Volume 13 Number 4 1985 Nucleic Acids Research

DNA methylation in thermophilic bacteria: N^4-methylcytosine, 5-methylcytosine, and N^6-methyladenine

Melanie Ehrlich, Miguel A.Gama-Sosa, Laura H.Carreira[1], Lars G.Ljungdahl[1], Kenneth C.Kuo[2] and Charles W.Gehrke[2]

Department of Biochemistry, Tulane Medical School, New Orleans, LA 70112, [1]Department of Biochemistry, University of Georgia, Athens, GA 30602, and [2]Department of Biochemistry, Experiment Station Chemical Laboratories, University of Missouri, Columbia, MO 65201, USA

Received 14 November 1984; Revised and Accepted 18 January 1985

ABSTRACT

While determining the minor and major base composition of the DNA from 17 types of thermophilic bacteria by high performance liquid chromatography (HPLC) of enzymatic digests, we have discovered a novel base, N^4-methylcytosine (m^4C). Its structure was proven by comparison of the DNA-derived nucleoside to the analogous authentic compound by HPLC, UV spectroscopy, and mass spectroscopy. Eight of the bacterial DNAs contained m^4C. Only two contained the common minor base, 5-methylcytosine (m^5C), and neither of these was from an extreme thermophile. The other prevalent modified base of bacterial DNA, N^6-methyladenine (m^6A), was found in nine of the DNAs. Restriction analysis revealed that four of the DNAs had dam-type (Gm^6ATC) methylation patterns. Due to the propensity of m^5C residues to be deaminated by heat to thymine residues and to inefficient repair of the resulting mismatched base pairs, thermophiles with optimal growth temperatures of ≥60°C generally may avoid having m^5C in their genomes. Instead, some of them have deamination-resistant m^4C residues.

INTRODUCTION

5-Methylcytosine (m^5C) and N^6-methyladenine (m^6A) are frequently found as minor bases in bacterial DNA (1,2). They are usually (1,3,4), but not always (5,6) part of restriction-modification systems. Two aspects of modification of bacterial DNA have been associated with DNA damage or repair. Compared to the analogous cytosine (C) residues, m^5C residues are preferred sites for spontaneous base substitutions in the lacI gene (7). This is probably partially due to m^5C residues in DNA being more susceptible to heat-induced deamination than are the corresponding C residues (8) and to a lower efficiency for repair of deaminated m^5C residues than for repair of deaminated C residues by the ubiquitous uracil-DNA glycosylase pathway (9,10). A link between DNA methylation and repair is seen in Escherichia coli, in which m^6A in GATC sequences (dam methylation) appears to be involved in strand discrimination during DNA mismatch correction in newly replicated DNA (11-16).

Since thermophiles might be especially vulnerable to heat-induced

deamination of bases in their DNA or prone to misincorporation of bases during DNA replication, we have studied the minor base composition of DNA from thermophilic bacteria. Depending on the species, thermophilic bacteria can grow at temperatures of >50°C and, in rare cases, even over 100°C (17,18). We proposed that thermophiles might avoid having m^5C in their DNA to minimize $m^5C \rightarrow T$ transitions at high temperatures or else they might possess mismatch repair systems which are unusually efficient at removing T residues from T·G mispairs in their DNA. In contrast, the presence of m^6A as a minor base in the DNA of thermophiles might pose no problem to bacteria living at high temperatures, but rather, as in E. coli (12,15,16), could be utilized to increase the efficiency of mismatch repair at the replication fork by directing it to the newly synthesized DNA strand. Alternatively, it could be part of restriction-modification systems.

By high performance liquid chromatography (19,20) we compared the deoxynucleoside composition of enzymatic digests of DNA from various thermophilic and mesophilic bacteria. Among the bacteria that we examined are cellulose-degrading, hemicellulose-degrading, starch-degrading, homoacetogenic (producing acetate from glucose), acetogenic (producing mixed acids), methanogenic (methane-producing), ethanogenic (ethanol-producing), and sulfate-reducing anaerobes. In the course of quantitating the major and minor base composition of their genomes, we identified a novel minor DNA base, N^4-methylcytosine (m^4C), in approximately half of the thermophiles.

MATERIALS AND METHODS

Propagation of bacteria.

The bacteria included in this study were obtained from several sources. The sulfate-reducing bacteria (21) were from the collection of H. Peck with the exception of Desulfovibrio sp. which is a newly isolated species obtained from M. Nacro. Two of the methanogens, Methanobacterium thermoautotrophicum (22) and Methanobacterium formicicum (23) as well as Clostridium thermosaccharolyticum (24) were obtained from the strain collection of J. Wiegel. The remaining methanogenic bacteria were supplied by W. Whitman. The cultures of Thermobacteroides acetoethylicus and Thermoanaerobium brockii (24) were originally obtained by J. Weigel from G. Zeikus. The clostridial species RBl, 2, 3 and 9 are newly isolated species from Icelandic hot springs (Ljungdahl et al., unpubl. results). All other anaerobic bacteria (24-29) were from stock cultures maintained in the laboratory of L.

Ljungdahl. Bacillus stearothermophilus strains were from New England Biolabs Co.

All bacteria were grown in 1-10 L batches at or near their temperature and pH optima. Anaerobes were under either argon, CO_2, or CO_2 and H_2 (20:80, v/v), and the facultatively anaerobic B. stearothermophilus and the aerobic E. coli and Micrococcus luteus under air. Clostridium thermocellum was grown in a medium containing cellobiose and the carbohydrate-utilizing anaerobes were grown on media containing starch and/or various sugars (30). Sulfate reducers were grown with lactate as the sole carbon source and methanogens on medium containing formic acid or pressurized with CO_2 and H_2 (20:80, v/v).

Isolation and purification of DNA. After washing with 50 mM Tris-HCl, 20 mM EDTA, pH 7.6, the bacterial pellets were resuspended in the same buffer and then egg white lysozyme (Calbiochem-Behring) was added to a final concentration of 0.4 mg/ml. The suspensions were incubated for 15 min at 25°C and then sodium dodecyl sulfate and proteinase K were added to final concentrations of 1% and 50 µg/ml, respectively, for 30 min of incubation at 65°C. Subsequently, the samples were extracted with phenol and chloroform/isoamyl alcohol (24:1, v:v) and then spooled upon the addition of sodium acetate to 0.3 M and ethanol to 70% (31). In some cases, the DNA did not have a high enough molecular weight to spool so that the precipitated DNA was collected by centrifugation. After dissolving the DNA precipitates in 15 mM NaCl, 1.5 mM sodium citrate, pH 7.0, contaminating RNA was digested with 40 µg/ml of RNase A and 12 units/ml of RNase T1 for 7 h at 37°C. The DNA was then incubated with proteinase K, extracted with organic solvents, precipitated with ethanol as above, dissolved in 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.6, and stored at 5°C.

After lysozyme treatment, M. thermoautotrophicum had to be incubated with 0.25 M β-mercaptoethanol for 30 min at 0°C. Lysis was obtained by subsequently treating with 1 mg/ml of bovine trypsin (Sigma, Type 2884) for 30 min at 37°C and then the DNA was isolated as described above.

As indicated below, some of the DNA samples were further purified by hydroxyapatite chromatography or density gradient centrifugation. In the former case, 200 µg of DNA was sheared by passing it three times through a 21 gauge needle and then applied to a 10 ml hydroxyapatite column and the double-stranded fraction was collected (32). This DNA was then dialyzed extensively against 0.3 M sodium acetate, 10 mM Tris-HCl, 1 mM EDTA, pH 7.6, and precipitated with ethanol, dissolved, and stored as described above.

Alternatively, the DNA solution was brought to a density of 1.68 to 1.71 g of CsCl per ml and then purified by isopycnic centrifugation. Only the peak DNA fractions (monitored by ethidium bromide-induced fluorescence of aliquots) were pooled. These were dialyzed and precipitated as above.

Restriction analysis. DNA samples (1 µg) were digested with 2-5 units of restriction endonuclease for 2 h according to the manufacturer's instructions (Bethesda Research Laboratories; New England Biolabs). They were then electrophoresed in 1% agarose gels.

DNA digestion and high performance liquid chromatography. DNA was digested to nucleosides with nuclease P1 and E. coli alkaline phosphatase (19). High performance liquid chromatography for quantitation of m^5dCyd, m^6dAdo, and the major deoxynucleosides utilized a reversed phase column (LC-18-DB, 250 x 4.6 mm, Supelco) at 30°C as previously described (19) with the following eluents (Solvent System 1) at a flow rate of 1.5 ml/min: 2.5% methanol (7.5 ml), 5.0% methanol (12 ml), and then 12.5% methanol (23 ml), all in constant 0.05 M potassium phosphate, pH 4.0. In this solvent system m^5dCyd and m^4dCyd comigrated but their approximate relative amounts could be determined by comparing the absorbance at 254 nm and at 280 nm. For accurate quantitation of m^4dCyd, under conditions in which it was resolved from m^5dCyd, the following solvents (Solvent System 2) were used with the above column: at 27°C and a flow rate of 1 ml/min: 2.5% methanol, 0.01 M ammonium phosphate, pH 5.3 (A; 12 ml); a concave gradient of A to 13% methanol, 0.01 M ammonium phosphate, pH 5.2 (B; 20 ml); a concave gradient of B to 20% methanol, 0.01 M ammonium phosphate, pH 4.9 (10 ml); and a concave gradient from 20% methanol, 0.01 M ammonium phosphate, pH 4.9, to 35% acetonitrile, 0.01 M ammonium phosphate, pH 4.9 (20 ml). The gradients for the second, third, and fourth steps were from programmed curves 2, 3, and 2, respectively, of a Perkin-Elmer Series-4 Solvent Delivery System.

RESULTS AND DISCUSSION

Major Base Content

By HPLC of DNA quantitatively digested to deoxynucleosides, we have determined the major base composition of 17 types of thermophilic bacteria of various species (Tables 1 and 2) including extreme thermophiles (those that grow well at ≥65°C). The base composition of a number of these species, had not been previously reported. For the DNA of the other species our HPLC analysis of enzymatic digests gives much more accurate determinations of their DNA base composition than do previous thin layer chromatographic

Table 1. Minor and major base composition and dam methylation of the DNA of extreme thermophilic bacteria

Species and strain	Optimum growth temp. (°C)	Mol%[a] m^5C	m^4C	m^6A	$A + m^6A + T$	dam methylation[b]
Thermobacteroides acetoethylicus	70	<0.01	0.16	0.08	65	+
Thermoanaerobium brockii	70	<0.02	<0.02	<0.02	64	NA[c]
Clostridium thermohydrosulfuricum (JW102)	68	<0.02	<0.02	<0.02	62	-
Clostridium sp. (RB2 and RB3)	70	<0.01	<0.01	<0.01	65	-
Thermoanaerobacter ethanolicus (JW200)	68	<0.01	<0.01	<0.01	64	-
Methanobacterium thermoautotrophicum (JW501)	68-70	<0.01	<0.01	0.06	51	-
Methanococcus thermolithotrophicum	65	<0.01	0.19	0.26	66	-
Acetogenium kivui (RW)	66	<0.02	<0.02	0.13	66	NA
Bacillus stearothermophilus (N)	65-70	<0.01	0.06	<0.01	57	-

[a]These data and those in the following tables were derived from replicate (2-4) determinations of the total deoxynucleoside composition of enzymatic digests of the indicated DNA. The mean mol% (percentage of total bases analyzed at the deoxynucleoside level) is given.

[b]dam methylation, N^6-methylation of adenine residues at GATC residues as assessed by sensitivity to digestion by DpnI and resistance to MboI; +, dam-type methylation; -, no detectable dam-type methylation.

[c]NA, not assayed.

analyses of acid digests or indirect determinations from the melting temperature or buoyant density (26,30,33). The accuracy of our data is indicated by the fact that the mean difference between deoxycytidine (dCyd; unmethylated plus methylated) and deoxyguanosine (dGuo) contents was only 1.3%. Similarly close were the deoxythymidine (dThd) and deoxyadenosine (dAdo; unmethylated plus methylated) contents of these DNAs.

As shown in Table 2, two so-called strains of Clostridium thermoauto-

Table 2. Minor and major base composition and dam methylation of the DNA of thermophilic bacteria[a]

Species and strain	Optimum growth temp. (°C)	Mol% m^5C	m^4C	m^6A	$A + m^6A + T$	dam methylation
Clostridium thermocellum (JW20)	60	<0.01	<0.01	<0.01	60	-
Clostridium thermoaceticum(72 and ATCC 39073)	60	<0.01	0.20	0.79	44	+
Clostridium thermoautotrophicum (701/3)	60	<0.01	0.07	0.93	42	+
Clostridium thermoautotrophicum (701/5)	60	<0.02	<0.02	<0.02	66	-
Clostridium thermosaccharolyticum	60	<0.02	<0.02	0.25	65	+
Clostridium sp. (RB1, RB9)	56	<0.03	0.05	<0.01	65	NA
Desulfotomaculum nigrificans	55	0.18	0.10	0.09	53	NA
Bacillus stearothermophilus (H3)[b]	45-50	0.09	0.39	0.07	42	-

[a]Determination of base composition, dam methylation, and symbols are as indicated in Table 1.

[b]The exact species designation for B. stearothermophilus H3 as well as of B. stearothermophilus N has not been determined (N. Welker, pers. commun.)

Table 3. Minor and major base composition and dam methylation of the DNA of mesophilic bacteria

Species and strain	Optimum growth temp. (°C)	Mol%				dam methylation[a]
		m^5C	m^4C	m^6A	A + m^6A + T	
Methanobacterium formicicum	37	0.16	<0.01	0.03	53	-
Methanococcus voltae	32-40	<0.01	<0.01	<0.01	70	NA
Methanococcus delta	38	<0.01	<0.01	0.12	67	-
Clostridium aceticum	30	<0.01	<0.01	0.10	64	-
Clostridium formicoaceticum	37	0.09	ND[b]	0.30	64	NA
Acetobacterium woodii	30	0.24	<0.01	0.18	53	-
Desulfovibrio sp.	37	<0.01	<0.01	<0.01	42	NA
Desulfovibrio species	37-39	<0.03	<0.03	0.17	42	-
D. vulgaris (Hildenborough)	37	0.08	<0.01	0.07	37	-
D. desulfuricans (Norway 4)	33	0.16	ND	0.06	43	-
D. desulfuricans (ATCC 22774)	37	0.29	ND	<0.02	42	-
D. gigas	37	0.22	ND	0.10	37	-
Desulfotomaculum ruminus	37	0.11	ND	0.06	64	-
Escherichia coli B[c]	37	<0.01	<0.01	0.44	50	+
Micrococcus luteus	37	0.25	<0.01	<0.01	28	NA

[a] dam methylation was assayed and the symbols used are as in Table 1.

[b] ND, not detected by HPLC of the DNA digest in Solvent System 1 as described in Materials and Methods.

[c] Only this bacterium had been previously analyzed for its minor base composition (ref. 2). The m^6A content of E. coli B DNA, reported in ref. 20, was erroneous due to an incorrect molar response factor for m^6dAdo in our chromatography system.

trophicum are probably actually different subspecies or species as indicated by their very different major base composition even though they are both homoacetogens with the same optimal pH, temperature, and buffer requirements. They do differ in their pH and temperature range and in the ability of 701/5 but not 701/3 to use xylose as a carbon source. On the other hand, the base compositions of Clostridium sp. RB2 and RB3 were essentially identical (Table 1). These data combined with morphological and physiological similarities suggest that RB2 and RB3 may be the same strain or very closely related strains. Similarly, Clostridium thermoaceticum strains ATCC 39073 and 72 (Table 2) as well as strain wood (data not shown) have identical minor and major base compositions which could reflect their common origin (34,35).

For a comparison to the minor base composition of thermophilic bacterial DNAs, the DNA from a number of mesophiles was also examined by HPLC at the deoxynucleoside level (Table 3). As had been previously established (30,36-38), both mesophilic and thermophilic bacteria show a wide range of genomic A + T contents.

A few of the bacterial DNA digests, for example, the digest of T.

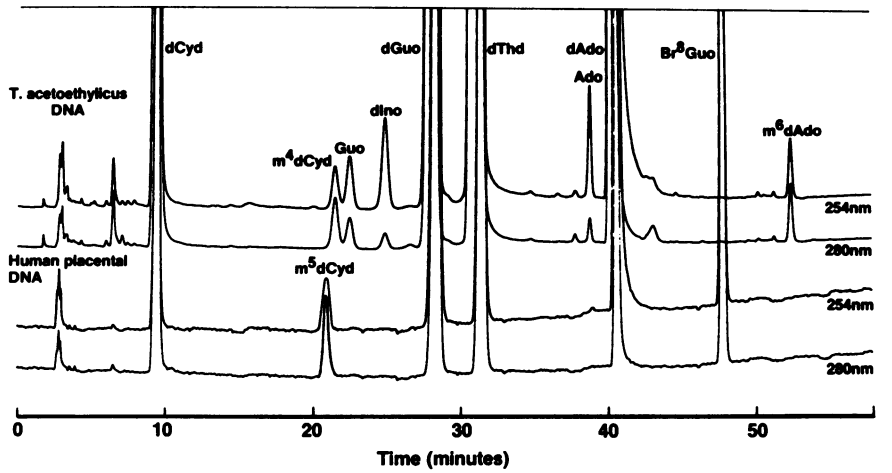

Fig. 1. HPLC of an enzymatic digest of T. acetoethylicus DNA. T. acetoethylicus DNA was purified by the standard methods described in the text. Like some of the other bacterial DNA samples, its digest contained a minor amount of ribonucleoside and dIno contaminants. The vertical axis shows absorbance at 254 or 280 nm with absorbance full scale as 0.02. The elution profiles shown in the figure are from one chromatographic run of the T. acetoethylicus DNA digest simultaneously monitored by absorbance at 254 and 280 nm and a separate run of a human placental DNA digest under identical conditions for comparison of m^5dCyd and m^4dCyd peaks. As an internal standard, 8-bromoguanosine (Br^8Guo) was included in the samples.

acetoethylicus DNA (Fig. 1), contained up to 6 mol% deoxyinosine (dIno). The content of dAdo, m^6dAdo, and dIno equalled that of dThd in these digests. Tests for deoxyadenosine deaminase activity in the enzymes used for digestion of these DNAs were negative so that this dIno may have resulted from degradation of dAdo by an unusually stable enzyme contaminating the DNA preparation. Isopycnic CsCl centrifugation of the routinely purified DNA samples resulted in essentially no dIno and negligible amounts of ribonucleosides being present in the ensuing digests (Fig.2).

N^6-Methyladenine Residues

Methylation of A residues is common in bacterial DNA. A compilation of various mesophilic bacterial strains whose DNA had been analyzed for m^6A showed that 23 out of 35 of these strains from various species contained this modified base (2). Of 12 other types of bacteria whose DNA methylases have been characterized as to their sequence specificity, six, including two thermophiles (Thermus aquaticus YTI and Thermus thermophilis), were shown to have enzymes that methylate A residues (3). In our study, ∿50% of the types of thermophilic bacteria including four types of archaebacteria and ∿70% of

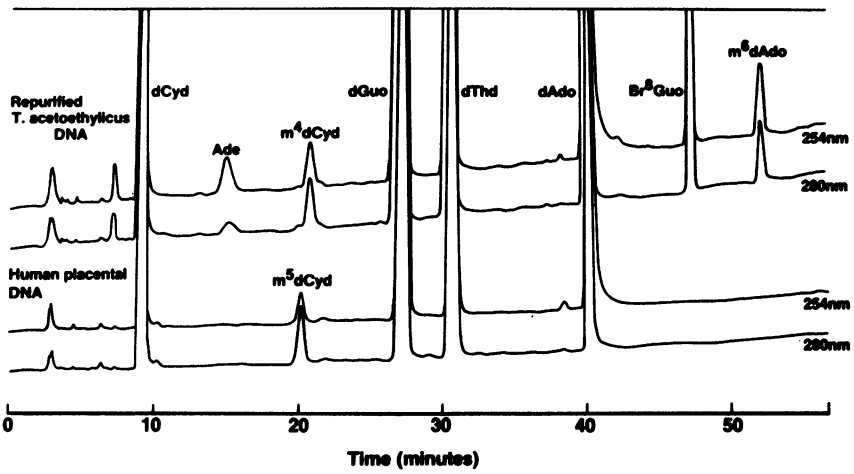

Fig. 2. HPLC of digested, repurified T. acetoethylicus DNA. The same batch of T. acetoethylicus DNA used for Fig. 1 was purified by isopycnic centrifugation on a cesium chloride gradient and quantitatively digested to deoxynucleosides. This digest and a human DNA digest were then chromatographed as in Fig. 1.

the types of mesophilic bacteria had m^6A in their DNA (Tables 1-3).

Since adenine methylation in GATC sequences in E. coli DNA has been shown to occur without an analogous restriction system and this methylation has been implicated in directing mismatch repair (5,14-16), we tested the thermophile DNAs for a similar type of methylation by assaying bacterial DNA for its susceptibility to digestion by DpnI, which cleaves only at Gm^6ATC sequences (39), and MboI, which cleaves only at GATC sites in which the A residue is unmethylated (4). From their resistance to MboI and hybridization to the cloned dam gene from E. coli, the genomes of seven species of the family Enterobacteriaceae and three species of Haemophilus were inferred to have dam-type DNA methylation unrelated to a restriction system (40). Four of the thermophilic bacterial DNAs which we tested were susceptible to hydrolysis by DpnI and not by MboI indicating a dam-type methylation pattern (Tables 1 and 2). However, as in the case of Moraxella bovis and, possibly, Anabaena variabilis (40), some of these bacteria might have a restriction modification system involving adenine methylation at GATC residues rather than E. coli's type of mismatch repair-associated methylation pathway.

The m^6A content of the four DpnI-sensitive DNA samples from thermophiles ranged from 0.08-0.93 mol% and the extent of digestion by DpnI varied in parallel. The 0.08 mol% m^6A found in T. acetoethylicus might indicate

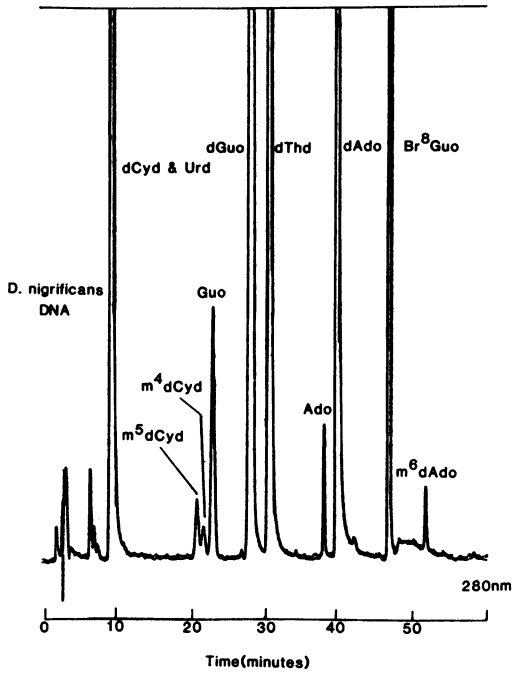

Fig. 3 HPLC of an enzymatic digest of D. nigrificans DNA. D. nigrificans DNA was purified by our standard method and then subjected to HPLC as in Fig. 1 with UV absorbance measured at 280 nm. The A_{254}/A_{280} ratios of the m^4dCyd and m^5dCyd peaks were the same as those of analogously analyzed authentic compounds. The ribonucleoside peaks were derived from an RNA contaminant constituting ∿2% of the nucleic acid in this sample. Two digests of other bacterial DNAs prepared and analyzed identically at the same time had equal levels of contaminating ribonucleosides but showed no peaks at the positions of m^4dCyd and m^5dCyd.

too low a frequency of Gm^6ATC in its genome for these sites to efficiently direct mismatch repair to the nascent strand rather than to the template strand (15,16). Although incubation with MboI did not give detectable digestion of this DNA, DpnI gave only a rather limited extent of digestion. These properties of T. acetoethylicus DNA and its low m^6A content in comparison to that of E. coli DNA (Table 3) suggest that GATC sequences are considerably underrepresented in this genome and those that are present are methylated at the A residue. On the basis of their relatively high extent of digestion by DpnI and their DNAs' m^6A content, we conclude that the tested C. thermoaceticum strains, C. thermosaccharolyticum, and C. thermoautotrophicum (701/3) have sufficiently high Gm^6ATC contents so that they might have E. coli's type of repair-associated dam methylation.

5-Methylcytosine Residues

Mesophilic bacterial DNAs often contain m^5C as a minor base. From 30 strains of various bacteria assayed for the minor base composition of their genomes, 18 contained m^5C in their DNA (2,6,41-47). Of the 14 mesophilic bacterial strains which we analyzed whose DNA's minor base composition had not been previously reported, eight had m^5C in their DNA (Table 3). Therefore, ∿60% of 44 strains of mesophilic bacteria studied by us or others had m^5C as a minor base in their DNA. Furthermore, ∿30% of the 256 characterized restriction endonucleases from different strains of bacteria have recognition sites with only C and G as invariant residues (3) implying that at least 30% (and presumably much more than 30%) of the bacterial hosts have methylated C residues in their DNA to prevent restriction of their own genomes.

In this first study of m^5C in thermophilic genomes, we found that only two types of thermophiles, Desulfotomaculum nigrificans and B. stearothermophilus H3, out of the 17 thermophiles analyzed, contained m^5C in their in DNA (Tables 1 and 2). Since m^5C residues in single-stranded DNA are deaminated at 95°C at three times the rate of C residues (8), many of the thermophiles may avoid having m^5C in their DNA in order to escape from the mutagenic consequences of $m^5C \rightarrow T$ transitions in transiently single-stranded regions of the genome. The only other way for these bacteria to bypass such a susceptibility to spontaneous (heat-induced) mutagensis at m^5C residues would be for them to have some unknown mechanism to depress heat-induced deamination of m^5C residues or to have some kind of mismatch repair system with a great preference for excising T residues at T•G mismatches (48). Whether D. nigrificans, which has an optimum growth temperature of 55°C, uses one of the above mechanisms or simply has an elevated spontaneous mutation rate due to its m^5C residues remains to be determined.

It is remarkable that six of the types of thermophilic bacteria and two of the mesophiles tested contained no detectable modified bases in their genomes (Tables 1-3). However, four of the five types of archaebacteria examined, Methanobacterium thermoautotrophicum, Methanococcus delta, Methanobacterium formicicum, and Methanococcus thermolithotrophicum contain m^6A in their DNA although methylated cytosine was found only in the genome of the latter two (Tables 1 and 3).

N^4-Methylcytosine Residues

In the course of analyzing the major base, m^5C, and m^6A content of DNA from T. acetoethylicus, Clostridium sp. RB1 and 9, C. thermoaceticum, and C.

themoautotrophicum strain 701/3, we noted an anomaly in the HPLC profiles of the deoxynucleosides in their enzymatic digests. In Solvent System 1, our standard chromatography system for DNA digests (19), a peak of UV light-absorbing material eluted with exactly the same retention time as for m^5dCyd but with an A_{254}/A_{280} ratio of 0.54 instead of 0.35. The latter ratio was characteristic of the m^5dCyd peak in hundreds of similarly chromatographed digests of mammalian DNA (32,49,50) and in digests of the m^5C-rich bacteriophage XP12 DNA (20), which we previously analyzed. The same absorbance ratio was found for the m^5dCyd peaks of the m^5C-containing mesophilic bacterial DNAs (Table 3).

The unusual compound in the DNA digests of T. acetoethylicus and in the above-mentioned Clostridium species and D. nigrificans and B. stearothermophilus strains H3 and N was resolved from m^5dCyd upon HPLC in Solvent System 2 (Fig. 1). Since this HPLC peak in the first thermophiles that we examined, Clostridium sp. RB1 and 9 and T. acetoethylicus, was very small compared to the other peaks (Fig. 1), we further purified the DNA to ascertain that the peak was not derived from a contaminant. DNA from these bacteria, which had been purified by our standard procedure and then isolated from a CsCl gradient centrifuged to equilibrium had the same content of the unknown peak upon digestion and HPLC as did samples not subjected to isopycnic centrifugation (Figs. 1 and 2). Also, a sample of C. thermoaceticum ATCC 39073 DNA which had been isolated by our standard procedure gave the same relative amount of the unusual HPLC peak as did the DNA further purified by hydroxyapatite chromatography. Further evidence that the novel peak was derived from the thermophile DNAs and not from a contaminant is that digestion of T. acetoethylicus DNA (with nuclease P1 alone) to deoxymononucleotides instead of to deoxynucleosides and HPLC in System 1 gave dAMP, dTMP, dGMP, dCMP, and one new minor peak with a different retention time than that of m^5dCyd (data not shown).

The UV absorption spectrum at pH 5.2 of the novel deoxynucleoside from T. acetoethylicus DNA had the same λ_{max}, 273 nm, and a similar, but not identical, shape to that of deoxycytidine. The corresponding λ_{max}, 280 nm, and UV absorption spectrum of m^5dCyd were quite different. Based on the spectral properties and chromatographic mobilities of the novel nucleoside we postulated that it was N^4-methyldeoxycytidine (m^4dCyd).

The novel nucleoside was isolated from an enzymatic digest of 2 mg of T. acetoethylicus DNA using our standard analytical reversed phase column and Solvent System #2, as described in Materials and Methods. A diode array

detector (1040A, Hewlett-Packard) was used in continuous monitoring of the UV spectrum of the eluted peak during collection to ensure the identity and homogeneity of this peak. Salt was removed by rechromatography of the collected fraction on the same column using 14% methanol in water as the mobile phase. This double chromatographically purified isolate was compared with chemically synthesized m^4dCyd prepared by a modification of the procedure of Wempen et al. (51; Gehrke et al., in preparation).

Under the two HPLC elution conditions that we used, the standard m^4dCyd had exactly the same retention time and UV absorbance spectrum as the unusual nucleoside from the DNA of T. acetoethylicus; C. thermoautotrophicum 701/3; C. thermoaceticum strains 72, ATCC 39073, and wood; Clostridium sp. RB1 and RB9; M. thermolithotrophicum, and B. stearothermophilus N (the source of the commercially available restriction endonuclease BstNI). In Solvent System 2 (at pH 5.2) the A_{254}/A_{280} ratio for m^4dCyd derived from these DNAs or synthesized chemically was 0.79. In contrast, under the same conditions m^5dCyd has a corresponding ratio of 0.52. Furthermore, trimethylsilyl (TMS) derivatives of the novel nucleoside from T. acetoethylicus DNA and the chemically derived reference compound had the same mass spectrum. Both had a molecular ion of 385 and a base + 1 ion of 126 when characterized by direct probe electron impact-mass spectroscopy. The TMS derivatives only formed at the 3' and 5' positions of the deoxynucleoside; no N^4-TMS group was found because of the steric hindrance of the N^4-CH_3. We conclude that the above thermophile DNAs, which contained no detectable m^5C, have m^4C residues instead. In addition, B. stearothermophilus H3 and D. nigrificans had both m^4C and m^5C in their genomes (Table 2; Fig. 3).

Analogous m^4C residues have not been found in RNA (52) and we have found no evidence for this base in vertebrate DNAs. Evidence for the presence of m^4C in CC(C/G)GG sequences in the DNA from Bacillus centrosporus, a mesophilic bacterium, and the prediction that B. stearothermophilus N would have m^4C in its genome were recently reported by Janulaitis and coworkers (53). It is noteworthy that the absence of other detectable modified bases in B. stearothermophilus N DNA (Table 1) implicates methylation of the amino group of a C residue at CC(A/T)GG sites in protection against restriction by BstNI in vivo.

The amounts of m^4C in the thermophilic bacterial DNAs that we examined ranged from 0.05-0.39 mol% (Tables 1 and 2). This range of m^4C contents is consistent with 4-methylation of C residues in various specific oligonucleotide sequences as part of restriction/modification systems. We propose an

evolutionary origin of m^4C residues in thermophilic bacteria so that the methylated cytosine residues would not be susceptible to heat-induced deamination to T residues as are m^5C residues. The recent discovery of an undefined cytosine derivative with similar properties to m^4C at the transposition-prone ends of trypanosome chromosomes (54) suggests that m^4C might not be limited to bacterial genomes.

ACKNOWLEDGEMENTS

We are very grateful to Dr. Ira Schildkraut and New England Biolabs Co. as well as to Drs. Jorgen Wiegel, Bill Whitman, Harry Peck, and Mouhoussini Nacro for supplying bacterial stocks or pellets.

This research was supported in part by National Science Foundation Grant PCM-8219000; the Research Assistance Act Higher Education Fund, University of Missouri, #315000-3 and #415002-1; and, at the University of Georgia, the U.S. Department of Energy Project De-A509-79ER10499.

REFERENCES

1. Brooks, J.E. and Roberts, R.J. (1982) Nucleic Acids Res. 10, 913-934.
2. Shapiro, H.S. (1976) In Handbook of Biochemistry, Sober, H.A. Ed., pp. 282-283, CRC Press, Cleveland.
3. Roberts, R.J. (1984) Nucleic Acids Res. 12, r167-r204.
4. McClelland, M. (1983) Nucleic Acids Res. 11, r169-r173.
5. Marinus, M.G. (1984) In DNA Methylation, Razin, A. Ed., pp. 81-109, Springer Verlag, New York.
6. Korch, C., Hagblom, P., and Normark, S. (1983) J. Bacteriol. 155,1324-1332.
7. Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Nature 274, 775-780.
8. Wang, R.Y.-H., Kuo, K.C., Gehrke, C.W., Huang, L.-H. and Ehrlich, M. (1982) Biochim. Biophys. Acta 697, 371-377.
9. Lindahl, T. (1979) Prog. Nuc. Acids Res. Molec. Biol. 22, 135-192.
10. Warner, H.R. (1983) J. Bacteriol. 154, 1451-1454.
11. Wagner, R.W., and Meselson, M. (1976) Proc. Natl. Acad. Sci. USA 73, 4135-4139.
12. Glickman, B., van den Elsen, P. and Radman, M. (1978) Molec. Gen. Genet. 163, 307-312.
13. Bale, A., d'Alarcao, M. and Marinus, M.G. (1979). Mutn. Res. 59, 157-165.
14. Glickman, B.W. and Radman, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1063-1067.
15. Lu, A.-L., Clark, S. and Modrich, P. (1983) Proc. Natl. Acad. Sci. 80, 4639-4643.
16. Pukkila, P.J., Peterson, J., Herman, G., Modrich, P. and Meselson, M. (1983) Genetics 104, 571-582.
17. Stetter, K.O., Konig, H., and Stackebraut, E. (1983) System. Appl. Microbiol. 4, 535-551.
18. Stetter, K.O. and Zillig, W. In The Bacteria, Gunsalus, J.C. Ed., Vol. 8. Academic Press, N.Y., in press.
19. Gehrke, C.W., McCune, R.A., Gama-Sosa, M.A., Ehrlich, M., and Kuo, K.C. (1984) J. Chromatog. 301, 199-219.
20. Kuo, K.C., McCune, R.A., Gehrke, C.W., Midgett, R., and Ehrlich, M. (1980) Nucleic Acids Res. 8, 4763-4776.

21. Postgate, J.R., Kentl, H.M., Robson, R.L., and Chesshyre, J.A. (1984) J. Gen. Microbiol. 130, 1597-1601.
22. Zeikus, J.G. and Wolfe, R.S. (1972) J. Bacteriol. 109, 707-713.
23. Mylorid, R.L. and Hungate, R.E. (1954) Canad. J. Microbiol. 1, 55-64.
24. Carreira, L.H. and Ljungdahl, L.G. (1984) in Liquid Fuels Development, Wise, D.L. Ed., pp. 2-29, CRC Series in Biotechnology, CRC Press, Boca Raton, Florida.
25. Ljungdahl, L.G. (1983) In Organic Chemicals from Biomass, Wise, D.L. Ed., pp. 219-248, Benjamin/Cummings Co., Menlo Park, California.
26. Corder, R.E., Hook, L.A., Larkin, J.M. and Frea, J.I. (1983) Arch. Microbiol. 134, 28-32.
27. Whitman, W.B., Ankwanda, E. and Wolfe, R.S., (1982) J. Bacteriol. 149, 852-863.
28. Huber, H., Thomm, M., Konig, H., Thies, G. and Stetter, K.O. (1982) Arch. Microbiol. 132, 47-50.
29. Peck, H.D. (1984) In Microbial Chemoautotrophy, Strohl, W.R. and Tuoninen, O.H. Eds., pp. 309-335, Ohio State Press, Columbus, Ohio.
30. Wiegel, J., Ljungdahl, L.G., and Rawson, J.R. (1979) J. Bacteriol. 139, 800-810.
31. Marmur, J. (1961) J. Mol. Biol. 3, 206-218.
32. Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A. and Gehrke, C. (1982) Nucleic Acids Res. 10, 2709-2721.
33. Schildkraut, C.L., Marmur, J. and Doty, P. (1962) J. Mol. Biol. 4,430-443.
34. Fontaine, F.E., Peterson, W.H., McCoy, E., Johnson, M.J., and Ritter, G.J. (1942) J. Bacteriol. 43, 701-715.
35. Ljungdahl, L.G. (1979) Advances Microbiol. Physiol. 19, 149-243.
36. Wiegel, J. and Ljungdahl, L.G. (1981) Arch. Microbiol. 128, 343-348.
37. Brock, T. (1978) Thermophilic Microorganisms and Life at High Temperatures, Spring-Verlag, New York.
38. De Ley, J. (1970) J. Bacteriol. 101, 738-754.
39. Vovis, G.F. and Lacks, S. (1977) J. Mol. Biol. 115, 525-538.
40. Brooks, J.E., Blumenthal, R.M., and Gingeras, T.R. (1983) Nucleic Acids Res. 11, 837-851.
41. Oda, K. and Marmur, J. (1966) Biochemistry 5, 761-773.
42. Degnen, S.T. and Morris, N.R. (1973) J. Bacteriol. 116, 48-53.
43. Razin, A. and Razin, S. (1980) Nucleic Acids Res. 8, 1383-1390.
44. Wachsman, J.T. and Irwin, V. (1970) J. Bacteriol. 104, 814-818.
45. Schein, A., Berdahl, B.J., Low, M., and Borek, E. (1972) Biochim. Biophys. Acta 272, 481-485.
46. Somogyi, P.A., Bel, M.M., and Foldes, I. (1982) Acta Microbiol. Acad. Sci. Hung. 29, 185-191.
47. Dybvig, K., Swinton, D., Maniloff, J. and Hattman, S. (1982) J. Bacteriol. 151, 1420-1424.
48. Ehrlich, M. and Wang, R.Y.-H. (1981) Science 212, 1350-1357.
49. Gama-Sosa, M.A., Midgett, R.M., Slagel, V.A., Githens, S., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983) Biochim. Biophys. Acta 740, 212-219.
50. Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983) Nucleic Acids Res. 11, 6883-6894.
51. Wempen, I., Duschinsky, R., Kaplan, L., and Fox, J.J. (1961) J. Am. Chem. Soc. 83, 4755-4762.
52. Sprinzl, M. and Gauss, D.H. (1984) Nucleic Acids Res. 12, r1-r57.
53. Janulaitis, A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983) FEBS Let. 161, 131-134.
54. Raibaud, A., Gaillard, C., Longacre, S. Hibner, U., Buck, G., Bernardi, G., and Eisen, H. (1983) Proc. Natl. Acad. Sci. USA 80,4306-4310.