

Molecular Cloning, Expression and Sequence Analysis of DNA Polymerase I from an Iranian Thermophilic Bacterium, *Bacillus* sp. G (2006)

E. Farashahi Yazd^a, M. Sadeghizadeh^{a,*}, S. Hosseinkhani^b, M. Khalaj-Kondori^a and R. Emamzadeh^b

^aDepartment of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, P.O. Box 14115-175, Tehran, Islamic Republic of Iran

^bDepartment of Biochemistry, Faculty of Biological Sciences, Tarbiat Modares University, P.O. Box 14115-175, Tehran, Islamic Republic of Iran

(Received 3 October 2008, Accepted 1 January 2009)

Thermostable DNA polymerases are widely used in DNA amplification reactions such as the Polymerase Chain Reaction (PCR), requiring the activity of the enzymes at high temperatures. The aim of the present study was to assess the potential biotechnological capabilities of Iranian thermostable DNA polymerases. To this end, we cloned the gene encoding a DNA polymerase from a novel thermophilic eubacterium, *Bacillus* sp. G (2006). Phylogenetic analysis of this gene revealed that the new isolate belongs to the genera *Bacillus*. Sequence analysis of the fragment produced by degenerate primers also showed that it consists of 2,631 bp encoding an 876 amino acid protein, and subsequent amino acid sequence analysis of this DNA polymerase showed that it belongs to family A-type DNA polymerases. The expression vector pET28a (+) was chosen for expression of the gene fragment in the mesophilic host bacterium *E. coli* BL21. This expression vector has some advantages such as attachment of a Poly-His tag to the N-terminus of the protein for the ease of purification and a powerful promoter of *lac-Z* induced by IPTG. The band corresponding to the protein product was observed in the molecular weight range of about 100KDa on the SDS-PAGE gel after heat and Ni²⁺-NTA column chromatography. Using the dot blot technique, the polymerase activity of the enzyme was qualitatively confirmed at 70 °C. Therefore, it is suggested that optimizations of this activity could make this enzyme appropriate for PCR processes in future.

Keywords: Cloning, Expression, Thermophile, DNA polymerase I

INTRODUCTION

Thermophilic and hyperthermophilic enzymes offer major biotechnological advantages over mesophilic enzymes [1]. The most salient reaction in which thermostable enzymes are widely used is the PCR reaction utilizing thermostable DNA polymerases [2]. The uses and advantages of employing thermophilic DNA polymerases for reverse transcription and

reverse transcription-coupled PCR amplifications (RT-PCR) have also been described [3].

Since the discovery and characterization of DNA polymerase I from *E. coli* by Kornberg and colleagues in the 1950s [4], a large number of DNA polymerases have been isolated and identified from prokaryotic and eukaryotic sources. *Taq* DNA polymerase from *Thermus aquaticus* was the first characterized thermostable enzyme [5]. Thereafter, many DNA polymerases from the *Thermus* strain were studied. *Tfl*, *Tth*, *Tfi*, and *Taq* polymerases have been applied

*Corresponding author. E-mail: sadeghma@modares.ac.ir

to PCR [6-9]. However, their base insertion fidelity is low due to the lack of the 3' → 5' exonuclease activity. High-fidelity of DNA polymerases, associated with the 3' → 5' exonuclease-dependent proofreading activity, is required for error correction during polymerization. Several thermostable DNA polymerases with proofreading activity (*Pfu*, *Vent*, *deep Vent*, and *Pwo*) were also studied and introduced for high fidelity PCR amplification [10-14]. DNA polymerases are classified into five major groups based on the alignment of their amino acid sequences [15,16]: families A, B, C, D and X represented by *E. coli* DNA polymerase I, II, III (α subunit), *Pyrococcus furiosus* DNA polymerase II, and eukaryotic DNA polymerase β , respectively [7].

Taq DNA polymerase and some other thermostable DNA polymerases possess measurable reverse transcriptase (RT) activity [17]; however, the prominent thermostable DNA polymerase with a remarkable reverse transcriptase activity has been identified from *Thermus thermophilus* which has had applications in RT-PCR [3]. Therefore, the presence of a thermostable DNA polymerase is an indispensable ingredient of PCR technology, and the increasing number of applications utilizing PCR has generated an increasing demand for various thermostable DNA polymerases [18]. In this study, we describe the identification, cloning, sequence analysis, activity measurement and homology modeling of deduced amino acid sequence of a novel thermostable DNA polymerase I derived from a thermophilic eubacterium isolated from a hot spring at Gheinarcheh, Ardabil, Iran (The Gheinarcheh hot spring, with a temperature of 85 °C, is the hottest hot spring in Iran.) [19].

EXPERIMENTAL

Bacterial Strains and Culture Conditions

Bacillus sp. G (2006) was kindly supplied by Dr. Fooladi, Microbiology Group, Alzahra University, Tehran, Iran. The cells were grown in a medium containing 3 g of bacto-peptone, 5 g of bacto-yeast extract, and 1 g of NaCl (per liter), and the pH was adjusted to 6.5 with (10 N) H₂SO₄. The cells were then inoculated with a syringe and grown aerobically at 70 °C for 2 days without shaking.

E. coli DH5 α was used for plasmid propagation and nucleotide sequencing. *E. coli* BL21(DE3) (Stratagene, USA), was used for gene expression. The *E. coli* cells were grown in

Luria-Bertani medium with appropriate antibiotics at 37 °C with vigorous shaking.

Genomic DNA Extraction & Molecular Phylogeny

The genomic DNA of *Bacillus* sp. G (2006) was extracted by the method of Marmur [20] with slight modifications. For the amplification of partial fragment of the 16S rRNA gene, two primers were used, Fu16: 5'-CCTACGGGAGGCAGCAG-3' as the forward and Ru16: 5'-GACGTCRTCCNCDCTTCCTC-3' as the reverse primers where R = (A/G), D = (A/G/T), N = (A/C/G/T). DNA amplification was carried out using 1.25 U of *Taq* DNA polymerase in a 25 μ l reaction mixture, consisting of 10 ng of *B. Sp. G (2006)* genomic DNA, 20 pmol of each primer, 200 μ M dNTPs, and PCR reaction buffer. The PCR reaction was performed under the following conditions: initial denaturation at 96 °C for 4 min, a 35-cycle amplification (94 °C for 30 s, 58 °C for 30 s, and 72 °C for 50 s), and a final extension for 8 min at 72 °C. A predicted ~850 bp PCR product was then purified by gel extraction kit (Bioneer Co., South Korea) from 1% agarose gel and the nucleotide sequence was determined using a 3700 ABI Capillary system (Macrogen Inc., South Korea). Alignment of the gene sequences was performed using Clustal W with a Gap Weight of 1.0 and a Gap Length Weight of 0.1. The multiple sequence alignment consisted of 850 nucleotides from the 16S rDNA gene sequence. Alignments of the gene sequences were subjected to parsimony analysis using AliBee-Multiple Alignment software [21] to generate the evolutionary distance matrix and phylogenetic tree. Sequences of the strains used in this study were taken from GenBank (Fig. 1). The 16S rDNA sequence of *Bacillus* Sp. G (2006) was deposited in GenBank under the accession number; DQ785316.

DNA Polymerase I Gene Amplification & Construction of Expression Plasmid

For the amplification of a *Bsg* DNA polymerase I gene fragment, two degenerate primers were designed based on both ends of DNA polymerase I gene sequence of the closest bacterial species to *Bacillus* sp. G (2006) [22], (GenBank accession no: DQ870752.1, NC_006510.1, AY952967.1, CP000557.1, M77484.1), which were aligned using Clustal W software, and two primers were designed, F1BS: 5'-

Cloning of a Thermostable DNA Polymerase

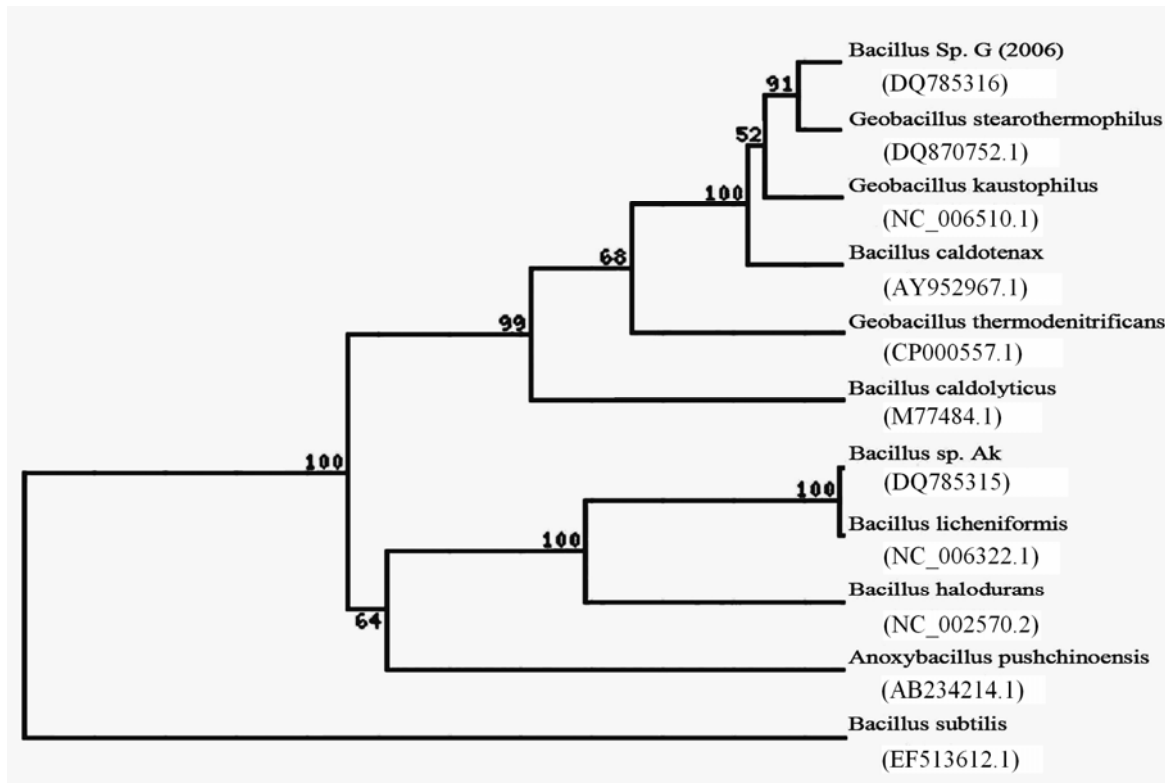


Fig. 1. Phylogenetic position of *Bacillus* sp. G (2006), the new isolate, among some other thermophilic members of family *Bacillaceae* based on the 16S rDNA partial sequences.

ATGAAAAAMAARCT GGTCTTAATTGAYGGCAA-3' as the forward (N-terminal) primer, and RBS2: 5'-TT ATTCGCATCATAACCATGTTGGGCCGTAA-3' as the 3' (C-terminal) primer. DNA amplification was carried out using 1.25 U of *Pfu* DNA polymerase in a 25 μ l reaction mixture, consisting of 10 ng of *Bacillus* Sp. G (2006) genomic DNA, 20 pmol of each primer, 200 μ M dNTPs, and the PCR reaction buffer. The PCR reaction was performed under the following conditions: initial denaturation at 96 $^{\circ}$ C for 4 min, a 35-cycle amplification (94 $^{\circ}$ C for 30 s, 47 $^{\circ}$ C for 30 s, and 72 $^{\circ}$ C for 2 min), and a final extension for 8 min at 72 $^{\circ}$ C. The amplified fragments were purified by gel extraction kit (Bioneer Co., South-Korea) from 1% agarose gel, and used for the following PCR steps as the template DNA. For the expression of the *Bsg* DNA polymerase I, the DNA polymerase gene was amplified by another new primer set: the 5' (N-terminal) primer, F3gpol

I: 5'-GCCGGGATCCATGAAAAACAAGTTGGTCTTAATTG-3' adding a unique *Bam*HI site, and the 3' (C-terminal) primer, R3gpol I: 5'-CTTCGCGGCCGCT TATTTTGCATCATAACCATGTTG-3' adding a unique *Not*I site to the amplified product. DNA amplification was carried out using 1.25 U of *Pfu* DNA polymerase (Fermentas, Lithuania) in a 25 μ l reaction mixture resembling the first PCR. The amplified fragment containing the *Bsg* DNA polymerase I gene was digested with *Bam*HI and *Not*I, purified by gel extraction kit (Bioneer Co., South-Korea) from 1% agarose gel, and ligated into the expression vector pET-28a(+) (Novagen, USA), which had been digested with the same enzymes, according to the previously reported method [12]. *E. coli* BL21(DE3) was transformed with the ligated product by CaCl₂ precipitation method. Clones with the correct construct were selected by the restriction enzyme analysis of plasmid minipreparations.

Purification of the Expressed *Bsg* DNA Polymerase I in *E. coli*

5 ml of the overnight culture of *E. coli* BL21(DE3) harbouring the relevant expression plasmid, grown in Luria-Bertani broth containing 50 $\mu\text{g ml}^{-1}$ kanamycin, was transferred to 500 ml of the same medium. This culture was grown at 37 °C until an A_{600} of 1.2 was reached. Gene expression was induced with 1.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG), and the culture was grown for another 9 h. The cells were harvested by centrifugation and resuspended in lysis buffer (NaH₂PO₄ 50 mM, NaCl 300 mM, Imidazole 10 mM, pH 8.0) containing 1 mM phenylmethylsulfonyl fluoride (PMSF). The resuspended cells were disrupted by sonication and the majority of heat labile *E. coli* proteins were precipitated by heat treatment at 60 °C for 20 min. The supernatant was applied onto a Ni²⁺-NTA chromatography column (Qiagen, USA), pre-equilibrated with washing buffer (NaH₂PO₄ 50 mM, NaCl 300 mM, imidazole 20 mM, pH 8.0). The column was washed with elution buffer (NaH₂PO₄ 50 mM, NaCl 300 mM, Imidazole 250 mM, pH 8.0). Major fractions containing the desired protein were pooled, dialyzed against enzyme storage buffer and analyzed with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [23].

DNA Sequencing and Sequence Analysis

The new DNA polymerase I gene sequence was submitted to GenBank (GenBank accession no. EF198253). All the handling and the analysis of nucleotide sequences were performed with tools available at the ExPASy Molecular Biology Server. Pair-wise and multiple sequence alignments were carried out using the Needleman-Wunch algorithm and Clustal W software, respectively. A model of the three dimensional structure of DNA polymerase I from *Bacillus* sp. G (2006) was predicted using Geno3D server [24,25] and the crystal structure of the large fragment of DNA polymerase I from *Bacillus stearothermophilus* within the Protein Data Bank (pdb 2hhvA) was used as the template. Analysis and comparison of the structures were carried out using the Swiss-Pdb Viewer software [26].

DNA Polymerase Activity Measurement

In order to confirm the polymerase activity of the

recombinant enzyme, dot blotting technique [27] using the probes synthesized by the recombinant *Bsg* was performed. Initially, the basic reaction mixture of samples (purified recombinant *Bsg* DNA polymerase I, *Bacillus* sp. G (2006) cell extract supernatant, Mesophilic transformed host cell extract supernatant, negative control and Klenow fragment with specific activity of 2 unit/ μl as the positive control) contained 3 μl (700 ng μl^{-1}) of the template DNA, 2 μl of random hexanucleotide, 2 μl of dNTP (1 mM each of dATP, dCTP, and dGTP, 0.65 mM dTTP, 0.35 mM DIG-11-dUTP pH 7.5), enzyme solution (7 μl , with the concentration of 216 mg/ml for the purified recombinant enzyme), ddH₂O (6 μl) and 2 μl of the Klenow enzyme (instead of the enzyme solution) as the positive control. All the samples except the positive control were incubated at 70 °C and the positive control was incubated at 37 °C for 2 h. The product of this reaction was used as the probe of hybridization step in the dot blot technique. Five aliquots of solution containing the template DNA (belonging to the plasmid pET28a(+)) 10 ng μl^{-1} were spotted separately onto a nitrocellulose paper (23 mm, Whatman, UK). The spotted nitrocellulose paper was washed in SCC Buffer (0.015 M sodium citrate and 0.15 NaCl) for 10 min and then dried in an oven at 80 °C. The synthesized probes were added to the spots for hybridization and incubated overnight. Subsequent vigorous washes were carried out using washing buffer (Maleic acid 0.1 M, NaCl 0.15 M, Tween 20 0.3% v/v). Anti-Digoxigenin-AP was added to the spots followed by substrate (NBT/BCIP) addition to detect the labeled probes.

RESULTS AND DISCUSSION

Molecular Phylogeny of *Bsg* DNA Polymerase I

The phylogenetic tree, which represents the deduced position of *Bacillus* sp. G (2006), the new thermophilic isolate, among some other members of *Bacillaceae* family, is depicted in Fig. 1. According to the phylogenetic tree, the new thermophilic isolate is in the same sister taxon of *Geobacillus stearothermophilus* (Fig. 1). Phylogenetic and distance matrix analysis (data not shown) revealed that the new isolate is related to the genus *Bacillus*; so, it was named *Bacillus* sp. G (2006).

Sequence Analysis and Homology Modeling of the *Bsg* DNA Polymerase I Gene

Sequence analysis revealed a 2631 nucleotide open reading frame (ORF) encoding an 876 amino acid polypeptide with an estimated mass of 99.42 KDa. Amino acid sequence analysis of *Bacillus* sp. G (2006) DNA pol I, using the BLASTp program, indicated 99% identity of this DNA pol I to DNA pol I of *Bacillus caldolyticus* strain *EA.1* (GenBank accession no. AAR11867). Amino acids variations were noted between *Bsg* DNA polymerase I and DNA polymerase of *B. caldolyticus* *EA.1*. According to the pair-wise alignment results, two residues of the *Bsg* DNA polymerase I, Arg 2 and Val 3, are omitted compared to that of *B. caldolyticus* *EA.1*, and the residues, Ser 13, Ala 166, Lys 208, Val 262, Val 548, Ser 872 are replaced by Asn, Thr, Arg, Ala, Ile and Pro, respectively. MotifScan analysis showed crucial motif and domains including DNA polymerase A signature, the 5' → 3' polymerase and 5' → 3' exonuclease domains, but not the typical 3' → 5' exonuclease domain. The presence of DNA polymerase A signature was verified using ScanProSite (702-RrqAKavnFGivYgiSdygL-721). It can, therefore, be proposed that *Bsg* polymerase I belongs to the family A of DNA polymerases.

The ribbon model of 3D structure for *Bsg* DNA polymerase I has been predicted using the deduced amino acid sequence, and the crystal structure of *Bacillus stearothermophilus* DNA polymerase I [28] large fragment (BF) (pdb 2hhvA) as template (Data not shown). Pair-wise sequence alignment between BF and the corresponding region of *Bacillus* sp. G (2006) DNA polymerase I (GF) revealed 99% sequence identity, and, as expected, the folding pattern strongly resembled the template. The root mean square (rms) deviation of C α positions between all residues of GF and BF was 0.07 Å, reflecting comparatively local and total structural similarities.

Expression and Purification of *Bsg* DNA Polymerase I

The *Bsg* DNA polymerase I gene was amplified and inserted into the *Bam*HI and *Not*I site of pET-28a(+). The resulting expression plasmid was designated pEBGP. *E. coli* BL21(DE3) cells harboring pEBGP were grown in a 1-liter LB medium. *Bsg* DNA polymerase I was expressed in a soluble form in the cytosol. As mentioned earlier, the

harvested cells (2 ml) were initially sonicated and most *E. coli* proteins were eliminated using the heat-resistance property of *Bsg* DNA polymerase I by heating at 60 °C for 20 min, and then all the denatured proteins were removed by centrifugation. Several *E. coli* proteins still remained soluble after the heating step. The soluble supernatant from the heating step was then applied on a Ni²⁺-affinity chromatography column. The purification of the enzyme was monitored by SDS-PAGE which revealed a single sharp protein band with a molecular mass of ~100 KDa that is consistent with the calculated molecular mass, 99.42 KDa, of *Bsg* DNA polymerase I from the deduced amino acid sequence (Fig. 2).

Bsg DNA Polymerase I Properties

The closest amino acid sequence to *Bsg* DNA polymerase I according to the result of BLAST software is DNA polymerase I from *Bacillus caldolyticus*. The preceding studies on the DNA polymerase I of *B. caldolyticus* showed its instability at temperatures above 70 °C [29]. Therefore, it is suggested that the recombinant DNA polymerase is also unstable at this temperature for a long time in a process such as PCR.

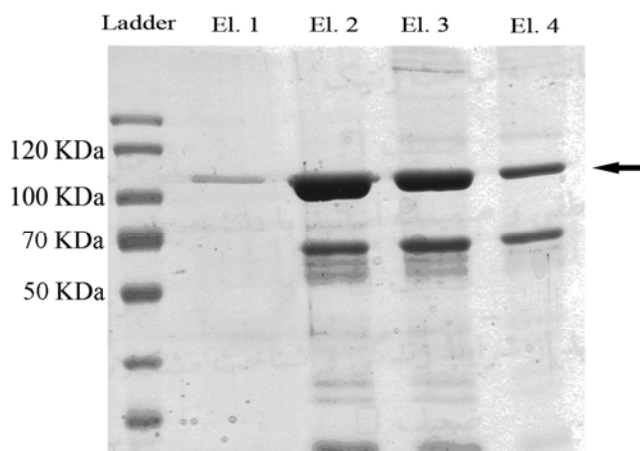


Fig. 2. SDS-PAGE analysis of recombinant *Bsg* DNA polymerase I. Lanes El.1, El.2, El.3, El.4 samples were collected at different stages of the elution phase of Ni²⁺-NTA chromaography. The arrow shows the defined band.

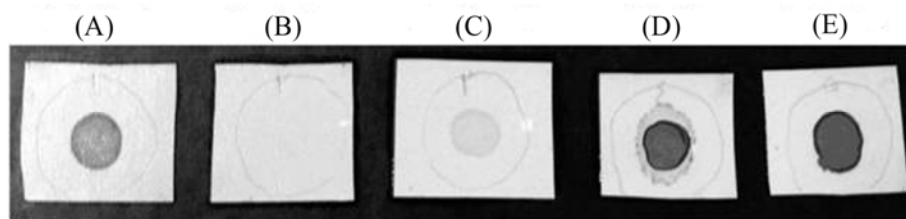


Fig. 3. Activity measurement of *Bsg* DNA polymerase I. (A) Purified recombinant *Bsg* DNA polymerase I: The blot shows labeled probes (synthesized by purified recombinant *Bsg* DNA polymerase I at 70 °C) hybridized to the fixed template DNA on nitrocellulose paper, (B) Negative control: In this sample no DNA polymerase enzyme was used in the probe synthesis stage and no blot is observed, (C) Mesophilic transformed host cell extract supernatant: The blot shows labeled probes (synthesized by recombinant *E. coli* (BL21) cell extract supernatant at 70 °C) hybridized to the fixed template DNA, (D) *Bacillus* sp. G (2006) cell extract supernatant : The blot shows labeled probes (synthesized by *Bacillus* sp. G (2006) cell extract supernatant at 70 °C) hybridized to the fixed template DNA, (E) Positive control: The blot shows labeled probes (synthesized by Klenow fragment at 37 °C) hybridized to the fixed template DNA.

Dot blotting technique was applied for DNA polymerase activity measurement of *Bsg* polymerase at 70 °C (Fig. 3), and the polymerase activity of this enzyme was confirmed.

As mentioned in the literature, there are differences in activities of native and recombinant proteins. The explanations for these differences in general, point to the effects of extrinsic factors such as special chaperones, salts and other cellular compounds which are characteristics of these thermophilic microorganisms [1].

Our data confirm the polymerase activity of the enzyme *Bsg* DNA polymerase I qualitatively at 70 °C. It is suggested that further quantitative analysis and optimization of the conditions would elucidate the potential capability of this enzyme for use in PCR experiments. It is also predicted, based on the closest DNA polymerase sequence, that this enzyme may possess a noticeable reverse transcriptase activity which remains to be investigated.

ACKNOWLEDGEMENTS

This work was supported by Ministry of Mines and Industries, High Tech. Department (Iran) grant 84/8/337 to M. S. We also thank Dr. S. Gharavi and A. Padeganeh for the reading of the manuscript.

REFERENCES

- [1] C. Vieille, G.J. Zeikus, *Microbiol. Mol. Biol. Rev.* 65 (2001) 1.
- [2] R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich, *Science* 239 (1988) 487.
- [3] T.W. Myers, D.H. Gelfand, *Biochemistry* 30 (1991) 7661.
- [4] A. Korenberg, T. Baker, *DNA replication*, Freeman Company, New York, 1992.
- [5] Chien, D.B. Edgar, J.M. Trela, *J. Bacteriol.* 127 (1976) 1550.
- [6] S. Kaledin, A.G. Sliusarenko, S.I. Gorodetskii, *Biokhimiia* 46 (1981) 1576.
- [7] C. Ruttimann, M. Cotoras, J. Zaldivar, R. Vicuna, *Eur. J. Biochem.* 149 (1985) 41.
- [8] S.E. Jung, J.J. Choi, H.K. Kim, S.T. Kwon, *Mol. Cells* 7 (1997) 769.
- [9] J.S. Kim, S. Koh, J.J. Kim, S.T. Kwon, D.S. Lee, *Mol. Cell* 8 (1998) 157.
- [10] N.F. Cariello, J.A. Swenberg, T.R. Skopek, *Nucleic Acids Res.* 19 (1991) 4193.
- [11] K.S. Lundberg, D.D. Shoemaker, M.W. Adams, J.M.

Cloning of a Thermostable DNA Polymerase

- Short, J.A. Sorge, E.J. Mathur, *Gene* 108 (1991) 1.
- [12] H. Kong, R.B. Kucera, W.E. Jack, *J. Biol. Chem.* 268 (1993) 1965
- [13] B. Frey, B. Suppmann, *Biochemica* 2 (1995) 34.
- [14] J. Cline, J.C. Braman, H.H. Hogrefe, *Nucleic Acids Res.* 24 (1996) 3546.
- [15] D.K. Braithwaite, J. Ito, *Nucleic Acid Res.* 21 (1993) 787.
- [16] I.K. Cann, Y. Ishino, *Genetics* 152 (1999) 1249.
- [17] M.D. Jones, N.S. Foulkes, *Nucleic Acids Res.* 17 (1989) 8387.
- [18] H.J. Shin, S.K. Lee, J.J. Choi, S. Koh, J.H. Lee, S.J. Kim, S.T. Kwon, *J. Microbiol. Biotechnol.* 15 (2005) 1359.
- [19] A. Saffarzadeh, Y. Noorollahi, *Proceedings World Geothermal Congress, Turkey, 2005.*
- [20] J. Marmur, *J. Mol. Biol.* 3 (1961) 208.
- [21] http://www.genebee.msu.su/services/malign_reduced
- [22] M. Khalaj-Kondori, M. Sadeghizadeh, K. Khajeh, H. Naderi-Manesh, A.M. Ahadi, A. Emamzadeh, *Appl. Biochem. Biotechnol.* 142 (2007) 200.
- [23] U.K. Laemmli, *Nature* 227 (1970) 680.
- [24] C. Combet, M. Jambon, G. Deleage, C. Geourjon, *Bioinformatics* 18 (2002) 213.
- [25] http://geno3dpbil.ibcp.fr/cgi-bin/geno3d_automat.plpage=/GENO3D/geno3d
- [26] <http://www.expasy.org/spdbv>
- [27] M.J. McPherson, S.G. Moller, PCR, BIOS Scientific Publishers, Oxford, 2000.
- [28] J.J. Warren, L.J. Forsberg, L.S. Beese, *Proc. Natl. Acad. Sci. USA* 103 (2006) 19701.
- [29] H. Shandilya, K. Griffiths, E.K. Flynn, M. Astatke, P.J. Shih, J.E. Lee, G.F. Gerard, M.D. Gibbs, P.L. Bergquist, *Extremophiles* 8 (2004) 243.