Nucleic Acids Research, 1998, Vol. 26, No. 20 **4657–4661**

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 ~1 h and then the reaction proceeds linearly. The synthesized DNAs have a variety of sequences; they are mostly tandem repetitive sequences, e.g. $(CATGTATA)_n$, $(TGTATGTATACATACATA)_n$ and $(TATACGTA)_n$. 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 $MgCl_2$, 0.1% (w/v) Triton X-100 and dNTPs (200 µM each of dATP, dTTP, dCTP and dGTP). In some experiments, [α-^{32}P]dNTP was added at a final specific activity of 36.6 MBq/µmol. The mixture was incubated

*To whom correspondence should be addressed. Tel: +81 6 382 3100; Fax: +81 6 382 1152

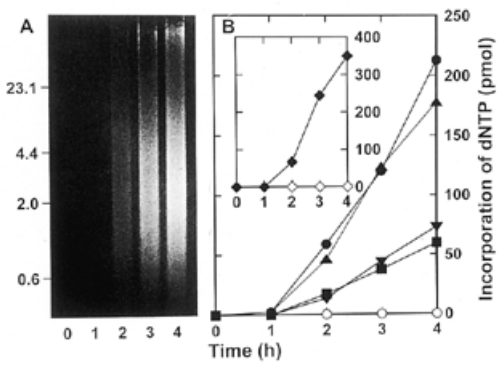

Figure 1. A time course of DNA synthesis without added primer and template DNAs by *Tth* DNA polymerase. (**A**) A standard reaction mixture (20 µl) (except that *Tth* DNA polymerase was 0.25 U) was incubated for various time intervals at 74°C, electrophoresed on an agarose gel and stained with ethidium bromide. (**B**) The reaction mixture as in (A) [except that one (dATP, circles; dTTP, triangles; dCTP, squares; or dGTP, inverted triangles) or all (diamonds) of dNTPs were labelled with ^{32}P in their α positions] was incubated at 0°C (open symbols) or 74°C (filled symbols) for various time intervals, and the radioactivity incorporated in acid-insoluble material was next counted. DNA size markers are shown on the left in kb.

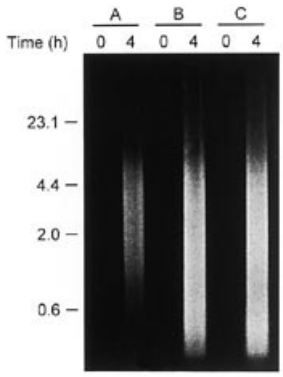

Figure 2. DNA synthesis without added primer and template DNAs by *Tth* DNA polymerase of various sources. *Tth* DNA polymerase (0.5 U) of various commercial sources was incubated at 74°C for 0 or 4 h in a standard reaction mixture, electrophoresed on a 0.8% agarose gel and stained with ethidium bromide. The *Tth* DNA polymerases used were (A) recombinant and expressed in *E.coli* (Perkin-Elmer), (B) native (Promega) and (C) recombinant and expressed in *E.coli* (Boehringer Mannheim). DNA size markers are shown on the left in kb.

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 (Seakem GTG, FMC) as described (12), or its 5% (w/v) trichloroacetic acid-insoluble radioactivity was determined after filtration on a glass microfiber filter (Whatman) (12) using a liquid scintillation counter. The gel was subsequently stained with 0.5 µg/ml ethidium bromide solution and photographed under ultraviolet illumination (12).

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 *Hin*dIII-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 (12). 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 *Hin*dIII-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 (9), and sequenced by a dideoxy chain-termination method (13) on both strands. DDBJ, EMBL and GenBank data searches were carried out by a BLAST search (14).

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 ^{32}P-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-^{32}P]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

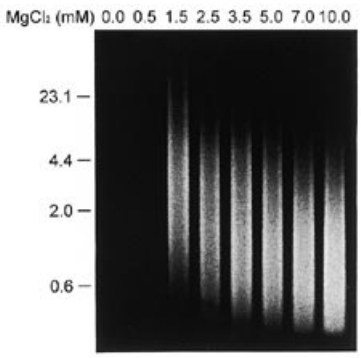

Figure 3. Effect of magnesium ion on DNA synthesis without added primer and template DNAs by *Tth* DNA polymerase. A standard reaction mixture containing various concentrations of $MgCl_2$ was incubated for 3 h at 74°C, and electrophoresed and stained as described in Figure 1. DNA size markers are shown on the left in kb.

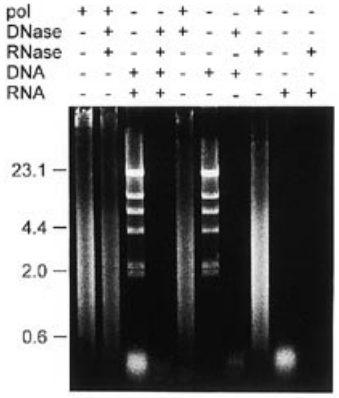

Figure 4. DNA synthesis without added primer and template DNAs by *Tth* DNA polymerase pretreated with DNase I and/or RNase A. A standard reaction mixture containing *Tth* DNA polymerase (pol) and dNTPs was pretreated with or without DNase I (DNase) and/or RNase A (RNase) at 37°C for 2 h, and next incubated at 74°C for 3 h. As a control, *Hin*dIII-cut λ phage DNA (DNA) and/or yeast tRNA (RNA) was likewise pretreated and next incubated at 74°C for 3 h. After the reaction, the mixtures were 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 supports 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 *Sma*I (a 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 neighbouring sequences. The majority of the sequences of the pol product clones showed tandem repeats of various unit lengths and, interestingly enough, purine and pyrimidine bases appeared alternately in many regions of the sequences. The most frequent sequences found were $(CATGTATA)_n$, $(TGTATGTATACATACATA)_n$, $(TATACGTA)_n$, $(TGTATGTATATACATACA)_n$, $(TGTACATATA)_n$, $(TATACGTATA)_n$, $(TGTATACATATA)_n$ and $(TGTATGTATACATACATACGTATGTATACATACATA)_n$. The lengths of each repeat unit were 8–36 bases. Although most of the pol product clones sequenced had tandem repeat sequences, there were some clones in which at least two kinds of the repetitive units were present in a single stretch of DNA. Such 'hybrids' of the repetitive unit 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 $(CATGTATA)_n$ (clone pTH752), which can also be written as $(TGTATACA)_n$, 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 ^{32}P-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.

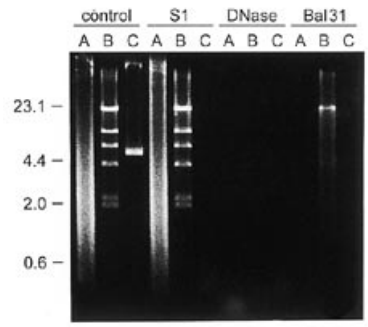

Figure 5. Characterization of DNA product synthesized without added primer and template DNAs by *Tth* DNA polymerase. (A) DNA product (pol product), (B) *Hin*dIII-cut λ phage DNA or (C) M13mp18 phage single-stranded DNA was digested without (control) or with S1 nuclease (S1), DNase I (DNase) or Bal31 exonuclease (Bal31), electrophoresed and stained as described in Figure 1. DNA size markers are shown on the left in kb.

Table 1. Nearest neighbour frequency of the DNA synthesized by *Tth* DNA polymerase without added primer and template DNAs

Frequency (%)		Frequency (%)	
AA	0.0	GA	0.0
AT	26.2	GT	11.3
AG	0.04	GG	0.0
AC	11.3	GC	0.2
TA	30.0	CA	8.9
TT	0.0	CT	0.04
TG	8.9	CG	3.0
TC	0.0	CC	0.0

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 of the bases 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), Qβ 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.litoralis* (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 early 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 fibroin gene of silk worm (*Bombyx mori*) (22); an 18-base repeat unit sequence represented as GGTGCTGGTGCTGGTTCA, 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 TTGGGG 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 $(TATCTAGA)_n$ (25% GC content) at 74°C and $(GATCGC)_n$ (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 ubiquitos* (*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

1 Alvager,T., Graham,G., Hilleke,R., Hutchinson,D. and Westgard,J. (1989) *Biosystems*, **22**, 189–196.
2 Zull,J.E. and Smith,S.K. (1990) *Trends Biochem. Sci.*, **15**, 257–261.
3 Kornberg,A. and Baker,T.A. (1992) *DNA Replication*, 2nd ed. Freeman, NY.
4 Goodman,M.F. and Fygenson,K.D. (1998) *Genetics*, **148**, 1475–1482.
5 Shilatifard,A. (1998) *Biol. Chem.*, **379**, 27–31.
6 Bridges,B.A. (1996) *Cancer Surv.*, **28**, 155–167.
7 Hampson,R. (1997) *Radiat. Oncol. Investig.*, **5**, 111–114.
8 Belkin,S. and Jannasch,H.W. (1985) *Arch. Microbiol.*, **141**, 181–186.
9 Ogata,N. and Miura,T. (1997) *Biochem. J.*, **324**, 667–671.
10 Ogata,N. and Miura,T. (1998) *Nucleic Acids Res.*, **26**, 4652–4656.
11 Oshima,T. and Imahori,K. (1974) *Int. J. Syst. Bacteriol.*, **24**, 102–112.
12 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
14 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
15 Krakow,J.S. (1967) *Proc. Natl Acad. Sci. USA*, **58**, 2094–2101.
16 Biebricher,C.K. and Orgel,L.E. (1973) *Proc. Natl Acad. Sci. USA*, **70**, 934–938.
17 Biebricher,C.K., Eigen,M. and Luce,R. (1986) *Nature*, **321**, 89–91.
18 Biebricher,C.K. and Luce,R. (1993) *Biochemistry*, **32**, 4848–4854.
19 Biebricher,C.K. and Luce,R. (1996) *EMBO J.*, **15**, 3458–3465.
20 Schachman,H.K., Adler,J., Radding,C.M., Lehman,I.R. and Kornberg,A. (1960) *J. Biol. Chem.*, **235**, 3242–3249.
21 Radding,C.M., Josse,J. and Kornberg,A. (1962) *J. Biol. Chem.*, **237**, 2869–2876.
22 Mita,K., Ichimura,S., Zama,M. and James,T.C. (1988) *J. Mol. Biol.*, **203**, 917–925.
23 Hsiao,K.-C., Cheng,C.-H.C., Fernandes,I.E., Detrich,W.H. and DeVries,A.L. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 9265–9269.
24 Greider,C.W. (1991) *Mol. Cell. Biol.*, **11**, 4572–4580.
25 Hardison,R. and Miller,W. (1993) *Mol. Biol. Evol.*, **10**, 73–102.
26 Ohno,S. (1987) *J. Mol. Evol.*, **25**, 325–329.
27 Ohno,S. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6486–6490.
28 Takagi,M., Nishioka,M., Kakihara,H., Kitabayashi,M., Inoue,H., Kawakami,B., Oka,M. and Imanaka,T. (1997) *Appl. Environ. Microbiol.*, **63**, 4504–4510.
29 Kainz,P., Schmiedlechner,A. and Strack,H.B. (1992) *Anal. Chem.*, **202**, 46–49.