P-ISSN 2586-9000
E-ISSN 2586-9027

Homepage : *https://tci-thaijo.org/index.php/SciTechAsia* Science & Technology Asia

Vol.27 No.1 January - March 2022 Page: [190-198]

Original research article

Thermodynamic Predictive Design for Single-Step Synthesis of Human Immunodeficiency Virus-1 Gene

Somchai Saengamnatdej*

Department of Microbiology and Parasitology, Faculty of Medical Science, Naresuan University, Phitsanulok 65000, Thailand

Received 3 September 2020; Received in revised form 3 September 2020
Accepted 2 February 2021; Available online 28 March 2022

ABSTRACT

Synthetic genes have many advantages over natural ones. Gene synthesis strategies that are cost-effective, simple, time-saving and accurate are often developed. Many previous works have suggested the use of numerous short oligonucleotides to improve accuracy. In this study, the prediction of secondary structure, shape and interaction of nucleic acids was studied in a gene synthesis with longer oligonucleotides. The oligonucleotides were designed for the *Escherichia coli* codon-optimized 318-bp gene coding for the HIV-1 protease and the extra seven amino acids located upstream and function as an auto-processing site. All the designed oligonucleotides were then analyzed for their interactions with NUPACK to predict synthesis, and with gquad and DNAShapeR to predict G-quadruplexes and shape, respectively. To test the prediction, the optimal number of thermal cycling rounds in the assembly step was determined. It was found that this synthesis could be done in a single step without errors as verified by sequencing. This report shows that analysis of minimum free energy secondary structures of the interacting oligonucleotides is useful for checking whether the components can hybridize correctly.

Keywords: DNAShapeR; Gene synthesis; Gquad; Minimum free energy secondary structure; NUPACK

1. Introduction

In this work, the synthesis of a gene encoding HIV-1 protease and its additional seven amino acids upstream, using long oligonucleotides, was studied. The G-quadruplex, other structures, and general DNA shape were examined. The predictive minimum free energy (MFE) structures and their interactions were calculated to help in designing a single-step synthesis. With this careful design for specificity of the oligonucleotides, the synthesis of this gene could be

*Corresponding author: somchais@nu.ac.th

doi: 10.14456/scitechasia.2022.19

achieved in one step. Emerging and re-emerging pathogens have become more common in past years. Many of them have neither specific treatments, nor prophylactic measures. Synthesis of genes, rather than cloning natural sequences, would allow molecular study of these virulent pathogens in a laboratory setting without the required high biosafety-level containment measures. In addition, propagation and genome extraction of the microorganisms would no longer be needed. Indeed, research on functional synthetic viral genes and genomes has increased, some of which has been used to redesign viruses to produce a new class of live vaccines, to study gene function and pathogenic potential and, quite controversially, to weaponize for military purposes [1].

Various strategies, protocols, and procedures for the synthesis of genes have been published [2-5]. The less complicated, more cost-effective and less labor-intensive methods appear to be the ones which use DNA polymerase. The pioneering works were performed in 1984 using in vivo DNA polymerase in the bacterial repair system by Rink, H. and colleagues to synthesize eglin c [6], and in 1988 by Adam, S.E. and colleagues, to synthesize tat gene from oligonucleotides of greater than 100 bp in length, which were connected to each other with linkers and to a vector with end adapters [7]. A much more simplified approach, however, was demonstrated using recursive PCR in 1992 by Prodromou and Pearl [8]. They used ten 54- to 86-nt long oligonucleotides with 17- to 20-nt overlaps to synthesize the 522-bp human lysozyme gene with Vent DNA polymerase, and had a satisfactory result with only one clone with a G to A transition out of 12 sequenced clones. Then, in 1995 Stemmer et al. [9] assembled fifty-six 40-nt long oligonucleotides with 20-nt overlaps in a 55-cycle PCR to synthesize a 1.1-kb TEM-1 *bla* gene, and also used 134 oligonucleotides with the same length and overlap as above, except for two, in a three-step (40-, 25-, 20- cycle) PCR to synthesize a 2.7-kb plasmid. In 2004, Young, L. and Dong, Q. synthesized a 470-bp DNA sequence containing a pro-insulin gene from twelve 50-nt long oligonucleotides with 10-nt overlaps on both ends, and another three 1.1-1.2 kb genes from up to ninety-four oligo-nucleotides with the same size and the overlapping length [10]. These two steps were called dual asymmetrical (DA) PCR and overlap extension (OE) PCR giving them only one correct gene out of four clones for the 470-bp gene and none for the other three longer genes. Moreover, the average deletion rate was estimated to be about one deletion per 200 nt. However, they showed that using ninety-four 25-nt long oligonucleotides without gaps between the oligonucleotides, which covered the whole length of both the sense and anti-sense strands with the use of T7 endo-nuclease-I, gave the correct products. For the synthesis of larger genes, oligonucleotides were assembled into fragments and the fragments, which were designed to have overlap, were then assembled into a full-length DNA sequence [11-12].

2. Materials and Methods

2.1 Gene design and synthesis strategy

The HIV-1 protease-coding DNA sequence (Accession Number AJ307332) was obtained from GenBank. The sequence was modified in some positions as required for our ongoing study. Furthermore, the DNA sequence for the viral auto-processing site, containing seven amino acids: (Gly-Thr-Val-Ser-Phe-Asn-Phe), was added upstream to the HIV-1 protease DNA sequence. The synthesis strategy was to combine the step of assembly of overlapping oligonucleotides into a gene using polymerase chain assembly (PCA) with the step of amplification. The program, GeneGenie [13], was used to design all oligonucleotides containing the overlaps with similar melting temperatures (Tm = 62 ˚C) and with the *Escherichia coli* codon usage. The outcome was checked and modified before use. The DNA sequences of the amplifying oligonucleotides (herein called adapters) containing a restriction site were designed manually.

2.2 DNA shape and G-quadruplex analyses

All of the designed oligonucleotides from 2.1 (Table 1) were analyzed for likeliness of forming non-B DNA structures with the R package, gquad [14]. Only four long oligonucleotides were also examined for general shape parameters with the package DNAShapeR [15].

Table 1. Designed oligonucleotides*.

Name	Size	Sequence (5'-3')
Adapter1	21	AAAGGATCCGGAACTGTATCC
Oligonucleotide1	94	GGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGAAACGTCCGCTGGTTACCATTAAAATTGGTGGTCAATTGAAAGAAGCACTGCTTG
Oligonucleotide2	95	CACCAATCATTTTCGGTTTCCAACGACCCGGCAGATTCATTTCTTCAATAACGGTATCATCTGCACCGGTATCAAGCAGTGCTTCTTTCAATTGA
Oligonucleotide3	94	TTGGAAACCGAAAATGATTGGTGGTATTGGTGGTTTTATTAAAGTTCGTCAGTATGATCAGATTATTGTTGAAATTTGTGGTCATAAAGCCATT
Oligonucleotide4	99	AAAATTTAAAGTGCAGCCAATCTGGGTCAACAGATTACGACCAATAATATTAACCGGGGTCGGACCAACCAGAACGGTACCAATGGCTTTATGACCACA
Adapter2	23	AAACTCGAGAAAATTTAAAGTGC

Note: *in the order as used in the NUPACK program.

2.3 Prediction of MFE secondary structures and the interaction

To predict the hybridization at the range of temperatures in the synthesis, all six oligonucleotides were used as an input in NUPACK version 3.0.5 [16]. The function MFE and the following parameters were used, material: DNA, sodium: 0.250, magnesium: 0.035. The output structures were visualized with VARNA version 3.93 [17] and the structures and interactions from different temperatures, ranging from 95 to 50 ˚C, at an interval of 5 ˚C, were compared and interpreted.

2.4 Oligonucleotides and adapters

All oligonucleotides were purchased as synthesized at 2 nmoles and purified with PAGE-200 polyacrylamide gel (Biobasic, Inc., Canada). The two adapters, one as a forward amplifying oligonucleotide and another as a reverse amplifying oligonucleotide, were also ordered from the above company, but at 25 nmoles.

2.5 Determination of optimal thermal cycle number

Determination of the optimal number of thermal cycles was accomplished by varying the number of cycles in the reaction of assembly step. The PCR master-mix was prepared using the 10xPCR buffer from the i-TaqTM plus DNA polymerase (Cat. No. 25152, Intron Biotechnology, Inc.) and 1 nM each of all four oligonucleotides, 10 mM dNTPs, and Taq-DNA polymerase. The master-mix was allocated into 8 tubes for reactions of 0, 1, 2, 3, 4, 5, 10, and 20 thermal cycle(s). For the synthesis without a prior assembly step (0 cycles), the tube was kept on ice. The thermal cycling conditions were 95 ˚C 2', N x (94 ˚C 30s, 20%Ramp, 58 ˚C 30s, 72 ˚C 1'), 72 ˚C 2', (N = positive integer) and performed with the Applied Biosystems Veriti® 96-well thermocycler. After the end of each cycling, all reactions were kept on ice. Then, 1 μl of the product from each reaction was used as a template for the amplification step. The two adapters, each at the final concentration of 0.8 μM, were used as amplifying oligonucleotides. The conditions for the amplification step were 95 ˚C 2', 35 x (94 ˚C, 30s, 52 ˚C 30s, 72 ˚C 1'), 72 ˚C 5', 4 ˚C ∞. The amplified products then were visualized using agarose gel electrophoresis.

2.6 Synthesis reaction, agarose gel electrophoresis and DNA Sequencing

Equimolar solutions (1 nM) of each oligonucleotide were prepared. The synthesis reaction was prepared in the same way as stated in section 2.5, except the two adapters were added into the reaction in order to combine the assembly and amplifying steps into one single step. The conditions were 95 ˚C 2', 35 x (94 ˚C, 30s, 52 ˚C 30s, 72 ˚C 1'), 72 ˚C 5', 4 ˚C ∞. The synthesized gene was detected with 0.8% agarose gel electrophoresis, stained with ethidium bromide, then visualized and photographed with gel doc systems and Genesnap software (Syngene). The electrophoresis buffer was 1xTAE. The 1-kb standard DNA ladder M11 was from SibEnzymes Ltd., Russia. Then, the synthesized products were purified from the agarose gel using QIAEXII (QIAGEN) and used for sequence verification. The nucleotide sequencing was performed in-house with the reverse adapter as the primer.

3. Results and Discussion

The Bioinformatics tool, GeneGenie, gave five oligonucleotide sequences with the shortest one being at the end of the target DNA sequence. The codons and sequences of the designed oligonucleotides were examined. The shortest one was then excluded in this study and the sequence of the fourth one was then extended to cover the rest of the gene sequence. The sequences of the two adapters were designed manually to include the desired restriction site at the 5'- or 3'-end for a downstream recombination process in another study. The sequences of all oligonucleotides and adapters used in this study are shown in Table 1.

The result of the prediction with the gquad() function in gquad package is shown in Table 2. A-phased DNA repeats, triplexes, slipped motifs, tandem repeats, and Z-DNA motifs are not found in any of the six sequences. However, weak G-quadruplex-forming sequences were predicted in oligonucleotides-1, 3, and 4.

Table 2. G-quadruplexes predicted by gquad.

Input_ID	Sequence_ position	Sequence	Sequence_ length	Likeliness
1	38	ggaaacgtccgctggttaccattaaaattggtgg	34	*
2	-	-	-	-
3	3	ggaaaccgaaaatgattggtggtattggtgg	31	*
4	24	gggtcaacagattacgaccaataatattaac cggggtcggaccaaccagaacgg	56	*
5	-	-	-	-
6	-	-	-	-

Note: * weak G-quadruplex-forming

In general, quadruplexes can be formed from one, two, or four separate DNA strands [18] and G-quadruplex-forming sequence can lead to the blocking of DNA synthesis [19].

Fig.1 shows the plots of the mean minor groove width generated with the getShape() and plotShape() functions in DNAShapeR package. Parameters determining the DNA shape were predicted directly from the sequence of the four long oligonucleotides. The widths of the minor groove are less than 5 angstroms in some regions of oligonucleotides-1 to -4 and greater than 5 angstroms in almost the entire length of oligonucleotide-5. The narrow groove width has lower electrostatic potential due to the backbone phosphates being close to the center of the groove[20]. Whether this electrostatic potential is involved in the process of gene synthesis is unknown.

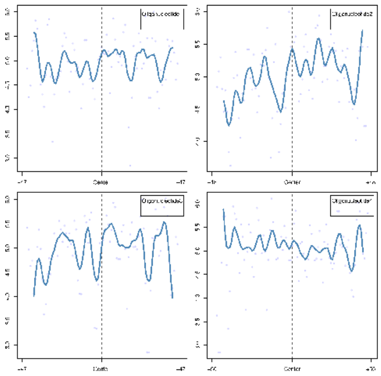


Fig. 1. The mean minor groove widths (in angstroms) of four long oligonucleotides predicted by DNAShapeR.

To better understand the nature and behavior of the nucleic acids in the synthesis reaction, and to examine whether the hybridization would occur correctly, the oligonucleotide components were analyzed thermodynamically. The sequences, structures, and their interactions from the executable MFE in NUPACK at the different temperatures were visualized, annotated, and titled using VARNA and are shown in Fig 2.

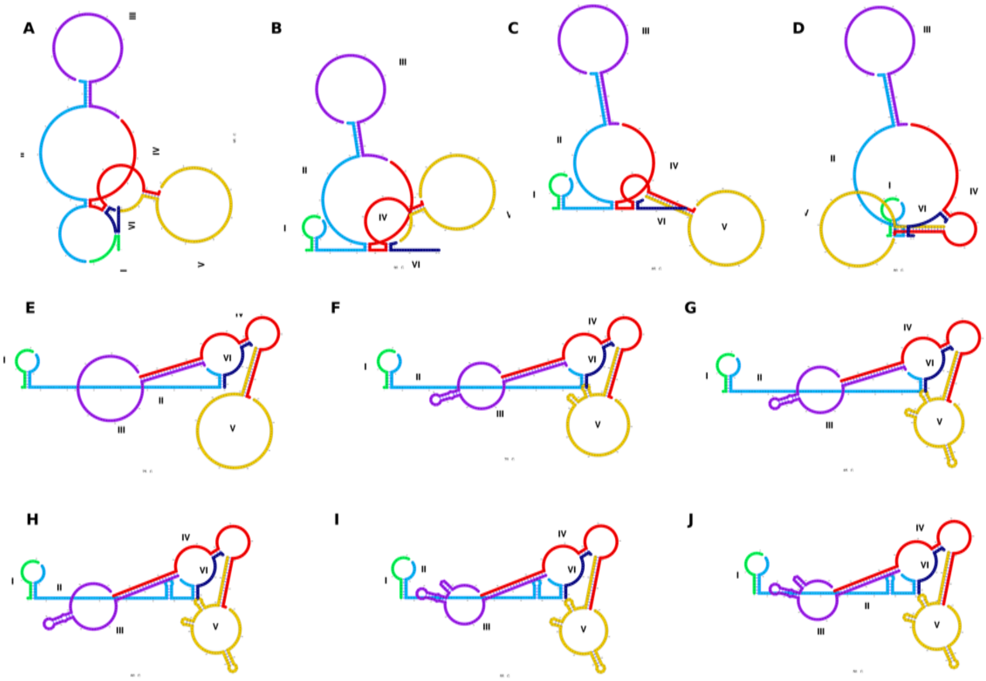


Fig. 2. The MFE secondary structures and interaction. All oligonucleotides fold into various secondary structures and hybridize with other oligonucleotides to different extents. The structures and interactions are shown for the temperatures 95 ˚C (A), 90 ˚C (B), 85 ˚C (C), 80 ˚C (D), 75 ˚C(E), 70 ˚C(F), 65 ˚C(G), 60 ˚C(H), 55 ˚C(I), and 50 ˚C(J). Each oligonucleotide is labeled in Roman numerals I to VI which are Adapter1 (in green), Oligonucleotide1 (in blue), Oligonucleotide2 (in violet), Oligonucleotide3 (in red), Oligonucleotide4 (in yellow), and Adapter2 (in navy), respectively.

All the assembly reactions with different numbers of thermal cycles and without thermal cycling gave the same size amplified DNA product on agarose gel, as shown in Fig. 3A. The one-step synthesized gene also appeared as a single band on a 0.8% agarose gel as shown in Fig. 3B. When compared with the standard DNA marker, their sizes measured approximately 300 bp. The DNA sequencing showed that the product was the desired gene and there were no mutations in the synthetic gene (Fig. 4).

The comprehensive aspects on nucleic acid structural energetics are available [21]. There are many prediction tools for nucleic acid structures and some tools work better than the others [22]; however, only MFOLD, UNAFOLD, and NUPACK have a specific algorithm to be used with DNA.

NUPACK was used in this study because it not only incorporates Santa Lucia's work (reference therein) in its algorithm, but it also allows the analysis of many oligonucleotides (more than two) at once.

The MFE secondary structures and interactions were predicted for the range of temperatures used in the synthesis reaction (95 ˚C to 50 ˚C). The concentrations of sodium and magnesium were equivalent to those in the reaction. The predicted free energy decreased proportionally with temperature from -25.073 kcal/mol at 95 ˚C to -70.970 kcal/mole at 50 ˚C (data not shown). The oligonucleotides barely hybridized at 95 ˚C (Fig. 2A) and DNA polymerase would not function at this temperature. At 90-80 ˚C, the oligonucleotide-1 hybridizes with oligonucleotide-2 (Fig. 2B, C, D). This allows T*aq* DNA polymerase, which works best at high temperatures, to fill in nucleotides in the 5' to 3' direction. The newly formed DNA duplex eliminates the secondary structure and frees the adapter1, which hybridizes weakly. When the temperature goes down to 75 ˚C or lower (Fig. 2 E-J), the oligonucleotide-3 hybridized to oligonucleotide-2 at the 5'-end, and to oligonucleotide-4 at the 3'-end. This showed the correct hybridization had occurred. The DNA polymerase then synthesizes the nucleotides on the hybrid sequence. Again, the duplex DNA strand will free the adapter2. Therefore, based on this minimum free energy prediction, the assembly of the first copy of the gene could occur before the first round of thermal cycling. This explanation, therefore, hypothesizes that a separate step for PCA is not necessary as previously thought, because thermo-stable DNA polymerase will begin to synthesize immediately while the oligonucleotide components move into the right geometry adopting the structure with minimum free energy.

The above prediction was concordant with the result of the optimal thermal cycle number determination experiment. As shown in Fig. 3A, the product could be amplified even in a reaction without a PCA step (thermal cycle = 0). This suggested that the PCA step could be omitted. When the adapters were added in the assembly reaction, and the two-step synthesis was reduced to one-step, the results appeared no different from those of other reactions (Fig. 3B). After sequencing the amplified product and examining its sequence with Artemis (Sanger Institute), it was found that the gene was synthesized successfully and without error. The sequence was aligned with all four oligonucleotides as shown in Fig. 4. Although Taq DNA polymerase has as high of an error rate as $2x10^{-4}$ [23], for sequences less than 1 kb, the probability of error is sufficiently low for synthesis. This, therefore, simplifies two steps into one. The strategy for synthesis in this work is different from that used in recursive PCR [8], by using adapters in place of the outermost oligonucleotides, and is different from that used in the two-step PCR method [2] as well, by integrating the two steps together.

As discussed above, when the free energy was taken into account, the actual process of gene synthesis might not be as explained before. In fact, unifying the two steps has been reported [24]. However, many mutations and deletions were found in that study, which may be partly due to the use of many (ten to forty-six) shorter oligonucleotides (40 bp long).

In this study, longer oligo- nucleotides were used and as such, the deletions and errors were able to be avoided. Because the synthesized gene showed no mutations, it therefore implies that the size of the oligonucleotides used in this study might be the appropriate size for producing the right geometry in the ensemble of synthesis. Further works would be required to confirm this.

4. Conclusion

This study shows that the accurate synthesis of a gene can be done in a single step by integrating the assembly and amplification steps together. The well-designed specificity and appropriate sizing of the oligonucleotides appear to promote error-free synthesis. The recently available R packages (DNAShapeR and gqud) are easy and useful to examine the general structure of the molecule and the possibility of forming interfering G-quadruplex structure. The designed oligonucleotides are less likely to form into any undesired structures, thus they are easy to assemble into the desired gene. The analysis of MFE secondary structures and the interactions among them can be used to examine the synthesis experiment and to design oligonucleotides.

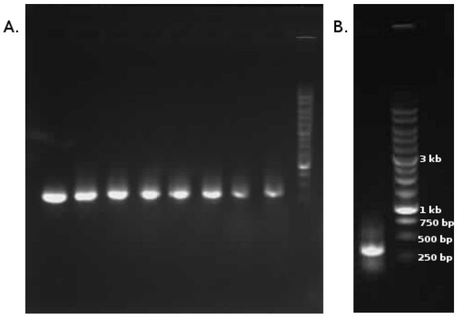
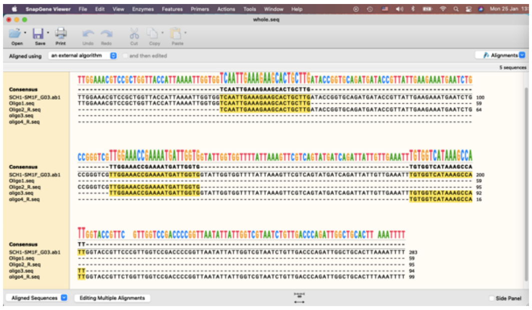


Fig. 3. The determination of the optical thermal cycle number and a single-step gene synthesis. A. The optimal number of thermal cycling experiment. Lane 9 was DNA ladders (M11, SibEnzymes, Inc.). Lanes 1-8 were the amplified products from the assembly steps of cycle number 0, 1, 2, 3, 4, 5, 10, and 20, respectively. The fading of DNA bands in lanes 7 and 8 is due to loading mistakes. Their DNA were lost because some of them did not go into the lanes. The size of the products were between 250 bp and 500 bp, when compared to the ladders. B. The one-step gene synthesis. Lane 1 was the synthesized gene and lane 2 was the DNA ladder. The size of the synthetic gene was around 300 bp.



Aligned using an external algorithm □ and then edited

```
                  TTGGAAACGTCCGCTGGTTACCATTAAAATTGGTGGTCAATTGAAAGAAGCACTGCTTGATACCGGTGCAGATGATACCGTTATTGAAGAAATGAATCTG
Consensus         ------------------------------------TCAATTGAAAGAAGCACTGCTTG-----------------------------------------
SCH1-SM1F_G03.ab1 TTGGAAACGTCCGCTGGTTACCATTAAAATTGGTGGTCAATTGAAAGAAGCACTGCTTGATACCGGTGCAGATGATACCGTTATTGAAGAAATGAATCTG 100
Oligo1.seq        TTGGAAACGTCCGCTGGTTACCATTAAAATTGGTGGTCAATTGAAAGAAGCACTGCTTG----------------------------------------- 59
Oligo2_R.seq      ------------------------------------TCAATTGAAAGAAGCACTGCTTGATACCGGTGCAGATGATACCGTTATTGAAGAAATGAATCTG 64
oligo3.seq        ----------------------------------------------------------------------------------------------------
oligo4_R.seq      ----------------------------------------------------------------------------------------------------

                  CCGGGTCGTTGGAAACCGAAAATGATTGGTGGTATTGGTGGTTTTATTAAAGTTCGTCAGTATGATCAGATTATTGTTGAAATTTGTGGTCATAAAGCCA
Consensus         --------TTGGAAACCGAAAATGATTGGTG-----------------------------------------------------TGTGGTCATAAAGCCA
SCH1-SM1F_G03.ab1 CCGGGTCGTTGGAAACCGAAAATGATTGGTGGTATTGGTGGTTTTATTAAAGTTCGTCAGTATGATCAGATTATTGTTGAAATTTGTGGTCATAAAGCCA 200
Oligo1.seq        ---------------------------------------------------------------------------------------------------- 59
Oligo2_R.seq      CCGGGTCGTTGGAAACCGAAAATGATTGGTG--------------------------------------------------------------------- 95
oligo3.seq        --------TTGGAAACCGAAAATGATTGGTGGTATTGGTGGTTTTATTAAAGTTCGTCAGTATGATCAGATTATTGTTGAAATTTGTGGTCATAAAGCCA 92
oligo4_R.seq      ------------------------------------------------------------------------------------TGTGGTCATAAAGCCA 16

                  TTGGTACCGTTC GTTGGTCCGACCCCGGTTAATATTATTGGTCGTAATCTGTTGACCCAGATTGGCTGCACTT AAATTTT
Consensus         TT---------------------------------------------------------------------------------
SCH1-SM1F_G03.ab1 TTGGTACCGTTCCCGTTGGTCCGACCCCGGTTAATATTATTGGTCGTAATCTGTTGACCCAGATTGGCTGCACTTAAAATTTT 283
Oligo1.seq        ----------------------------------------------------------------------------------- 59
Oligo2_R.seq      ----------------------------------------------------------------------------------- 95
oligo3.seq        TT--------------------------------------------------------------------------------- 94
oligo4_R.seq      TTGGTACCGTTCTGGTTGGTCCGACCCCGGTTAATATTATTGGTCGTAATCTGTTGACCCAGATTGGCTGCACTTTAAATTTT 99
```

Fig. 4. The sequence was aligned with all four sequences using SnapGene® software (from GSL Biotech; available at snapgene.com).

Acknowledgements

I thank Professor Gavin P. Reynolds for reading my manuscript and giving valuable suggestions. The support and help from my colleagues are appreciated. This work was financially supported with the Faculty Personal Account (FPA). Chemicals and materials were the leftovers from the research grants from the National Research Council of Thailand (NRCT) and the National Center for Genetic Engineering and Biotechnology (BIOTEC).

References

[1] Wimmer E, Paul AV. Synthetic poliovirus and other designer viruses: what have we learned from them? Annu Rev Microbiol. 2011;65:583-609.

[2] Dillon PJ, Rosen CA. Use of polymerase chain reaction for the rapid construction of synthetic genes. Methods Mol Biol Clifton NJ. 1993;15:263-8.

[3] Czar MJ, Anderson JC, Bader JS, Peccoud J. Gene synthesis demystified. Trends Biotechnol. 2009 Feb;27(2):63-72.

[4] Gibson DG. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol. 2011;498:349-61.

[5] Xiong A-S, Peng R-H, Zhuang J, Liu J-G, Gao F, Chen J-M, et al. Non-polymerase-cycling-assembly-based chemical gene

synthesis: strategies, methods, and progress. Biotechnol Adv. 2008 Apr;26(2):121-34.

[6] A large fragment approach to DNA synthesis: total synthesis of a gene for the protease inhibitor eglin c from the leech Hirudo medicinalis and its expression in E. coli | Nucleic Acids Research | Oxford Academic [Internet]. 2017 [cited 2017 Feb 11]. Available from: https://academic.oup.com/nar/article/12/16/6369/2385165/A-large-fragment-approach-to-DNA-synthesis-total

[7] Adams SE, Johnson ID, Braddock M, Kingsman AJ, Kingsman SM, Edwards RM. Synthesis of a gene for the HIV transactivator protein TAT by a novel single stranded approach involving in vivo gap repair. Nucleic Acids Res. 1988 May 25;16(10):4287.

[8] Prodromou C, Pearl LH. Recursive PCR: a novel technique for total gene synthesis. Protein Eng. 1992 Dec;5(8):827-9.

[9] Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene. 1995 Oct 16;164(1):49-53.

[10] Young L, Dong Q. Two-step total gene synthesis method. Nucleic Acids Res. 2004 Jan 1;32(7):e59-e59.

[11] Xiong A-S, Yao Q-H, Peng R-H, Li X, Fan H-Q, Cheng Z-M, et al. A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucleic Acids Res. 2004;32(12):e98.

[12] A Simple and Accurate Two-Step Long DNA Sequences Synthesis Strategy to Improve Heterologous Gene Expression in Pichia [Internet]. 2017 [cited 2017 Feb 11]. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036607

[13] Swainston N, Currin A, Day PJ, Kell DB. GeneGenie: optimized oligomer design for directed evolution. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W395-400.

[14] Ajoge HO. gquad: Prediction of G Quadruplexes and Other Non-B DNA Motifs [Internet]. 2017 [cited 2017 Feb 11]. Available from: https://cran.r-project.org/web/packages/gquad/index.html

[15] Chiu T-P, Comoglio F, Zhou T, Yang L, Paro R, Rohs R. DNAshapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. Bioinformatics. 2016 Apr 15;32(8):1211-3.

[16] NUPACK: Analysis and design of nucleic acid systems - Zadeh - 2010 - Journal of Computational Chemistry - Wiley Online Library [Internet]. 2017 [cited 2017 Feb 11]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jcc.21596/abstract

[17] Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009 Aug 1;25(15):1974-5.

[18] Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006;34(19):5402-15.

[19] Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015 Oct 15;43(18):8627-37.

[20] Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. The role of DNA shape in protein–DNA recognition. Nature. 2009 Oct 29;461(7268):1248-53.

[21] Vieregg JR, Vieregg JR. Nucleic Acid Structural Energetics, Nucleic Acid Structural Energetics. In: Encyclopedia of Analytical Chemistry, Encyclopedia of Analytical Chemistry [Internet]. John Wiley & Sons, Ltd, John Wiley & Sons, Ltd; 2016 [cited 2017 Mar 5]. Available from:

http://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a1418.pub3/abstract, http://onlinelibrary.wiley.com/doi/10.1002/9780470027318.a1418.pub3/abstract

[22] Lai D, Meyer IM. A comprehensive comparison of general RNA-RNA interaction prediction methods. Nucleic Acids Res. 2016 Apr 20;44(7):e61.

[23] V. Potapov and J. L. Ong, "Examining Sources of Error in PCR by Single-Molecule Sequencing," PLoS One, vol. 12, no. 1, Jan. 2017, doi: 10.1371/journal.pone.0169774.

[24] Wu G, Wolf JB, Ibrahim AF, Vadasz S, Gunasinghe M, Freeland SJ. Simplified gene synthesis: A one-step approach to PCR-based gene construction. J Biotechnol. 2006 Jul;124(3):496–503