagococcus fluvialis	30	1486	52.7	54.5	64.1
Weissella kandleri	30	1528	50.5	56.2	64.8
Weissella paramesenteroides	30	1527	50.4	56.5	64.9

mesophiles, and 3 sequences from pyschrophiles. Thermophiles grow above 55˚C and psychrophiles grow below 20˚C. There are 4 types of nucleotides; hence, 16 types of base pairs are possible. However, an A·U base pair is identical to a U·A base pair. Therefore, 10 types of base pairs were considered; they were as follows: A·U, G·C, G·U, A·G, C·C, U·U, C·U, A·C, A·A, and G·G. Interestingly, G·C, A·U, and G·U base pairs were dominant, and the sum of other 7 base pairs equaled only approximately 5% of the total base pairing. Therefore, only the hydrogen bonds comprising G·C, A·U, and G·U base pairs were taken into account in this study. The number of hydrogen bonds for G·C and A·U base pairs was estimated as 3 and 2, respectively. The number of hydrogen bonds for G·U base pair has been reported to be 1 or 2 [11]. Therefore, the number of hydrogen bonds was estimated by the following equations:

$$\text{Hydrogen bonds} = \text{G}\cdot\text{C pairs}\times 3 + \text{A}\cdot\text{U pairs} \times 2 + \text{G}\cdot\text{U pairs}\times 1$$

or

$$\text{Hydrogen bonds} = \text{G}\cdot\text{C pairs}\times 3 + \text{A}\cdot\text{U pairs} \times 2 + \text{G}\cdot\text{U pairs}\times 2$$

We also calculated the hydrogen bonds consisting of G·C and A·U base pairs as follows:

$$\text{Hydrogen bonds} = \text{G}\cdot\text{C pairs}\times 3 + \text{A}\cdot\text{U pairs}\times 2$$

Using these equations, we calculated the number of hydrogen bonds in 3 ways. The percentage of hydrogen bonds was calculated as the number of hydrogen bonds divided by the length of 16S rRNA. The percentage of base pairs was calculated as the sum of G·C, A·U, and G·U base pairs divided by the length of 16S rRNA. OGTs were retrieved from the web site http://www.dsmz.de/species/strains.htm. Mono- and dinucleotide compositions of 16S rRNAs were calculated. Expected dinucleotide compositions were calculated using the mononucleotide compositions, and the ratio of observed/calculated compositions was thus obtained. The average compositions were also calculated for both thermophiles and mesophiles. The average of mesophiles was calculated, including data from three psychrophiles. The 3 psychrophiles examined did not differ significantly from the mesophiles with regard to the G + C content or the percentage of hydrogen bonds.

3. RESULTS

3.1. Hydrogen Bonds in 16S rRNA versus OGT

The plot of the percentage of hydrogen bonds consisting of G·C and A·U base pairs versus OGT expressed in degrees Celsius (˚C) showed the highest correlation (correlation coefficient, 0.65) (**Figure 1**). When the hydrogen bonds from the G·U base pairs were considered, the correlation coefficient was found to be 0.63 and 0.57, when the G·U base pair was assumed to contain 1 and 2 hydrogen bonds, respectively. Several G·U base pairs were observed in the 16S rRNA secondary structures; however, hydrogen bonds from G·U base pairs did not increase the correlation with the OGT. G + C content of the 16S rRNA versus the OGT is shown in **Figure 2**. This resulted in a correlation coefficient of 0.59, which was lower than the hydrogen bonds. The ratio of G·C base pairs in the G·C and A·U base pairs increased with OGT, and the ratio of G·C base pairs showed a high correlation with the G + C content (correlation coefficient, 0.96). This result indicates that a correlation between

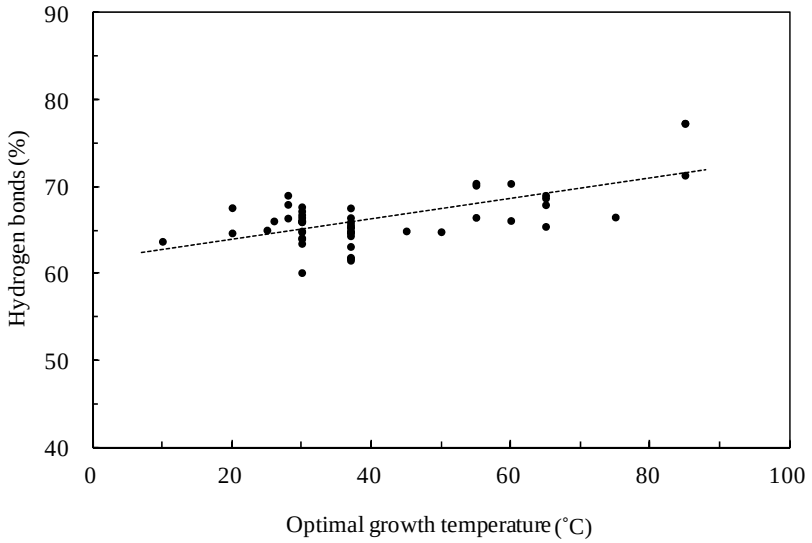

Figure 1. Hydrogen bonds (%) in 16S rRNAs against optimal growth temperature. The hydrogen bonds consist of G·C and A·U base pairs.

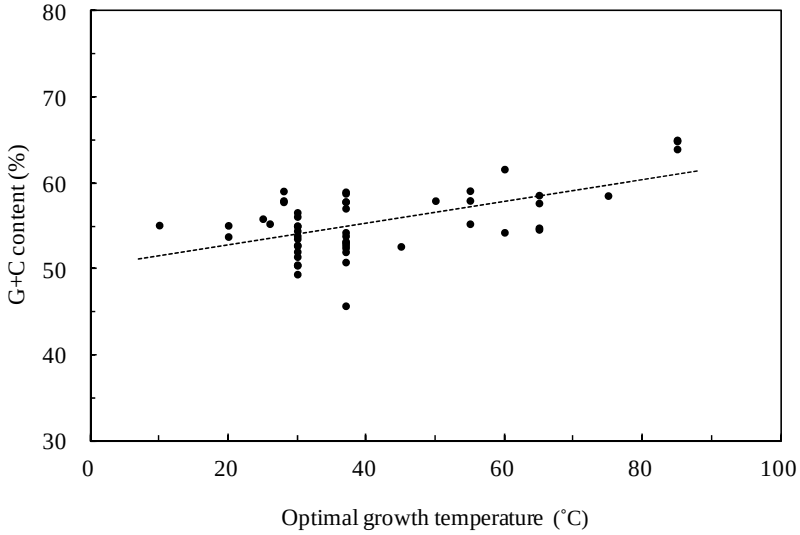

Figure 2. G + C contents (%) of 16S rRNAs against optimal growth temperature.

RNA G + C content and OGT is a secondary effect of the hydrogen bonds and OGT.

3.2. Nucleotide Composition of 16S rRNA

To make maximum base pairing, the guanine content should be equal to the cytosine content, and the adenine content should be equal to the uracil content. However, the guanine content was higher than that of cytosine, and the adenine content was higher than that of uracil, except in the case of *Propionibacterium acnes*. To show the difference in the nucleotide content, we calculated the ratios of guanine/cytosine and adenine/uracil. The average ratios of guanine/cytosine and adenine/uracil were 1.36 and 1.25, respectively. This result indicates that purines are more abundant than pyrimidines in 16S rRNA sequences. The numbers of paired and unpaired nucleotides were estimated for both thermophiles and mesophiles by using the average compositions as the length of 16S rRNA, which was assumed to be 1500 nucleotides. **Table 2** shows the comparison between mesophiles and thermophiles, with the data for thermophiles represented within parentheses. The deviation of nucleotide components in the whole sequence between mesophiles and thermophiles was roughly identical with the deviation in paired nucleotides. For example, the deviation of adenine in the whole sequence was 25 and that in paired was 24. The number of unpaired nucleotides was roughly identical between mesophiles and thermophiles, with the exception of uracil. This result indicates that unpaired nucleotides are independent of the OGT. Adenine was the most abundant, and the cytosine was the least in the unpaired nucleotides. This result was consistent with the high percentage of unpaired adenine in 16S rRNA structure models [6,12]. It is reported that nearly 75% of AA dinucleotides are found in loops in rRNA sequences [10]. We found that the AA dinucleotide is the most favorable on the basis of the ratios of observed to calculated composition (see below). Total base-paired nucleotides in thermophiles were slightly higher than those in mesophiles, however, the base pair components were quite different between the 2 groups. For example, base-paired cytosine in thermophiles was 38 higher than that in mesophiles. In contrast, paired uracil was higher in mesophiles than in thermophiles. This result indicates that, in contrast to mesophiles, G·C base pairs were abundant and A·U base pairs were few in thermophiles. Thus, thermophiles increase the amount of G·C base pairs in rRNAs to adjust to high temperatures.

The relationship between G+C content and OGT can be expressed as shown in **Eq.1** from our previous study [2];

$$\mathrm{OGT} = 2.91\times(\mathrm{G}+\mathrm{C})-103 \qquad (1)$$

where OGT is estimated in degrees Celsius (˚C), and G + C refers to the percentage of guanine and cytosine content in 16S rRNA. In this study, the relationship obtained by least-square regression analysis was slightly different

Table 2. Comparisons of paired and unpaired nucleotides of 16S rRNA between mesophiles and thermophiles. Data for thermophiles are indicated within parentheses.

Nucleotide	Whole	Paired	Unpaired
A	376 (351)	110 (86)	266 (265)
C	343 (381)	253 (291)	90 (90)
G	471 (503)	315 (344)	156 (159)
U	310 (265)	171 (139)	139 (126)
total	1500 (1500)	849 (860)	651 (640)

from **Eq.1** and was expressed as **Eq.2**:

$$\mathrm{OGT} = 2.75 \times (\mathrm{G} + \mathrm{C}) - 111 \quad (2)$$

Favorable and unfavorable dinucleotides were estimated in terms of ratios of the observed to calculated compositions. The AA and UG dinucleotides showed average ratios greater than 1.1 and were considered favorable, and AU and UC had average ratios less than 0.9 and were considered unfavorable. The other 12 dinucleotides showed average ratios in the range of 0.9 - 1.1.

4. DISCUSSION

The number of hydrogen bonds was estimated by the secondary structures of the paired bases. Therefore, the accuracy of base pair is very important. Gutell *et al*. evaluated the 16S and 23S rRNA structure models against the crystal structures, and they reported approximately 97% - 98% of the base pairings predicted were indeed corresponding to their experimental data [13].

A simple way to attain thermal stability of nucleic acids is to increase the number of hydrogen bonds, *i.e*., the G + C content. This is indeed observed for the 16S rRNAs of thermophiles, in which the G + C content is increased with OGT. If the same strategy for thermal stability as observed in RNA was applicable to DNA as well, the increased G + C content would have significant effects on the amino acid composition. For example, amino acids encoded by G + C rich codons such as Ala, Arg, Gly, and Pro would be abundant, whereas those encoded by G + C poor codons such as Lys, Ile, Tyr, and Phe would be less represented in thermophilic proteins. We found no correlation between the G + C content of DNA and OGT. Instead, the dinucleotide composition of DNA was found to be correlated with OGT [2]. rRNAs do not encode proteins; therefore, G + C content of rRNA seemed to have no restrictions. In fact, the G + C content of RNA depend on the G + C content of genomic DNA in mesophiles [14], whereas no correlation was observed between the G + C contents of RNA and genomic DNA in thermophiles [2]. As shown in **Figure 2**, the G + C contents from mesophiles showed greater deviations from the regression line than the plots in hydrogen bonds (**Figure 1**). This result suggests that mesophiles have more freedom to have various G + C content in 16S rRNA than hydrogen bonds.

Hydrogen bonds of G·U base pairs did not increase the correlation to OGT. The hydrogen bond of a G·U base pair was observed in the *syn* conformation of guanine, on the edge of the bulge loop in the hairpin loop structure [11]. If G·U base pairs are located toward the center of the secondary structure, which forms the helical structure, it is difficult for guanine to assume the *syn* conformation; hence, it might be difficult to form hydrogen bonds between the G·U base pairs. It is reported that G·U base pairs are found in different conformations in different chemical and structural environments, and the RNA double helix can be more easily altered at sites of G·U base pairs [15]. Future studies are needed to examine the relationship between hydrogen bonds of G·U base pairs and thermal stability.

REFERENCES

[1] Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K. and Suzuki, M. (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 14257-14262. doi:10.1073/pnas.97.26.14257

[2] Nakashima, H., Fukuchi, S. and Nishikawa, K. (2003) Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures. *The Journal of Biochemistry*, **133**, 507-513. doi:10.1093/jb/mvg067

[3] Galtier, N. and Lobry, J.R. (1997) Relationships between genomic G + C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *Journal of Molecular Evolution*, **44**, 632-636. doi:10.1007/PL00006186

[4] Galtier, N., Tourasse, N. and Gouy, M. (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science*, **283**, 220-221. doi:10.1126/science.283.5399.220

[5] Hurst, L.D. and Merchant, A.R. (2001) High guanine-cytosine content is not an adaptation to high temperature: A comparative analysis amongst prokaryotes. *Proceeding of the Royal Society B. Biological Sciences*, **268**, 493-497.

[6] Wang, H.-C. and Hickey, D.A. (2002) Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Research*, **30**, 2501-2507. doi:10.1093/nar/30.11.2501

[7] Khachane, A.N., Timmis, K.N. and Martins dos Santos, V.A.P. (2005) Uracil content of 16S rRNA of thermophilic and psychrophilic prokaryotes correlates inversely with their optimal growth temperatures. *Nucleic Acids Research*, **33**, 4016-4022. doi:10.1093/nar/gki714

[8] Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5A resolution. *Science*, **292**, 883-896. doi:10.1126/science.1060089

[9] Brodersen, D.E., Clemons Jr., W.M., Carter, A.P., Wimberly, B.T. and Ramakrishnan, V. (2002) Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16S RNA. *Journal of Molecular Biology*, **316**, 725-768. doi:10.1006/jmbi.2001.5359

[10] Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi,

L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N. and Gutell, R.R. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 1-31. doi:10.1186/1471-2105-3-1

[11] Cheong, C., Varani, G. and Tinoco Jr., I. (1990) Solution structure of an unusually stable RNA hairpin, 5'G-GAC(UUCG)GUCC. *Nature*, **346**, 680-682. doi:10.1038/346680a0

[12] Gutell, R.R., J.J. Cannone, J.J., Shang, Z., Du, Y. and Serra, M.J. (2000) A story: Unpaired adenosine bases in ribosomal RNAs. *Journal of Molecular Biology*, **304**, 335-354. doi:10.1006/jmbi.2000.4172

[13] Gutell, R.R., Lee, J.C. and Cannone, J.J. (2002) The accuracy of ribosomal RNA comparative structure models. *Current Opinion in Structural Biology*, **12**, 301-310. doi:10.1016/S0959-440X(02)00339-1

[14] Muto, A. and Osawa, S. (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 166-169. doi:10.1073/pnas.84.1.166

[15] Varani, G. and McClain, W.H. (2000) The G·U wobble base pair. *EMBO Reports*, **1**, 18-23. doi:10.1093/embo-reports/kvd001

