Chapter 147
Elongation of Trinucleotide Repeats by DNA Polymerase

Yang Wang, Ping Dong and Xingguo Liang

Abstract The wide present simple repetitive sequences in genomes are significant for studying molecular evolution, genetic diversity, and some hereditary diseases. Here, elongation of 64 (3^4) kinds of 18 nucleotides long sequences with trinucleotide tandem repeats by DNA polymerase was studied. The result showed that all of the repeats, apart from several strands made up with single nucleotide, could be elongated, although the efficiency depended on the sequences and reaction temperatures. More GC content required a higher temperature for efficient elongation. For double strands, elongation products had a narrow size distribution, and the length of synthesized DNA increased linearly with reaction time. Single-stranded sequences could also be elongated with even higher efficiency, and products longer than 10 kb were obtained in several hours. Our results are promising to be used for explanation of molecular evolution, as well as unusual amplification of repetitive sequences during gene amplification and detection.

Keywords Repetitive sequences · Isothermal amplification · Molecular evolution

147.1 Introduction

Tandem short repetitive sequences are strands of DNA made up with short units connected with each other, which are also known as satellite DNA [1, 2], such as $(AT)_n$ and $(CAG)_n$. They are widely distributed in eukaryotic and prokaryotic organisms and show polymorphism in length even in closely related individuals.

Y. Wang · P. Dong · X. Liang (✉)
College of Food Science and Engineering, Ocean University of China, Qingdao 266003, People's Republic of China
e-mail: liangxg@ouc.edu.cn

T.-C. Zhang et al. (eds.), *Proceedings of the 2012 International Conference on Applied Biotechnology (ICAB 2012)*, Lecture Notes in Electrical Engineering 251,
DOI: 10.1007/978-3-642-37925-3_147,

Tandem repeats are closely related to many aspects of biological and medical science like molecular evolution, genetic diversity, molecular markers, and genetic diseases [3]. It has been revealed that more than 10 serious diseases, such as Huntington's disease and Myotonic dystrophy, are caused by expansion of trinucleotide repetitive sequences [4]. As the result of easy mutation of these short repeats, they contribute greatly to biological evolution and diversity of species. Moreover, these repeats have already been widely used as gene specific markers [5]. On the other hand, although human genome project has just been completed 10 years ago, what we know about repetitive sequences is definitely not enough to understand their abundant presence as well as to solve relevant problems. Accordingly, in-depth study of these repetitive sequences, especially the mechanism of abnormal expansion is required.

There are many proposed mechanisms for repetitive sequence expansion, such as intra or intermolecular recombination, slipped strand mispairing, and illegitimate elongation model [6–8]. Ogata studied the amplification of repeats with palindromic sequences, considering that formation of hairpin structures plays an important role in the initiation stage for elongation, and further elongation is carried out in a slippage style [9, 10]. After noticing these mechanisms were inadequate to explain the high efficiency of expansion, especially at low concentrations, Liang et al. proposed a mechanism to show that the hairpin formation at the end of a DNA duplex contributed to the efficient and continuous elongation, although further experiments were required to clarify this model [11, 12]. At presents, we believe that expansion of repetitive sequence may be carried out according to various mechanisms, which depend on the sequences and the reaction conditions.

Up to now, some scattered researches have shown that many simple repetitive sequences such as $(TA)_n$, $(GAA)_n/(TTC)_n$, $(CCTG)_n$, and $(dG)_n/(dC)_n$ could be expanded by polymerase to 10 kb long [13, 14]. However, no reasonable explanation was provided. On the other hand, repetitive sequences were synthesized ab initio (i.e., in the absence of any initial nucleic acid) from dNTPs by thermophilic DNA polymerases [15]. More and more results supported that modern coding sequences of DNA might evolve from primordial simple sequence repeats. Thus, for both clarifying the mechanism of synthesis of repetitive sequences and understanding their biological significance, systematic study on elongation of simple repetitive sequences is essential. Here, abnormal expansion of all of the 18 nt trinucleotide repetitive sequences by DNA polymerase is studied comprehensively for understanding the expansion mechanism, providing some hints for molecular evolution, as well as explaining the nonspecific amplification in gene detection and diagnose.

147.2 Materials and Methods

147.2.1 Materials

Sixty four kinds of sequences of 18 bp trinucleotide repeats used in this study were purchased from Integrated DNA Technologies. $(AGC)_6$ and $(GCT)_6$ indicate the single-stranded oligonucleotides; and $(AGC)_6/(GCT)_6$ and $(ATA)_6/(TAT)_6$. indicate corresponding duplexes. The oligonucleotides were dissolved in H_2O to 100 μM as stock solutions. The oligos were then diluted to 1.0 μM as the stock for elongation (to a final concentration of 100 nM in reaction). Vent and Vent (exo^-) DNA polymerase, restriction enzyme ApeKI (recognition sequence: G↓CWGC, W = A or T), AluI (recognition sequence: AG↓CT), SmaI (recognition sequence: CCC↓GGG), plasmid PUC18, and RNase-free DNase I were purchased from New England Biolabs. Deoxyribonucleoside Triphosphates (dNTPs) was obtained from Tiangen, T4 DNA ligase was from Fermentas.

147.2.2 Elongation of Short Repetitive Sequences

In principal, standard reaction solution contains 50 nM or 100 nM short repeats, 0.5 mM each dNTPs and 20 U/mL Vent or Vent (exo^-) DNA polymerase. 1 × Thermolpol buffer contains 20 mM Tris-HCl, 10 mM $(NH4)_2SO_4$, 10 mM KCl, 2.0 mM $MgSO_4$, 0.1 % Triton X-100, pH 8.8 at 25 °C. For elongation of double strands, the repeats were pretreated by incubating in 1 × ThermolPol buffer at 90 °C for 3 min followed by slowly cooling to room temperature. The reaction mixture were added to the 96 well plates and performed on life express Thermal cycler (BIOER). Aliquots of the reaction solution were next loaded on a 0.8 % nondenaturing agarose or 8.0 % polyacrylamide gel and then stained with ethidium bromide.

147.2.3 Digestion of Elongated Products by Restriction Enzyme

Elongated products were purified by PCI (phenol: chloroform: isoamyl alcohol = 25:24:1), CIA (chloroform: isoamyl alcohol = 24:1) treatment and precipitated with ethanol. The products were then digested by restriction enzymes ApeKI and AluI for sequence analysis. The reaction was carried out at 75 °C by 100 U/mL ApeKI in 20 μL 1 × NEBuffer 3 (100 mM NaCl, 50 mM Tris–HCl, 10 mM $MgCl_2$, 1.0 mM Dithiothreitol, pH 7.9 at 25 °C) and at 37 °C by 250 U/mL AluI in 20 μL 1 × NEBuffer 4 (50 mM KAc, 20 mM Tris-Ac, 10 mM $Mg(Ac)_2$, 1.0 mM DTT, pH 7.9 at 25 °C).

147.2.4 Molecular Cloning of the Elongated Products

After purification by PCI and CIA treatment and precipitated with ethanol, elongated products were digested slightly with 0.4 U/mL of DNase I in the presence of $MnCl_2$ to generate fragments with blunt ends. The products were then purified again, cloned into the SmaI site of a PUC18 plasmid vector. The strands of the insert DNA were sequenced by the dideoxy chain termination method.

147.2.5 T_m Measurement

1.0 μM 18 nt or 18 bp short repeats were added to 1 × Thermopol Detergent Free buffer (10 mM KCl, 10 mM $(NH4)_2SO_4$, 20 mM Tris-HCl, 2.0 mM $MgSO_4$, pH 8.8 at 25 °C). Change of absorbance at 260 nm with temperature was recorded on UV-1800 equipped with temperature-controlling device (Shimadsu).

147.3 Results and Discussion

147.3.1 Elongation of Single-Stranded Sequences by DNA Polymerase

Elongation of all the single strands of repetitive sequences were first performed at 70 °C using 100 nM short repeats seeds in the presence of 0.5 mM dNTPs by polymerase. The elongation products were analyzed by 0.8 % agarose gel. For single strands, The sequences with higher GC content, such as $(TCG)_6$ were inclined to expand efficiently. Products up to 10 kb were observed in some cases within 1 h (Fig. 147.1). For most of other sequences, this kind of phenomenon was also observed. Similar results were also obtained when a denaturing agarose gel (30 mM NaOH) was used.

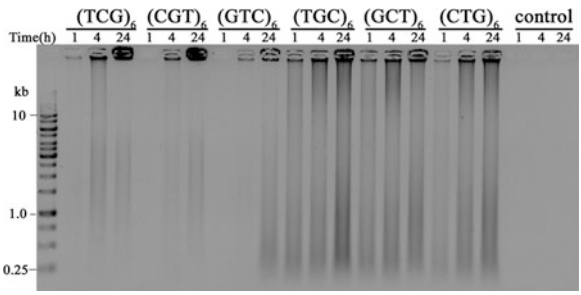

Fig. 147.1 Gel electrophoresis analysis of expansion products of several single strands repeats at 70 °C for 4 h

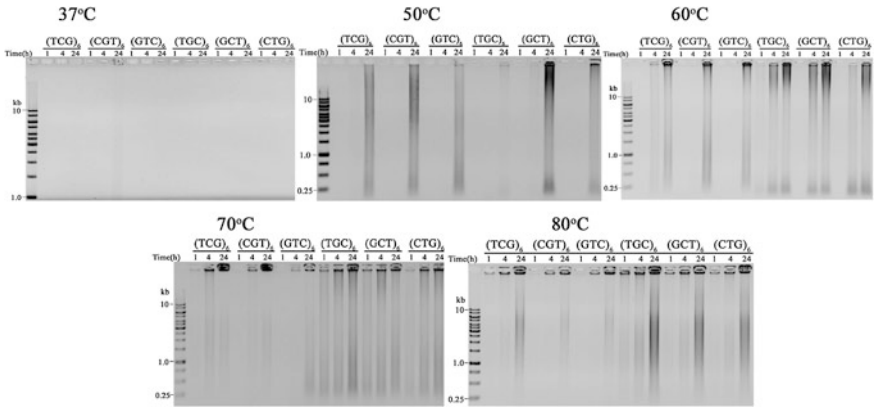

Fig. 147.2 Extensions of single-stranded short trinucleotide repeats at various temperatures

Next, we checked the effect of reaction temperature on the elongation of short repeats. As shown in Fig. 147.2, for $(TCG)_6$ very long DNA was obtained within 1 h at 70 and 80 °C; at 50 and 60 °C, the speed of elongation became slower, and products could only be observed after 4 or 24 h; at 37° C, basically no elongation was observed under the conditions we used. Other short repeats were also analyzed and similar results were obtained, showing that temperature had great effect on their expansion. Repeats with different GC content have different optimal reaction temperatures. For all the short repeats, the suitable temperatures were increased according to the increasing GC content.

147.3.2 Elongation of Double-Stranded Sequences by DNA Polymerase

To catalyze single strands elongation, we used Vent (exo^-) to substituted Vent, because there have been reports describing primer and template-free de novo synthesis of DNA by Vent DNA polymerase [10, 16, 17], especially in high temperature (70–80 °C) for several hours. However, we found that short duplex with repetitive sequence could produce nearly uniform long DNA with Vent polymerase better than with Vent (exo^-). So for double strands repeats, Vent DNA polymerase was used in our study, and the smear band at high temperature can be regarded as de novo synthesis or single strands elongation.

For double strands repeats, the products after elongation had a narrow size distribution (Fig. 147.3), especially for those with more AT content, such as $(AAG)_6$/$(AGG)_6$ and $(ATG)_6$/$(CTA)_6$. When the GC content is 2/3, although more smeared products were observed, the narrow bands around 500 bp could also be clearly seen. The repeats with 100 % AT contents did not give lots of products at 70 °C after 4 h of reaction; on the other hand, except for $(GGG)_6$/$(CCC)_6$, short

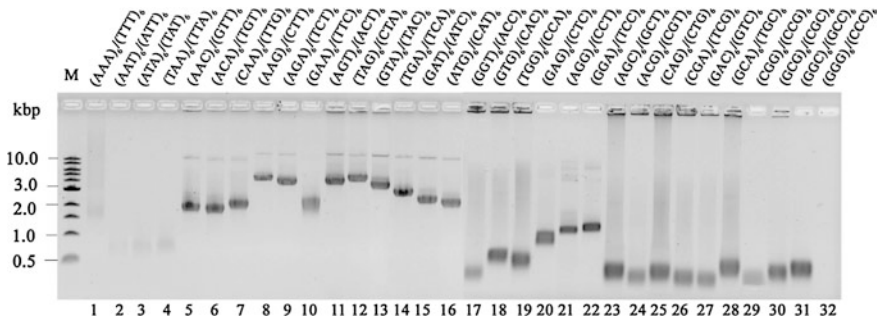

Fig. 147.3 Gel electrophoresis analysis of expansion products of 32 duplexes of trinucleotide repeats at 70 °C for 4 h. 1–16: Short repeats with 0 or 1/3 GC content; 17–32: short repeats with 2/3 or 100 % GC content. Reaction conditions: 50 nM short repeats, 10 U/mL Vent, 0.5 mM dNTPs, in 20 μL of 1 × Thermopol Buffer. The products were analyzed on a 0.8 % agarose gel

repeats with 100 % GC content can be elongated and narrow bands were observed (Fig. 147.2). Obviously, the average elongation speed of each repeats changed with the variation of the sequences. Repeats with 1/3 GC content grew fastest, with a speed of more than 8.0 bp/min. Repeats which have the same GC content but different sequence grew similarly.

In order to find out the effect of temperature to amplification double stands, only duplex $(AGC)_6/(GCT)_6$ and $(GAT)_6/(ATC)_6$ were studied specially. For $(AGC)_6/(GCT)_6$ at all temperatures, narrow bands were observed, although smear bands were also present in some cases. At 37 and 50 °C, the reactions were very slow and weak bands could only be observed after 20 h; at a temperature higher than 60 °C, the products were observed in just 1 h. For $(GAT)_6/(ATC)_6$, narrow bands could be observed from 37 to 70 °C, although the smear bands were obviously present at 50 and 60 °C. As compared with $(AGC)_6/(GCT)_6$, whose GC content is 2/3, elongation speed for $(GAT)_6/(ATC)_6$ (1/3 GC content) was up to 8.5 bp/min at 70 °C. Unexpectedly, elongation was carried out for both $(AGC)_6/(GCT)_6$ and $(GAT)_6/(ATC)_6$, although the efficiency is not so high. It should be noticed that the optimum reaction temperature of Vent DNA polymerase is 70–80 °C.

147.3.3 Sequence Analysis of Elongated Products by Restriction Enzyme Digestion and Sequencing of Elongated Products

To check whether the long products were surely elongated from short repeats, the elongated DNA expanded from $(AGC)_6$, $(GCT)_6$, and $(AGC)_6/(GCT)_6$ were digested by restriction enzyme ApekI with recognition site G↓CWGC and AluI with recognition site AGCT (Fig. 147.4). The expansion products of single and double strands were digested efficiently and almost no DNA was left with ApekI after 1 h,

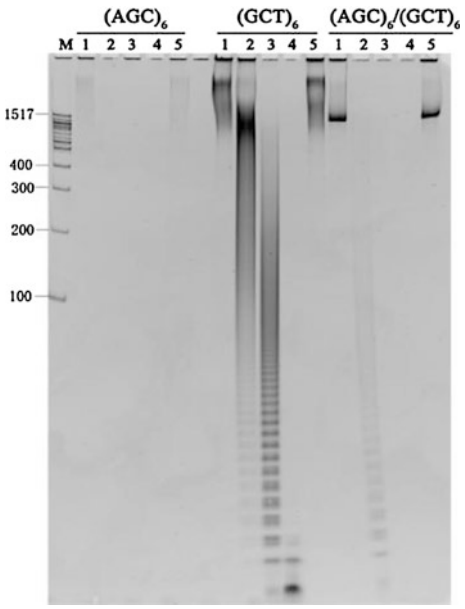

Fig. 147.4 Digestion of elongated products of short repeats by restriction enzyme. *Lane 1* Elongated products without. *Lane 2–4* ApeKI treatment for 10, 30, 60 min, respectively. *Lane 5* AluI treatment for 60 min. The standard reaction conditions were used and the DNA products were analyzed by nondenaturing polyacrylamide gel (6 %) electrophoresis

indicating that the products have a large amount of GCAGC and/or GCTGC sequence at short intervals. On the other hand, AluI could not digest these kinds of long DNA after 1 h showing that little AGCT sequence existed (Fig. 147.4).

The products sequencing of $(CAG)_6$ showed that the sequences of elongated products were basically tandem repeats of seeds (data not shown). There were up to 64 CTG in the long products without any interruption, indicating that slippage occured during elongation. It can be inferred from the result that the long DNA were made up of a large amount of trinucleotide repetitive sequences. Although several point mutations appeared. These may be the result of misincorporation by Vent (exo^-) DNA polymerase, which lacks proof-leading activity.

147.3.4 Proposed Mechanism of Elongation of Short Repetitive Sequences

From above results, we concluded that trinucleotide repetitive sequences had the potential to expand generally. Almost all of the trinucleotide repeats could be elongated, but the sequence had a great effect on it. It showed that no matter for single strands or for double strands, the repeats with more GC contents could easily proceed. For single strands, the expansion was so quick that the products of more than 10 kb could be got even in 1 h. We proposed a new model to explain this unusual elongation (Fig. 147.5a). The 18 nt short repeats can form transient hairpin at 3′ end, then DNA polymerase catch it to complete the template/primer

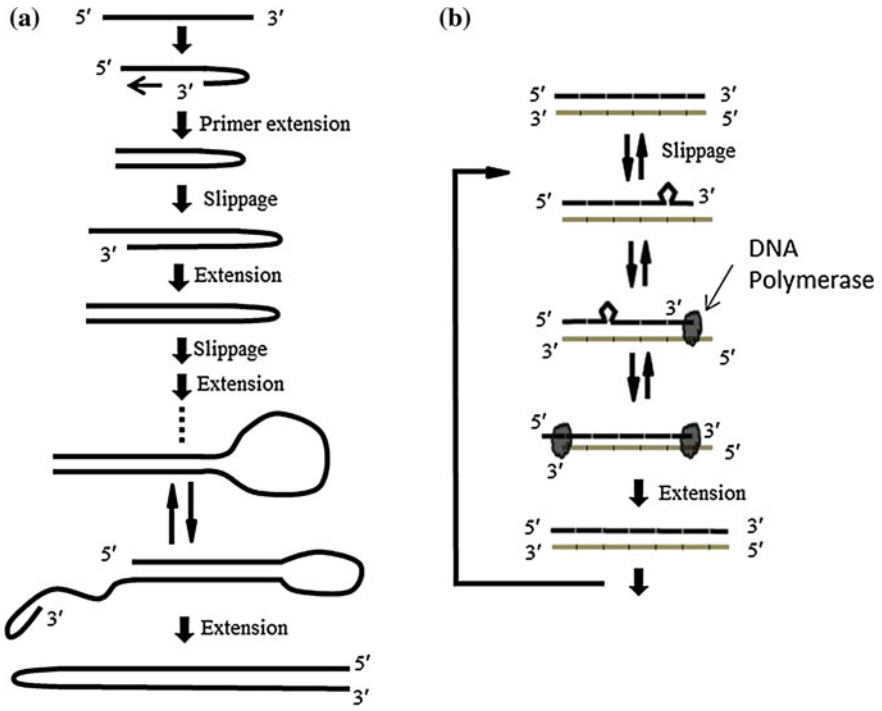

Fig. 147.5 Proposed mechanism for elongation of tandem repetitive sequence. **a** Mechanism proposed of the slippage and hairpin extension model for single strands repeats. **b** Possible mechanism for duplex repeats. The $(ATC)_6/(GAT)_6$ was set as an example to explain this slippage synthesis model

extension to form a bigger hairpin. Slippage happened in this duplex followed by filling the 5′ by DNA polymerase. After several slippage-extension cycles, the 3′ end slips to form a overhang, which was long enough to be in favor of a new hairpin. In this situation the self-priming extension happened again to prolong the repeats, then accessed to next expansion cycle.

According to this model, most of the results can be explained. The elongated products of $(AGC)_6$ and $(GCT)_6$ can be digested completely by ApeKI to give very clear bands, indicating that the elongated products consist of only repetitive sequence of short repeated seeds. At the beginning, for example, the hairpin of $(GAT)_6$ at the end is difficult to form at the reaction temperatures (60–70 °C). However, in the present of DNA polymerase, which can strongly bind to the template/primer to form the expansion complex, the transient hairpin formation can be caught by a DNA polymerase and carry out primer extension quickly. On the other hands, with the products becoming longer, the T_m can be up to 75.9 °C, in which the slippage can be easier to process and expansion become quick evidently. At lower temperatures, the duplex is difficult to slip; at higher temperatures, the transient hairpin formation is difficult to be caught by DNA polymerase so that the efficiency of primer extension becomes lower.

For duplex extension, the obvious bright single band showed that something different from single strands elongation had happened. Some other studies have discovered similar phenomenon and put forward several mechanisms [13, 17]. The slippage model is the orthodox view but few practical data and theoretical explanation could give supports. Moreover, this slippage required very high activation energy for inter-strands slippage when the reaction temperature is much lower than the T_m of repeats. We found that almost all of the double trinucleotide repetitive sequences increased linearly with time in some extent, and the extension speed is not related to the length of prolonged products, which is unacceptable to explain with the existing model. We proposed a model as shown in Fig. 147.5b. It was reported that several base pair opening (breathing) could happen transiently blew the T_m of duplex [18, 19]. Generally, base pairs are much easier to open at the end of the duplex [20]. Strand slippage inclines to happen for repetitive sequences. If the six base pair in the end of duplex opens to take a bulge form, the trinucleotide slippage will happen, then DNA polymerase binds to the 3′ end to prevent it from returning back to normal duplex. This kind of unstable trinucleotide bulge will slip to the other end to form a template-primer complex and filled by polymerase. The slippage-loop-filling cycle makes duplex extend gradually. The duplex expanded linearly as the even speed of loop forming, which could account for the result of our research. When product is elongated to more than 1.0 kb, the loop becomes easier to slip to left and right and need more time to wave to the end, which makes the expansion to slow down.

147.4 Conclusions

From above results, we concluded that all of the trinucleotide repetitive sequences, no matter single strand or duplex, can be used as the seeds for synthesizing long DNA by DNA polymerase under isothermal conditions, except several strands with uniform nucleotide. The temperature and sequence of repeats had great effect on it. The products were made up of a large number of trinucleotide repeats. New model on these two kinds of elongation was proposed. Single strands repeat expanded quickly to more than 10 kb. The high efficiency may be due to the continuous slippage and hairpin formation for self-priming and primer extension. For duplex repeats, as a result of slippage of duplex and spread of loop, the elongation product prolonged gradually with a narrow size distribution, and the length of synthesized DNA increases linearly with reaction time. These models may be helpful for clarifying the molecular evolution of nucleic acids and the nonspecification during DNA detection.

Acknowledgments Recruitment programs of “Wanren Plan”, “Fund for Distinguished Young Scholars” of Shandong province, and “National Youth Qianren Plan”, “Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT)” are acknowledged.

References

1. Hastings PJ, Lupski JR, Rosenberg SM et al (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10:551–564
2. Heidenfelder BL, Topal MD (2003) Effects of sequence on repeat expansion during DNA replication. Nucleic Acids Res 31:7159–7164
3. Treangen TJ, Salzberg SL (2012) Repetitive DNA and next generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46
4. Mohmood S, Sherwani A, Khan F et al (2003) DNA trinucleotide repeat expansion in neuropsychiatric patients. Med Sci Monitor 9:237–245
5. Kumar RS, Parthiban KT, Rao MG (2009) Molecular characterization of jatropha genetic resources through inter-simple sequence repeat (ISSR) markers. Mol Biol Rep 36:1951–1956
6. Heale SM, Petes TD (1995) The stabilization of repetitive tracts of DNA by variant repeats requires a functional DNA mismatch repair system. Cell 83:539–545
7. Aquilina G, Hess P, Branch P et al (1994) A mismatch recognition defect in colon carcinoma confers DNA microsatellite instability and a mutator phenotype. P Natl Acad Sci USA 91:8905–8909
8. Elizabeth FS, Linda J (2000) Dinucleotide repeat expansion catalyzed by bacteriophage T4 DNA polymerase in vitro. J Biol Chem 275:31528–31535
9. Ogata N, Miura T (1998) Creation of genetic information by DNA polymerase of the thermus thermophilus. Nucleic Acids Res 26:4657–4661
10. Ogata N, Miura T (2000) Elongation of tandem repetitive DNA by the DNA polymerase of the hyperthermophilic archaeon thermococcus litoralis at a hairpin-coil transitional state: a model of amplification of a primordial simple DNA sequence. Biochemistry 39:13993–14001
11. Liang XG, Jensen K, Frank-Kamenetskii MD (2004) Very efficient template/primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease. Biochemistry 43:13459–13466
12. Liang XG, Kato T, Hiroyuki A (2008) Mechanism of DNA elongation during de novo DNA synthesis. Nucleic Acids Res Supple 52:411–412
13. Kotlyar AB, Borovok N, Molotsky T et al (2005) In vitro synthesis of uniform poly (dG)-poly(dC) by klenow exo$^-$ fragment of polymerase I. Nucleic Acids Res 33:525–535
14. Tuntiwechapikul W, Salazar M (2002) Mechanism of in vitro expansion of long DNA repeats: effect of temperature, Repeat length, repeat sequence, and DNA polymerases. Biochemistry 41:854–860
15. Ogata N, Miura T (1998) Creation of genetic information by DNA polymerase of the archaeon thermococcus litoralis: influences of temperature and ionic strength. Nucleic Acids Res 26:4652–4656
16. Liang XG, Kato T, Hiroyuki A (2007) Unexpected efficient ab initio DNA synthesis at low temperature by using thermophilic DNA polymerase. Nucleic Acids Res supple 51:351–352
17. Jorg SH, Eric TK (2005) Efficient isothermal expansion of human telomeric and minisatellite repeats by thermococcus litoralis DNA polymerase. Nucleic Acids Res 33:4922–4927
18. Krueger A, Protozanova E, Frank-Kamenetskii MD (2006) Sequence-dependent base pair opening in DNA double helix. Biophys J 90:3091–3099
19. Frank-Kamenetskii M (1987) How the double helix breathes. Nature 328:17–18
20. Smolina IV, Demidov VV, Soldatenkov VA et al (2005) End invasion of peptide nucleic acids (PNAs) with mixed base composition into linear DNA duplexes. Nucleic Acids Res 33:e146