Creation of genetic information by DNA polymerase of the thermophilic bacterium Thermus thermophilus

Norio Ogata* and Takanori Miura

Taiko Pharmaceutical Co., Ltd, Uchihonmachi 3-34-14, Suita, Osaka, Japan

Received June 25, 1998; Revised and Accepted September 7, 1998 DDBJ/EMBL/GenBank accession nos Y17475–Y17498

ABSTRACT

Genetic information encoded in a template of a genome is replicated in a complementary way by DNA polymerase or RNA polymerase with high fidelity; no creation of information occurs in this reaction unless an error occurs. We report here that DNA polymerase of the thermophilic bacterium Thermus thermophilus can synthesize up to 200 kb linear double-stranded DNA in vitro in the complete absence of added primer and template DNAs, indicating that genetic information is actively created by protein. This ab initio DNA synthesis occurs at 74°C and requires magnesium ion. There is a lag time of about 1 h and then the reaction proceeds linearly. The synthesized DNAs have a variety of sequences; they are mostly tandem repetitive sequences, e.g. (CATGTATA)_n, (TGTATGTA-TACATACATA)_n, and (TACCTC)n. Some degenerate sequences of these basic repeat units are also found. The similar repetitive sequences are found in many natural genes. These results, together with similar results found using DNA polymerase of archaeon Thermococcus litoralis, suggest that creative, non-replicative synthesis of DNA by protein was a driving force for diversification of genetic information at a certain stage of the evolution of life on the early earth.

INTRODUCTION

Genetic information is encoded as 5’ to 3’ direction-specific sequences of nucleotide bases in DNA or RNA strands that constitute the genome of many organisms (1,2). The information thus encoded in the genome is replicated with extremely high fidelity by reading each strand in a complementary way during a cell division or virus replication (3). The replication of the genetic material is catalyzed by DNA polymerase (4) or RNA polymerase (5). The former copies the pre-existing genetic information on a single-stranded template DNA with aids of pre-existing short primer DNA or primer RNA and four deoxyribonucleoside triphosphates (dNTPs) (3). Crucial points of this reaction are that no genetic information is made ab initio unless an error occurs; this is in fact an extremely rare event (6,7).

We have recently reported that DNA polymerase of Thermococcus litoralis, a hyperthermophilic anaerobic archaeon (archaebacterium) found in a submarine thermal vent in the Bay of Naples (8), can synthesize long stretches of DNA in the complete absence of added primer and template DNAs, demonstrating that genetic information is created by protein (9,10). This finding suggests that there is potential transfer of genetic information from protein to DNA. To further substantiate whether such creative synthesis of genetic information is a common phenomenon among DNA polymerases of many species (especially eubacteria), we have screened DNA polymerase of many species and have found that DNA polymerase of Thermus thermophilus (Tth), a thermophilic aerobic bacterium (eubacterium) discovered in a hot spring (Mine Spa) in Izu Peninsula (Japan) by Oshima and Imahori (11), can also synthesize very long stretches of complex DNA sequences in the complete absence of added primer and template DNAs. We report here the details of the reaction by this DNA polymerase and characterization of the reaction product together with the implication of this phenomenon in terms of the evolution of genetic information on the early earth. In the accompanying paper (10) we demonstrate that the nucleotide sequences of the DNA thus synthesized without added primer and template by DNA polymerase of archaeon are markedly influenced by environmental factors.

MATERIALS AND METHODS

DNA accession numbers

The EMBL DNA accession numbers of the sequences of the 24 clones in this paper are Y17475–Y17498.

Tth DNA polymerase reaction without primer and template DNAs

The standard reaction mixture (20 µl) contained 5.2 ng (0.5 U) of Tth DNA polymerase, which is recombinant and produced in Escherichia coli (Boehringer Mannheim, 99% pure as determined by SDS–PAGE) unless otherwise specified, in a ‘polymerase buffer’ containing 50 mM KCl, 10 mM Tris–HCl buffer (pH 9.0 at 25°C and pH 7.5 at 74°C), 1.5 mM MgCl2, 0.1% (w/v) Triton X-100 and dNTPs (200 µM each of dATP, dTTP, dCTP and dGTP). In some experiments, [α-32P]dNTP was added at a final specific activity of 36.6 MBq/µmol. The mixture was incubated...
at 74°C for various time intervals specified and the reaction was terminated by adding 1 μl of 500 mM EDTA (pH adjusted to 8.0 with NaOH). The mixture was then electrophoresed on a 0.8% agarose gel and stained with ethidium bromide. A liquid scintillation counter. The gel was subsequently stained under ultraviolet illumination.

Reaction with pretreated Tth DNA polymerase

The reaction mixture contained 5.2 ng of Tth DNA polymerase, or 10 μg of yeast tRNA (Sigma) and/or 480 ng of HindIII-cut λ phage DNA (Toyobo) as a control, in 20 μl of the polymerase buffer with 25 ng of deoxyribonuclease I (DNase I) (Boehringer Mannheim) and/or 20 ng of ribonuclease A (RNase A) (Worthington). The mixture was first incubated at 37°C for 2 h and then at 74°C for another 3 h. Each sample was subsequently electrophoresed and stained as above.

Characterization of a Tth DNA polymerase reaction product

The Tth DNA polymerase reaction product was prepared in the standard reaction mixture for 3 h as described above, except that the scale of the reaction was 1000-fold (20 ml), and the product was then purified as described. A 620 ng aliquot of the reaction product was digested with 1 U of DNase I, 0.26 U of S1 nuclease (Takara) or 1 U of Bal31 exonuclease (Toyobo) at 37°C for 10 min in 10 μl of a solution recommended by each manufacturer. As a control, 240 ng of HindIII-cut λ phage DNA or 200 ng of M13mp18 phage single-stranded DNA was likewise digested. Each reaction mixture was then electrophoresed and stained as above. Another 100 μg aliquot of the reaction product was cloned as described, and sequenced by a dideoxy chain-termination method on both strands. DDBJ, EMBL and GenBank data searches were carried out by a BLAST search.

RESULTS

DNA-like substance (hereafter called ‘pol product’) was synthesized by Tth DNA polymerase in the absence of added primer and template DNAs as judged by an agarose gel electrophoresis and 32P-incorporation in acid-insoluble material (Fig. 1). The amount and the size of the pol product increased when the reaction time increased; the size was 0.1–200 kb after 4 h (Fig. 1A). When the time course of the reaction was measured as acid-insoluble radioactivity incorporated in macromolecules from [α-32P]dNTP, there was a lag time of ~1 h; only 2.3 fmol of dATP and 2.7 fmol of dTTP were incorporated, while dCTP or dGTP was not incorporated in the acid-insoluble material in the first hour. The reaction then proceeded almost linearly and all four dNTPs were incorporated; a total of 350 pmol (2.2% of the substrates used) dNTPs were incorporated after 4 h (Fig. 1B). The nearly stoichiometric ratios of A:T and G:C in acid-insoluble material suggest the formation of A–T and G–C base pairs as are found in common double-stranded DNA; this finding suggests that the pol product is double-stranded DNA. The GC content of the pol product was calculated to be 25.3% after 4 h (Fig. 1B).

We next examined whether the pol product is synthesized by Tth DNA polymerases of different sources. All the Tth DNA polymerases, one native and two recombinants (expressed in E.coli), could synthesize the pol product, albeit the amounts of the pol product synthesized differed markedly between different sources (Fig. 2). The reaction of the pol product synthesis absolutely required magnesium ion. The pol product was not detected without it, but was detected in the presence of 1.5–10 mM; the size of the pol product became shorter and the amount of pol product synthesized increased when the concentration of magnesium ion increased (Fig. 3). The synthesis of the pol product was observed even though
in a single stretch of DNA. Such 'hybrids' of the repetitive unit clones in which at least two kinds of the repetitive units were present sequenced had tandem repeat sequences, there were some each repeat unit were 8–36 bases. Although most of the pol product mixture containing various concentrations of MgCl₂ was incubated for 3 h at 74°C, and electrophoresed and stained as described in Figure 1. DNA size markers are shown on the left in kb.

The reaction mixture was pretreated with DNase I and/or RNase A (Fig. 4). The result strongly excludes the possibility that the pol product is synthesized owing to a small amount of DNA or RNA that might have been present as a contaminant in the reaction mixture. In other words, the result suggests that the reaction of the pol product synthesis actually occurs in the absence of added primer and template DNAs.

Next we characterized the pol product. When the pol product was treated with endonuclease DNase I or exonuclease Bal31, it was digested almost completely (Fig. 5). On the other hand, it was not digested with single strand-specific endonuclease S1 nuclease, although the enzyme could completely digest M13mp18 phage single-stranded DNA used as a control under the same reaction condition. Based upon the substrate specificities of the enzymes employed, these results indicate that the pol product is indeed DNA and is double-stranded and linear.

To determine the nucleotide sequence of the pol product DNA, it was partially digested with DNase I in the presence of manganese ion to generate blunt-ended DNA molecules of appropriate sizes and then cloned into a plasmid vector pUC19 in its unique blunt-end cutter site. We obtained 24 insert-positive clones. After sequencing, the lengths of the insert DNA were 26–98 bp. Two clones, pTH279 and pTH952, showed, for an unknown reason, double bases (T and C) at nucleotides 23 and 41, respectively, on a sequence ladder. We assigned one of them by comparison with template DNAs by nuclease P1 (24.0%, data not shown). This fact supports the idea that there is almost no 'cloning bias', i.e. cloning of only those DNAs of the clones calculated as double-stranded (total 2724 bases) from all clones was calculated to be 23.5%, while when the GC content was calculated for each clone individually and an average of 24 clones was taken, it was 23.5 ± 2.3%, demonstrating that the variations of the GC contents among the clones are very small. In addition, these values agreed well with those calculated from 32P-incorporation in pol product (25.3%, Fig. 1), and determined by an analysis of the pol product by column chromatography after complete hydrolysis into four mononucleotides by nuclease P1 (24.0%, data not shown). This fact supports the idea that there is almost no 'cloning bias', i.e. cloning of only those populations of DNA molecules that are preferentially cloned, and that the clones obtained are real representatives of the pol product DNA. The sizes of the insert DNAs were very short, although we expected their lengths to be 0.5–2 kb from the cloning procedure employed. Nearest neighbor frequency (Table 1) of the insert DNAs of the clones calculated as double-stranded (total 2724 neighbours) clearly demonstrated that AA, TT, GG, CC, TC and GA sequences never appeared, and that AG, GC and CT sequences never appeared, and that AG, GC and CT sequences appeared in extremely low frequency. When a similarity search of the sequences of the pol product was carried out against natural DNAs in DNA databases in EMBL, DDBJ and GenBank with BLAST search (14), very similar repetitive sequences were found in genes of many organisms (data not shown). These sequences appear mostly in non-coding regions and introns of natural genes.

The lengths of each repeat unit were 8–36 bases. Although most of the pol product sequences were found in clones pTH298, pTH674 and pTH975. It is to be noted that the DNA sequence of some clones was not necessarily a perfect repeat but there was some degeneracy in the sequence, e.g. nucleotide 24 in pTH262. It is worth noting that most of the repeat unit sequence has a structure of a palindrome. For instance, the sequence (CATGTATAₙ)ₙ, has the center of the palindrome between italicized A and T. The GC content of the total bases sequenced (1386 bases) from all clones was calculated to be 23.5%, while when the GC content was calculated for each clone individually and an average of 24 clones was taken, it was 23.5 ± 2.3%, demonstrating that the variations of the GC contents among the clones are very small. In addition, these values agreed well with those calculated from 32P-incorporation in pol product (25.3%, Fig. 1), and determined by an analysis of the pol product by column chromatography after complete hydrolysis into four mononucleotides by nuclease P1 (24.0%, data not shown). This fact supports the idea that there is almost no 'cloning bias', i.e. cloning of only those populations of DNA molecules that are preferentially cloned, and that the clones obtained are real representatives of the pol product DNA. The sizes of the insert DNAs were very short, although we expected their lengths to be 0.5–2 kb from the cloning procedure employed. Nearest neighbour frequency (Table 1) of the insert DNAs of the clones calculated as double-stranded (total 2724 neighbours) clearly demonstrated that AA, TT, GG, CC, TC and GA sequences never appeared, and that AG, GC and CT sequences appeared in extremely low frequency. When a similarity search of the sequences of the pol product was carried out against natural DNAs in DNA databases in EMBL, DDBJ and GenBank with BLAST search (14), very similar repetitive sequences were found in genes of many organisms (data not shown). These sequences appear mostly in non-coding regions and introns of natural genes.
The nearest neighbour frequency was calculated from all the insert DNAs (as double-stranded) in the pol product clones.

DISCUSSION

Most, if not all, of the genetic information, whether DNA or RNA and whether small or large, is now believed to be a 5′ to 3′ direction-specific defined sequence of four kinds of nucleotide bases. This sequence of the bases is replicated by DNA polymerase or RNA polymerase from generation to generation in many organisms. Our results demonstrate that such defined sequences can be made ab initio by protein in the absence of pre-existing genetic information encoded in DNA or RNA. Such protein-catalyzed creation of genetic information was first reported regarding RNA synthesized by RNA polymerases of Azotobacter vinelandii (15), E.coli (16), Qφ phage (17,18) and T7 phage (19). However, in the first two the enzymes used were only partially purified and the possibility of contaminating DNA and/or RNA, which might have worked as a primer or a template, is not excluded (15,16). Regarding the latter two cases, the implication of the findings in terms of the early appearance and evolution of genetic information is limited, because the enzymes are derived from phages, which have somewhat special mechanisms of replication and must have appeared after host organisms had appeared on the earth. In addition, the product RNAs are only 60–120 bases long (19), Schachman et al. (20) reported synthesis of alternating co-polymer polyd(AT) and Radding et al. (21) reported synthesis of homopolymer polydG/polydC without added primer and template DNAs using partially purified DNA polymerase of E.coli. However, we could not detect such ‘de novo’ synthesis (20,21) of DNA with highly purified enzyme of this bacterium (unpublished data). Thus the meaning of their observations remains unclear unless the results are reproduced by a highly purified enzyme preparation. Contrary to the above enzymes, those of T.thermophilus (bacterium) and T.thermosphaera (archaeon) (9,10), which are 99 and >95% pure, respectively, would tell us about what happened on the early earth. The fact that DNA polymerases of both of these species can synthesize DNA ab initio, suggests that such creative DNA synthesis by protein actually occurred on the earth by some primitive organism(s) that appeared before divergence of archaeon and bacterium.

Tandem repetition of a short repeat unit is found in many genes. For instance, it is found in a coding region of a fibron gene of silk worm (Bombbyx mori) (22); an 18-base repeat unit sequence represented as GGTGCTGGTCGTGTTCA, which is translated to Gly-Ala-Gly-Ala-Gly-Ser, appears very frequently in its coding region as tandem clusters. At a region of nucleotide 37–180, a perfect repeat of the unit sequence appears eight times in tandem (22). In a gene of antifreeze protein of an antarctic cod Notothenia coriiceps neglecta, most of the coding region starting at nucleotide 112 and ending at the termination codon is composed of almost perfect 46 tandem repeats of a 51-base (coding a stretch of 17 amino acids) repeat unit (23). A tandem repetitive sequence in the noncoding region is well known in telomeres, in which a sequence TTAGGG appears as a unit sequence (24). Such a tandem repetitive structure in many genes might have arisen not by gene duplication as commonly believed about many genes, e.g. globin genes (25), but by ab initio DNA synthesis as shown in this paper. Ohno proposed that the first set of coding sequences that arose in the prebiotic world were repeats of nucleotide oligomers (26,27). In the light of his idea, it may be plausible that a primitive polypeptide having polymerase-like activity, synthesized long stretches of DNAs as simple repetitive sequences and that they then gradually ‘evolved’ into degenerate sequences by accumulating mutations during replication by primordial enzyme that might have been error-prone at that time.

In the accompanying paper (10) we demonstrate that DNA polymerase of archaeon T.litoralis can synthesize repetitive DNAs ab initio, where the lengths of the repeat units are 4–18 nucleotides, and that the nucleotide sequence of each repeat unit differs markedly depending upon a change of temperature; e.g. the sequences are (TATCTAGA)_6 (25% GC content) at 74°C and (GATCCG)_6 (67% GC content) at 94°C (10). This means that if such ‘temperature sensitivity’ holds true about DNA polymerase of T.thermophilus, a complex and long stretch of DNA may have arisen by changes in the environmental temperature. We have recently found that DNA polymerase of thermophilic bacteria Pyrococcus sp. strain KOD1 (KOD DNA polymerase) (28) and Thermus flavus ubiquitus (Tub DNA polymerase, Amersham) (29) can also synthesize DNA in the absence of added primer and template DNAs (unpublished data), demonstrating that creative synthesis of DNA by protein is a fairly common phenomenon at high temperature. It may be time now to think about the
importance of protein-driven creation of genetic information in the 'protein world' before we think about diversification of RNA genes in the so-called RNA world.

REFERENCES