3 Biotech (2017) 7:128
DOI 10.1007/s13205-017-0745-2

SHORT REPORTS

High-fidelity PCR enzyme with DNA-binding domain facilitates de novo gene synthesis

A. S. Dolgova[1] · **O. A. Stukolova**[1,2]

Received: 19 May 2016 / Accepted: 30 January 2017 / Published online: 1 June 2017

Abstract Nowadays enzymatic synthesis of genes is the most powerful tool for fast resolution of the various tasks in the field of basic and applied biological research. PCR-based gene assembly from overlapping oligonucleotides has become a widely used strategy. However, all the methods described in the literature are not perfect and need an extra processing step. In this study we are verifying Phusion high-fidelity polymerase as a tool to reduce nucleotide mismatches in de novo gene synthesis, thus facilitating subsequent cloning. To test the efficiency of the polymerase, we selected Fel d 4 gene, which is a 581 bp DNA sequence encoding the lipocalin allergen protein, one of the major cat allergens. The approach described here, therefore, would be useful in DNA sequences creation.

Keywords De novo gene synthesis · High-fidelity polymerase · One-step PCR · Fel d 4

Introduction

The easiest way to obtain the DNA sequence for the future protein expression is to synthesize it with the help of PCR using template DNA extracted from the organism of interest. But template DNA is not always available or sometimes its extraction from unusual sources could be challenging. Because of that in vitro chemical synthesis of DNA has become one of the principal methods of modern synthetic biology. This method is especially preferred when the optimization of the codon composition is needed to achieve high-level expression in certain expression systems (Zhao et al. 2014; Itkonen et al. 2014; Öberg et al. 2011). Overlap extension (OE) PCR method (one-step PCR) was originally described by Stemmer et al. (1995). In this method a set of oligonucleotides had been designed to cover both DNA strands and the full-length molecule was obtained in a single PCR reaction. PCR method has become a widely used tool, but during the synthesis of long DNA sequences some difficulties may appear. Increasing length and complexity of DNA sequences results in increasing probability of nonspecific matches that may occur among oligonucleotides thus increasing number of errors in the resulting sequence. This process may lead to the decreasing yield or even to lack of the target PCR product.

Some modifications have been developed to improve one-step PCR. First, it was shown that purification of oligonucleotides by a high performance liquid chromatography (HPLC) (Andrus and Kuimelis 2001) or polyacrylamide gel electrophoresis (PAGE) (Ellington and Pollard 2001) noticeably reduces the number of mismatches. Also nowadays there are software for the oligonucleotide design: DNAWorks, Gene2Oligo, GeMS and others (Gould et al. 2014). But even if well-designed

Electronic supplementary material The online version of this article (doi:10.1007/s13205-017-0745-2) contains supplementary material, which is available to authorized users.

✉ A. S. Dolgova
annadolgova@inbox.ru

[1] Federal Budget Institution of Science "Central Research Institute of Epidemiology" of The Federal Service on Customers' Rights Protection and Human Well-being Surveillance, 3A, Novogireyevskaya st, Moscow 111123, Russia

[2] Federal Budget Institution of Science "Research Institute of Occupational Health", Prospect Budennogo 31, Moscow 105275, Russia

and additionally purified oligonucleotides are used it is not recommended to generate DNA molecules larger than 500 bp by one-step. For the synthesis of longer chains there are several strategies, mainly based on the formation of long sequences from short blocks by the second step of PCR (Xiong et al. 2004, 2006; Young and Dong 2004; Li et al. 2006). Despite all well known advantages of these methods, an additional editing of obtained DNA sequences is required using OE-PCR, T7 endonuclease I treatment or by other approaches (Ma et al. 2012).

Until recently during de novo gene synthesis using any of the methods mentioned above only two different DNA polymerases were used: Taq and Pfu. However, nowadays the new generation of high-fidelity enzymes for PCR amplification of DNA has become commercially available. It has been shown that Phusion High-Fidelity DNA Polymerase when used for DNA synthesis has dramatically lower error rates compared to the other enzymes previously used for such applications: its fidelity is 2–6× higher and approximately 50× higher than the fidelity of Pfu and Taq polymerases, correspondingly (Li et al. 2006; McInerney et al. 2014). Here we compare the features of Taq, Pfu and Phusion polymerases using an example of one-step assembly of an 581 bp-long DNA sequence corresponding to the cat lipocalin allergen protein (Fel d 4) for expression in *E. coli*.

Fel d 4 is one of the major cat allergens. The prevalence of cat sensitization ranges from 10 to 15% of the population in Northern Europe to 60% in Western countries (Gulbahar et al. 2003). Early and accurate diagnosis of cat allergy takes an important part in further patient's treatment. The most modern approach to allergy diagnostics is component-resolved and based on individual molecules. This new approach allows predicting clinical symptoms and their severity. In this connection Fel d 4 have been described as a useful tool for predicting asthma (Uriarte and Sastre 2016) and for study the differences and possible cross-reactivities in immune responses between Fel d 4 and other mammalian allergens of the lipocalin protein family (Smith et al. 2004). Thus, Fel d 4 is a required part of cat allergy distinguish. In recombinant form it would be available for immunological investigations, diagnosis and treatment.

Materials and methods

DNA and oligonucleotide design

Protein sequence of Fel d 4 was taken from GenBank: AAS77253.1. DNA sequence with codon optimization for expression in *E. coli* was designed using DNA works software (http://helixweb.nih.gov/dnaworks/). Oligonucleotides were designed using the same software and then edited manually using OligoAnalyzer software (https://eu.idtdna.com/calc/analyzer). The length of oligonucleotides did not exceed 62 bp, with an 18–24 bp overlap at the end of each fragment. All oligonucleotides were checked for GC content, the probability of primer dimer and loop formation. Resulting oligonucleotide sequences are listed in the Fig. 1. Designed oligonucleotides were synthesized and PAGE-purified (CRIE, Russia).

Gene synthesis

DNA was assembled from fragments by one-step PCR. A mix of primers was prepared in advance. The reaction was carried out in a 25 µL volume containing 200 mM of each dNTP, 1 µL of oligomix (1.5 pmol per primer), 30 pmol of outer primers, and either 2.5 µL of 10× Phusion GC buffer with 1 U of Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) or 10× Pfu buffer with 1 U of Pfu polymerase (Thermo Fisher Scientific) or 10 µL of Taq mix (CRIE, Russia). PCR thermal cycling was set as following: initial denaturation at 94 °C for 1 min, followed by 25 cycles of 94 °C for 30 s, 61 °C for 40 s and 72 °C for 50 s, followed by the final extension at 72 °C for 10 min. The products of Phusion and Pfu enzymes reactions were gel purified by Gel Extraction Centrifugal Filter Units (Merck Millipore, Darmstadt, Germany), followed by the dA tailing (Taq DNA polymerase, Evrogen, Russia) and cloning into pGEM®-T Easy vector (Promega, Madison, USA). Five positive clones resulting from the ligation of the product obtained by each polymerase were selected and sequenced by Applied Biosystems 3500 Series Genetic Analyzers (Thermo Fisher Scientific). Final nucleotide sequence of Fel d 4 gene was submitted to GenBank with the Accession Number KT334370.

Cloning and expression

The gene was cut out from pGEM®-T with EcoRI/XhoI, ligated into the expression vector pET24a and transformed into *E.coli* BL21 (DE3), followed by cultivation and subsequent protein isolation and purification using a Chelating Sepharose Fast Flow (GE Healthcare, USA) and DEAE Sepharose Fast Flow (GE Healthcare, USA) according to the manufacturer's instructions. Recombinant protein rFel d 4 was obtained in buffer: 50 mM Tris/HCl, pH 8.0, 0.3 M NaCl, 10% glycerol at a final concentration 1.4 mg/ml. The purity and molecular size were verified by SDS-PAGE analysis (Fig. 3). The reactivity to various sera was ascertained by microarray method (Sup. 1) using sera from cat-allergic patients and normal controls (Fig. 3).

A

Feld4_A	ACT GAA TTC GAT GAA ACT TCT GCT GCT GTG C
Feld4_1	ATG AAA CTT CTG CTG CTG TGC CTG GGC CTG ATT CTG GTG TGC GCG CAT GAA GAA GAG
Feld4_2	GCC ACT GAT CTT GCT GAT ATC GAT GTT GCT ACG CAC CAC GTT CTC TTC TTC ATG CGC GCA
Feld4_3	TCG ATA TCA GCA AGA TCA GTG GCG AAT GGT ATA GCA TTC TGT TAG CGA GCG ATG TGA AAG A
Feld4_4	GTT CCA CAA ACA CAC GCA TAC TGC CGT TTT CTT CAA TCT TCT CTT TCA CAT CGC TCG CTA
Feld4_5	AGT ATG CGT GTG TTT GTG GAA CAT ATT AAA GCG CTG GAT AAC AGC AGC CTG AGC TTT GTG
Feld4_6	GAT TTC GGT GCA TTT GCC GTT TTC TTT GGT ATG AAA CAC AAA GCT CAG GCT GCT GTT AT
Feld4_7	AAC GGC AAA TGC ACC GAA ATC TTC CTG GTG GCG GAT AAA ACC AAA GAT GGC GTG TAT ACC
Feld4_8	TTC CAC AAT GCT AAA CAC GTT ATA GCC ATC ATA CAC CAC GGT ATA CAC GCC ATC TTT GGT
Feld4_9	GCT ATA ACG TGT TTA GCA TTG TGG AAA CCG TGT ATG ATG AAT ATA TTC TGC TGC ATC TGC T
Feld4_10	ACC AGC TGA AAT GGA CGT GTC TTA TCA AAG TTC AGC AGA TGC AGC AGA ATA TAT TCA TC
Feld4_11	GAC ACG TCC ATT TCA GCT GGT GGA GTT CTA TGC GCG TGA ACC GGA TGT GAG CCA GAA ACT GA
Feld4_12	ACA ATA CCA TGT TCC TGG CAA TAC TTC ACA AAC TTC TCT TTC AGT TTC TGG CTC ACA TCC G
Feld4_13	GTA TTG CCA GGA ACA TGG TAT TGT GAA CAT TCT GGA TCT GAC CGA AGT GGA TCG TTG CT
Feld4_14	ACA GAG CTA TCC TGC GCA ACT TCA GAG CCA CGA GCC TGT AAG CAA CGA TCC ACT TCG GTC
Feld4_B	ACG TCT CGA GTT ATT CAA CAG AGC TAT CCT GCG CAA C

B

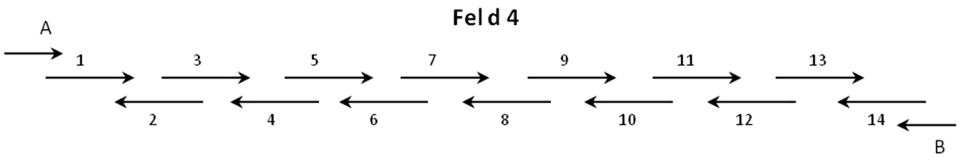

Fig. 1 **a** Oligonucleotides designed for PCR-based assembly of the cat lipocalin allergen Fel d 4. **b** A schematic diagram of the one-step PCR based synthesis of the Fel d 4 from 14 oligonucleotides (1.5 pmol of inner primers and 30 pmol of external primers were taken for reaction)

Quantification of anti-Fel d 4 immunoglobulin E

Capacity of obtained rFel d 4 protein for specific recognition of IgE in human serum samples was determined using special microarray method developed in our laboratory (Sup. 1). Sensitization to Fel d 4 was considered significant when specific IgE level was $\geq$0.35 ME/mL.

Serum samples

Serum samples were obtained from 200 patients: 174 with a positive specific IgE test response to a cat dander extract (ImmunoCAP; Phadia, Uppsala, Sweden) and 26 with a negative response. Serum samples were collected in the Center of Molecular Diagnostic of the Federal Budget Institution of Science "Central Research Institute of Epidemiology" of The Federal Service on Customers' Rights Protection and Human Well-being Surveillance in Moscow, Russia.

Results

There were tested three polymerases: Taq, Pfu and Phusion. DNA fragment corresponding to a full-sized Fel d 4 gene (581 bp) was synthesized by one-step PCR. Amplification was performed with some modifications of the original method. In contrast to several PCR-assembly methods successfully using Taq polymerase (Andrus and Kuimelis 2001; Yehezkel et al. 2012) we have not obtained any PCR product of the required size using this polymerase (Fig. 2). Pfu polymerase gave clearly visible band about 581 bp (as was anticipated) but there was a number of smeary diffuse bands above and one extra 300 bp band. Phusion polymerase produced one pure band, corresponding to the PCR-product of the correct size.

Pfu and Phusion amplification products consisting of $\sim$580 bp were cloned into pGEM vector. After PCR screening five positive clones for each polymerase were sequenced. In the case of Pfu polymerase there were no exact clones of Fel d 4 gene: one clone had one, two clones had two and two clones had three mismatches. In the case of Phusion there were two correct sequences and three sequences with one nucleotide mismatch (Sup. 2).

Plasmid carrying codon optimized gene was transformed and expressed in *E. coli* in order to verify sequence efficiency of the 21.4 kDa lipocalin Fel d 4 protein. Recombinant Fel d 4 molecular weight deduced comparing standards matched values calculated from the nucleotide sequence (Fig. 3). The ability to bind sIgE was verified by microarray method with cat-positive serum pool (Fig. 3). Anti-Fel d 4 IgE in the serum samples characterized by

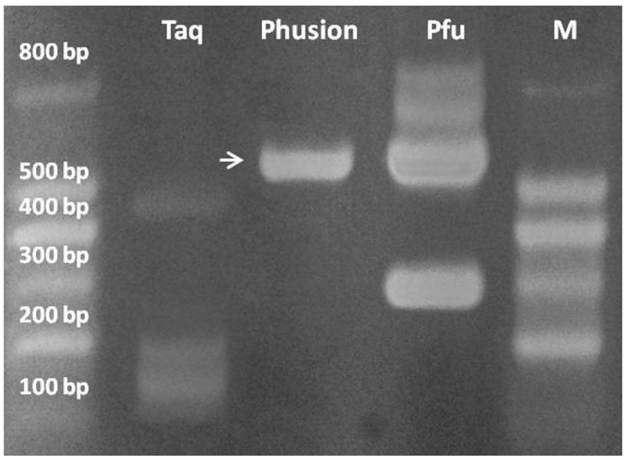

Fig. 2 Analysis of OE-PCR performance of three polymerases Taq, Phusion and Pfu using gel electrophoresis. Full-size 581-bp Fel d 4 gene is indicated by an *arrow*

specific IgE response to a cat dander extract was quantitated by the microarray assay. Thirty-nigh percent of the samples from cat-allergic subjects group (68/174) had Fel d 4-specific IgE (Sup. 3).

Discussion

A variety of approaches for DNA synthesis have been developed (Xiong et al. 2004, 2006; Yang et al. 2012) but the resulting DNA sequences obtained using any of these techniques need to be edited during the following steps. Perhaps the problem-solving approach is to use the new generation polymerases which have higher fidelity. We tested PCR performance of three different DNA polymerases (Taq, Pfu and Phusion) in one-step PCR reaction. Mean error rates for the amplification of 581 bp DNA fragment for Pfu and Phusion were 2.2 ± 0.837 and

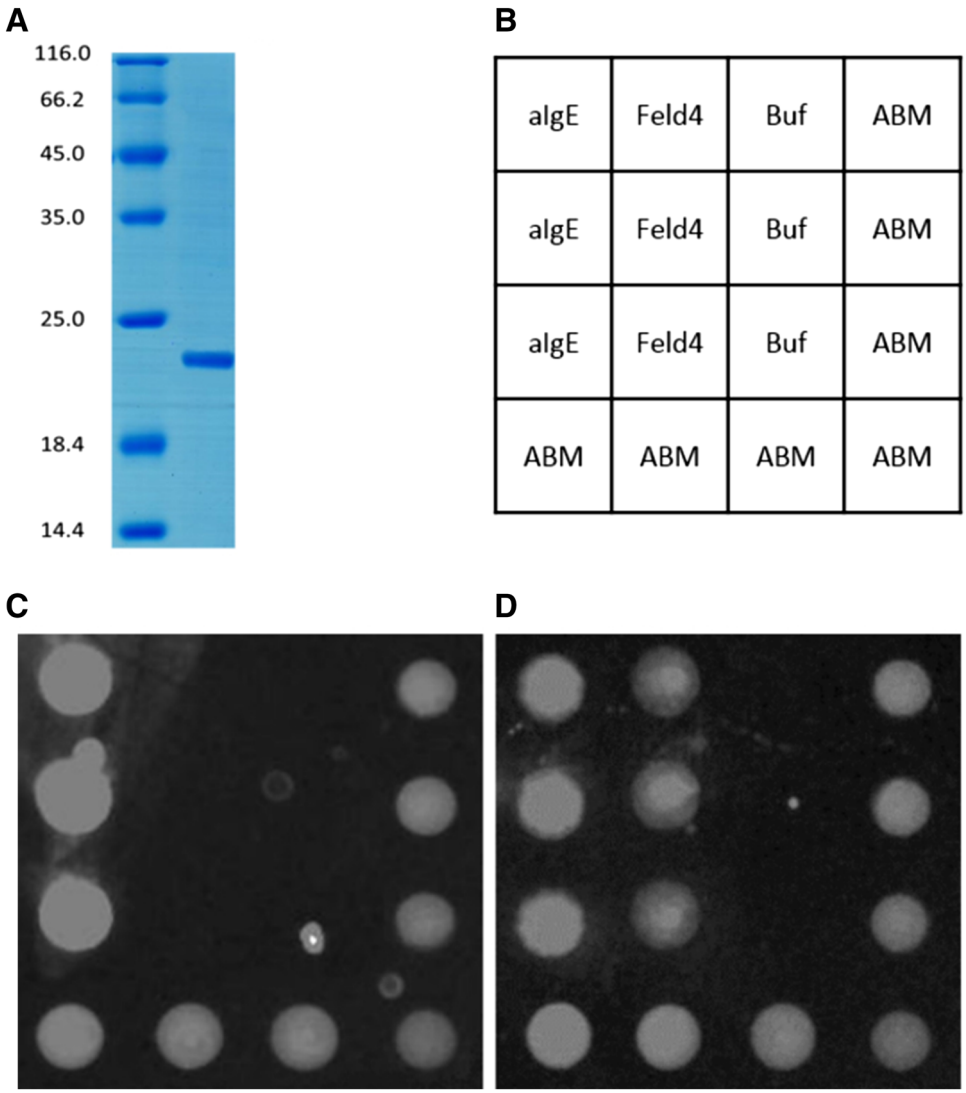

Fig. 3 Analyses of recombinant protein Fel d 4 properties and capacity to bind specific IgE. **a** Sodium dodecyl sulphate polyacrylamide gel electrophoresis of recombinant Fel d 4 (purity >95%). Molecular weight markers (kd) are on the *left* and recombinant Fel d 4 on the right side. **b**–**d** Microarray data confirming the reactivity of IgE binding. **b** Scheme of microarray. *ABM* array border markers; *Buf* 1× PBS buffer, negative control; *Feld4* recombinant protein Fel d 4 plotted in concentration 200 µg/mL, *aIgE* anti-human IgE, positive control. **c** Microarray showing the reactivity of cat negative serum pool. **d** Microarray showing the reactivity of cat-positive serum pool

0.6 ± 0.548 respectively ($p < 0.05$, Student's t test). So, we can conclude that the Phusion polymerase is three times more efficient than then Pfu polymerase in such type of PCR reaction. Such a significant difference may be explained by the fact that only Pusion polymerase has DNA-binding domain fused to a Pyrococcus-like proof-reading polymerase. Usage of Phusion polymerase reduces the number of clones required to find the correct sequence of target gene and will eliminate the step of nucleotide editing.

Furthermore, Pfu polymerase produced an additional PCR band of a lower molecular weight. This disadvantage complicates subsequent cloning, because in this case gel purification of PCR-product is required. Moreover, if further assembly of a long sequence from short blocks is planned (as described by Xiong et al. 2006) PCR-products having incorrect molecular weight would append one more step in the procedure. The use of Phusion polymerase will allow taking fragments from the first step of PCR to the second without purification. This would be time-saving and economically advantageous as well.

Purified recombinant protein rFel d4 was tested to IgE binding capability. Fel d 4-specific IgEs were found in 39% of serum samples from cat-allergic patients. This results are in a well agreement with 34% of Fel d4 positive samples obtained in 2014 years study (Konradsen et al. 2014) and differ from another study in which the percentage of patients sensitized to Fel d 4 reached 63% (Smith et al. 2004). Differences in IgE binding could be caused by various factors. The results obtained in Konradsen's study are closer to the results of current study because of more similar conditions of tests carrying out. During Konradsen's study IgE sensitization measurement was performed using ImmunoCAPTM ISAC system (based on microarray technology) and in Smith's DELFIA® immunoassay, the modification of traditional ELISA, was used. Apart from that Konradsen's study took place on the basis of Sweden population and Smith's on Australian. It is very likely that IgEs profiles can vary in populations of different climate. Anyway our results are comparable (more or less) with previous investigations allowing us to assume that Fel d 4 which DNA sequence was synthesized by one-step PCR can be used for molecular based allergy diagnostics and will provide the means for allergen-specific immunotherapy in the future.

Conclusion

It was shown that mismatches in OE-PCR gene assembly occurred three times less often when Phusion polymerase was used compared to the usage of Pfu polymerase. Moreover, Pfu also produced unwanted PCR-products with incorrect molecular weight that could lead to subsequent meshing of the fragment cloning, while Phusion did not have such a disadvantage. Thus, usage of Phusion polymerase is able to facilitate gene synthesis and reduce the time of DNA assembly. Phusion polymerase had been used in synthesis of the rFel d 4 gene that was successfully expressed and its reaction with IgE in the serum of 39% cat-allergic patients was observed suggesting the obtaining of the fully-functional protein. Usage of Phusion or other hi-fidelity polymerases would certainly be a great approach in the future DNA synthesis for biomedical or biotechnology purposes.

References

Andrus A, Kuimelis RG (2001) Analysis and purification of synthetic nucleic acids using HPLC. Curr Protoc Nucleic Acid Chem 1:10.5.1–10.5.13. doi:10.1002/0471142700.nc1005s01

Ellington A, Pollard JD Jr (2001) Introduction to the synthesis and purification of oligonucleotides. Curr Protoc Nucleic Acid Chem Appendix 3, Appendix 3C. doi: 10.1002/0471142700.nca03cs00

Gould N, Hendy O, Papamichail D (2014) Computational tools and algorithms for designing customized synthetic genes. Front Bioeng Biotechnol 2:41

Gulbahar O, Sin A, Mete N, Kokuludag A, Kirmaz C, Sebik F (2003) Sensitization to cat allergens in non-cat owner patients with repisratory allergy. Ann Allergy Asthma Immunol 90:635–639

Itkonen JM, Urtti A, Bird LE, Sarkhel S (2014) Codon optimization and factorial screening for enhanced soluble expression of human ciliary neurotrophic factor in *Escherichia coli*. BMC Biotechnol 14:92

Konradsen JR, Nordlund B, Onell A, Borres MP, Grönlund H, Hedlin G (2014) Severe childhood asthma and allergy to furry animals: refined assessment using molecular-based allergy diagnostics. Pediatr Allergy Immunol 25:187–192

Li M, Diehl F, Dressman D, Vogelstein B, Kinzler KW (2006) BEAMing up for detection and quantification of rare sequence variants. Nat Methods 3:95–97

Ma S, Saaem I, Tian J (2012) Error correction in gene synthesis technology. Trends Biotechnol 30:147–154

McInerney P, Adams P, Hadi MZ (2014) Error rate comparison during polymerase chain reaction by DNA polymerase. Mol Biol Int 2014:287430. doi:10.1155/2014/287430

Öberg F, Sjöhamn J, Conner MT, Bill RM, Hedfalk K (2011) Improving recombinant eukaryotic membrane protein yields in *Pichia pastoris*: the importance of codon optimization and clone selection. Mol Membr Biol 28:398–411

Smith W, Butler AJL, Hazell LA, Chapman MD, Pomes A, Nickels DG, Thomas WR (2004) Fel d 4, a cat lipocalin allergen. Clin Exp Allergy 34:1732–1738

Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995) Single step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164:49–53

Uriarte S, Sastre J (2016) Clinical relevance of molecular diagnosis in pet allergy. Allergy 71:1066–1068

Xiong AS, Yao QH, Peng RH, Li X, Fan HQ, Cheng ZM, Li Y (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucl Acids Res 32:e98

Xiong AS, Yao QH, Peng RH, Duan H, Li X, Fan HQ, Cheng ZM, Li Y (2006) PCR-based accurate synthesis of long DNA sequences. Nat Protoc 1:791–797

Yang JK, Chen FY, Yan XX, Miao LH, Dai JH (2012) A simple and accurate two-step long DNA sequences synthesis strategy to improve heterologous gene expression in Pichia. PLoS One 7:e36607

Yehezkel TB, Linshiz G, Shapiro E (2012) De novo DNA synthesis using single-molecule PCR. Methods Mol Biol 852:35–47

Young L, Dong Q (2004) Two-step total gene synthesis method. Nucl Acids Res 32:e59

Zhao H, Blazanovic K, Choi Y, Bailey-Kellogg C, Griswold KE (2014) Gene and protein sequence optimization for high-level production of fully active and aglycosylated lysostaphin in *Pichia pastoris*. Appl Environ Microbiol 80:2746–2753