VOL.2 NO.2September 2013 - ISSN 2233 – 1859

Operations Research Society in Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing

Available online:www.scjournal.com.ba

Efficient Algorithm for Primer Design

Faruk Berat Akcesme

International University of Sarajevo, Faculty of Engineering and Natural Sciences, HrasnickaCesta 15, Ilidža71210 Sarajevo, Bosnia and Herzegovina

Article Info

Article history:
Received 17 Sep.2013
Received in revised form 17 Oct 2013

Keywords:
primer, PCR, primer design

Abstract

PCR is one of the most popular technique that enable amplify specific region of the genome. It allows performing variety of analyses and application including DNA cloning for sequencing; functional analysis of genes; the diagnosis of hereditary diseases. Primer design is one of the most fundamental in PCR-based methods. Many different parameters need to be taken into account. There is many commercial program, some of them are online available, for the primer design. In this paper we present an algorithm which is not located in analyzing large sequence but for infrequent users.

1. INTRODUCTION

In 1983, biochemist KaryMulis invents a technique that called polymerase chains reaction (PCR) to amplify a single or few copies of piece of DNA. This technique which brings Nobel Prize in Chemistry to him enables to generate millions of copies of a particular DNA sequence. It allows performing variety of analyses and application including DNA cloning for sequencing; functional analysis of genes; the diagnosis of hereditary diseases (Kary, 1993).

The methods, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic reaction of the DNA, relies on thermal cycling. Short DNA fragments that are complementary to the target region along with DNA polymerase are key components to perform repeated amplification (Yuryev, 2002).

Since the technique relies on thermal there is need to employ a heat stable DNA polymerase. It is naturally exist in bacterium Thermusaquaticus (Boyle, 2009). The heat stable DNA polymerase called Taq polymerase and it is commercially available.

Taq Polymerase like the other polymerases can only add new nucleotides to an existing strand of DNA. For that reason there is needs for primer that serve as a starting point for DNA synthesis. Primers are generally short chemically synthesized oligonucleotides with a short length of about 20 base pairs. Since the primer is crucial component for the PCR application, primers needs to be design very carefully with considering some important biological and mathematical aspect. Because the primers are functioning in pairs it has to designed as so called forward and reverse primers. They have designed as they will be extended towards each of cover the given target region.

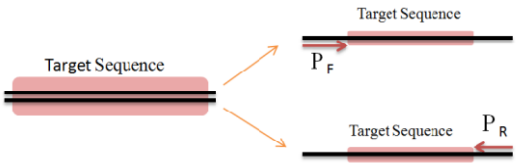

Figure 1Primer annealing

PCR begins with approximately 95 C0. This step called denaturation. In this step the double strandedDNA converted into single stranded DNA. In the second step of the PCR, annealing step, primers hybridize to the sequence, this step has temperature ranging between 45-62 0C. The final step called extension, in this step intermediate temperature about 72 0C applied to extend the hybridized primers in both directions. This all steps are called one cycle and each of the PCR application typically performed in 25-45 cycles.

1.1. Principle of Primer Design

After the usage of PCR in research and clinic, designing primer becomes critical. There is many software and desktop programs, even some of them are online, that are aimed to design a primer in efficient way. There are several criteria to design a efficient primer and these criteria`s are directly effecting the efficiency of PCR.

• Lengths: Generally between 18-22 base pairs.

• GC contents: No large GC rich or deficient content.

• Complementarily: No dimer potential.

• Secondary structure in primer: no hairpin secondary structure.

• Tm (melting temperature) of primers: Annealing temperature within user set limits (Kampke, 2001).

To perform PCR consistently in every cycle Melting Temperature of the primers must be similar. Generally programs use the original nearest neighbor method (Breslauer, 1986) to determine Tm. Beside the nearest neighbor method, programs also use same method with empirically determined thermodynamics values to determine Tm with the DNA target according to the following equation (Sugimoto, 1996). The other one of the important criteria is G/C content. Since G-C content contains three hydrogen bonds whereas an A-T content has two. Three hydrogen bonds make the G-C content stronger so the melting temperature needs to be increase in order to break these bounds. Generally, G/C content of the assumption is to arrange GC content between 40 and 60% (Lowe, 1990).

Self-complementation is also one of the biological facts that must be avoided when designing primers. Primers enable either the formation of secondary structure in the single stranded oligonucleotides, or binding another copy of itself (Haas, 1998). This self-complementation called primer dimer. Both of the complementation must be considered when designing the primer.

Primer length is a criterion which is directly related to the uniqueness. Also hybridization stability is related to the length of the primer. While increasing the length of primer cause mismatch tolerance, shorter primer has specificity problem. Especially according to the real PCR data, short primer don't show specificity to the target region of the sequence and it manipulates the result of PCR.

1.2. Algorithm in primer design

Software that is developed for biological application gives insight to the field of bioinformatics. There are many different programs for designing primers and some of them are available online. These programs designed by some companies that are engaged in bio software development include: AlkamiBiosystems, Molecular Biology Insights, PREMIER Biosoft International, IntelliGenetics Inc., Hitachi Inc., DNA Star, Advanced American Biotechnology and Imaging. Some scientists have also developedAlgorithm and computer programs for various purposes of primer design (Abd-ElsalamKamel, 2003).

In 2012 BLAST introduce a new software toll called Primer BLAST. With this program BLAST claim that program design target-specific primers. It uses global alignment algorithm to ensure a full primer target. The very useful features of the program is checking the pre-existing primers and able to place primers based on exon intron location and excluding single nucleotide polymorphism sites in primers. Primer BLAST uses Needleman-Wunsch algorithm to perform pairwise alignment through the entire length of the sequences. This program was implemented using the C++ toolkit and the primer3 C programming interface (Jian Ye et al, 2012).

1.3. Parameters of primer design

1.3.1. Primer length:

Specificity, temperature and annealing time is partially depends on the length so length of primer is very important for PCR. Practices have shown that primers typically design in 18-30 nucleotides in length (Wu, 1991). Technically it is not difficult to arrange the primer length but the other variable needs to be correlated with it.

1.3.2. Melting temperature

In molecular biology oligonucleotide melting temperature is important for many techniques. For these reason several formulas have been developed to estimate melting temperature. One of the very popular one is nearest-neighbor model. Generally it gives highest accuracy for Tm prediction. İt is not certain for adjusting these models for the effects of reagents commonly used in PCR. Ahsenet all design an experiment that investigate 475 matched or mismatched target-probe duplexes. This data set was used to evaluate the contributions of[Mg21], [dNTPs], and [DMSO] in N-N calculations. This study succeeded to predict accurate melting temperature with using N-N calculation or empirical formulas (Wallacem, 1979).

1.3.3. GC Content:

Strength if annealing can be determined with looking GC content of primer since it is highly depends on GC content. Primers that have less than 50% GC content should be extended beyond 18 bases to keep the melting temperature above the recommended lower limit of 50 0C. Rychliket all describe the dissociation temperature. Since some algorithm are based solely on length of the oligonucleotide but reliable only when hybridizations are performed in some specific solution (Dieffenbach,1995). More practical way to determine the dissociation temperature has been described by Sugget all as in the following formula (Suggs, 1981):

$$T_d = 2°C \times \text{number of AT bp} + 4°C \times \text{number of GC bp}$$

Tool name	Description	www
CODEHOP	**C**onsensus **D**egenerate **H**ybrid **O**ligonucleotide **P**rimers; degenerate PCR primer design; will accept unaligned sequences.	http://blocks.fhcrc.org/codehop.html
Gene Fisher	Interactive primer design tool for standard or degenerate primers; will accept unaligned sequences.	http://bibiserv.techfak.uni-bielefeld.de/genefisher/
DoPrimer	Easily design primers for PCR and DNA sequencing.	http://doprimer.interactiva.de/
Primer3	Comprehensive PCR primer and hybridization probe design tool; many options but easy to accept defaults at first.	http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi http://www.basic.nwu.edu/biotools/Primer3.html http://www.justbio.com/primer/index.php
Primer Selection	Select PCR primers from nucleotide sequence.	http://alces.med.umn.edu/rawprimer.html
Web Primer	Allow alternative design of primers for either PCR or sequencing purpose.	http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer
PCR Designer	For restriction analysis of sequence mutations.	http://cedar.genetics.soton.ac.uk/public_html/primer.html
Primo Pro 3.4	Reduces PCR noise by lowering the probability of random primering.	http://www.changbioscience.com/primo/primo.html
Primo Degenerate 3.4	Primo Degenerate 3.4 designs PCR primers based on a single peptide sequence or multiple alignments of proteins or nucleotides.	http://www.changbioscience.com/primo/primod.html
PCR Primer Design	An application that designs primers for PCR or sequencing purposes.	http://pga.mgh.harvard.edu/servlet/org.mgh.proteome.Primer
The Primer Generator	The program analyzes the original nucleotide sequence and desired amino acid sequence and designs a primer that either has a new restriction enzyme site or is missing an old one.	http://www.med.jhu.edu/medcenter/primer/primer.cgi
EPRIMER3	Picks PCR primers and hybridization oligos (EMBOSS).	http://bioweb.pasteur.fr/seqanal/interfaces/eprimer3.html
PRIMO	Prediction of forward and reverse oligonucleotide Primers.	http://bioweb.pasteur.fr/seqanal/interfaces/primo.html3 http://atlas.swmed.edu/primo/primo_form.html
PrimerQuest	A primer design tool.	http://www.idtdna.com/biotools/primer_quest/primer_quest.asp
MethPrimer	Design primers for methylation PCRs.	http://itsa.ucsf.edu/~urolab/methprimer/index1.html
Rawprimer	A tool for selection of PCR primers.	http://alces.med.umn.edu/rawprimer.html
MEDUSA	A tool for automatic selection and visual assessment of PCR primer pairs.	http://www.cgr.ki.se/cgr/MEDUSA/
The Primer Prim'er Project	Software suite that completely automates the PCR primer design process.	http://www-nmr.cabm.rutgers.edu/bioinformatics/Primer_Primer_ Project/Primer.html
Oligonucleotides for the PCR	Seek oligonucleotides on both sides of an area.	http://www.citi2.fr/bio2/Oligo2lib.html
GAP	Genome- wide Automated Primer finder servers.	http://promoter.ics.uci.edu/Primers/

table1 online primer design program

Rychliket all describes an alternative computer program that called OLIGO which computes the Td values based on

nearest neighbor thermodynamic parameter. OLIGO is described as follows [17]:

$$T_d = \frac{\Delta H}{\Delta S + R \times \ln(C/4)} - 273.15°C - t$$

where AH and AS are the enthalpy and entropy for helix formation, respectively, R is molar gas constant [1.987 (cal/°C x mol)], and C is concentration of the probe.

1.3.4. 3'-End Sequence:

Mis-priming need to be minimizes or completely be avoided, if possible, to perform good PCR. Kwok and his friends investigate the effects of primer-template mismatches on PCR. Since the 5` ends and 3` ends should be sticky. These ends needs to be rich in GC, and sites could potentially anneal at multiple sites in the template. They investigate whether the effects of mismatches on PCR were symmetrical. For instance, would a G: T, primer-template, has the same effect on PCR as T: G, template-primer mismatches. They figure out that there is a sympatric effect of mismatches and they indicate that could be due to flanking sequences (Kwok,1990).

2. METHODS

The algorithm designed with considering 4 criteria. Each of the criteria is listed as follow; The DNA nucleotide sequence is taken as input of character array. In this algorithm both end of the sequence are considered as a target not the entire DNA sequence; to illustrate in this algorithm. The first 100 nucleotide of the sequence are taken as possible target. User allowed enlarging the target area. The length of the sequence is determined by the user, which can be range between 18-25. In this basis the number of GC are counted and their proportion has taken to the entire bases. Since the 50 % has strong certainty, it is counted in range between 45% and 55%. The basis which satisfies this percentage, they are taken as possible candidate of primers.

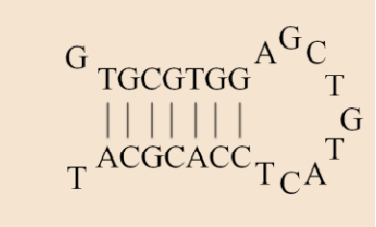

Figure 2self complementation

Figure 3 primer dimer

The algorithm avoids having complementary content in each ends of the sequence. For instance if the one of the candidate includes a strand ATC the other pair of primer must be free of complementary of it which is TAG. The last part of the algorithm is considering the possible hairpin loop composition, to avoid this each of the pair of primers must be free of inter complementation. For example, if the primer includes a strand ATC the in its content, it has to be free of complementary nucleotides, which is TAG.

Designed algorithm can be summarized as

No repeat structure in the sequence

- No large GC rich or deficient regions
- No nucleotide stretches
- No hairpin secondary structure
- No dimer potential
- Annealing temperature within user set limits

3. DISCUSSION AND RESULT

PCR is one of the most fundamental techniques in biology. There is many different approach to develop a good algorithm for designing PCR primer. In this paper many biological aspect was considered to design an efficient algorithm to design proper primer to increase the efficiency of PCR. Algorithm has been run successfully with considering the mentioned biological facts. This algorithm allow user to adjust the target sequence manually to protect the functional parts of the sequence. However, there is need to use the primer that are already designed. Since BLAST is using its own database which contains entire genome of many different species, it has a big advantage over the other algorithms that are designed for the same purpose. There is need for more accurate and powerful program to design efficient primer, since the biological knowledge are expending day by day.

REFERENCES

Kary B. Mullis, Michael Smith The Nobel Prize in Chemistry 1993,Nobel Lecture, December 8, 1993

Yuryev A, Huang J, Pohl M, Patch R, Watson F, Bell P, Donaldson M, Phillips MS, Boyce-Jacino MT: Predicting the success of primer extension genotyping assays using statistical modeling. Nucleic Acids Res 2002, 30(23):e131.

Boyle B, Dallaire N, MacKay J: Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. BMC Biotechnol 2009, 9:75.

Kampke, T, Kieninger, M and Mecklenbug, M. (2001) Efficient primer design algorithms. Bioinformatics. 17(3), 214-225.

Breslauer KJ, Frank R, Blocker H, Marky LA.(1986) Predicting DNA duplex stability from the base sequence.ProcNatlAcadSci U S A. 83(11), 3746-3750.

Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Research. 24, 4501-4505.

Lowe, T. et al. (1990) A computer program for selection of oligonucleotide primers for polymerase chain reactions. Nucleic Acids Research. 18(7), 1757-1761.

Haas, S., Vingron, M., et al. (1998) Primer Design for large scale sequencing. Nucleic Acids Research. 26(12), 3006-3012.

Abd-ElsalamKamel A, 2003 Bioinformatics tools and guideline for PCR primer design,

Jian Ye et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction, BMC Bioinformatics 2012, 13:134 Page 2 of 11http://www.biomedcentral.com/1471-2105/13/134

Wu DYW, Ugozzoli L, Pal BK, Qian J, Wallace RB (1991). The effect of temperature and oligonucleotide primer length on specificity and efficiency of amplification by the polymerase chain reaction. DNA Cell Biol. 10: 233-238.

Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res. 6: 3543-3557.

15.Dieffenbach CW, Lowe TMJ, Dveksler GS (1995). General Concepts for PCR Primer Design. In: PCR Primer, A Laboratory Manual, Dieffenbach CW, Dveksler GS Ed., Cold Spring Harbor Laboratory Press, New York, 133-155.

16. Suggs, S.V., Hirose, T., Miyake, E.H., Kawashima, M.J., Johnson, K.I., and Wallace, R.B. (1981) In

D.D. Brown (ed.), ICN-UCLA Symp. Dev. Biol. Using Purified Genes. Academic Press Inc., New York,Vol. 23, pp., 683-693.

Suggs, S.V., Hirose, T., Miyake, E.H., Kawashima, M.J., Johnson, K.I., and Wallace, R.B. (1981) In

D.D. Brown (ed.), ICN-UCLA Symp. Dev. Biol. Using Purified Genes. Academic Press Inc., New York,Vol. 23, pp., 683-693.

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky J (1990). Effects of primer-template mismatches on the polymerase chain reaction: Human Immunodeficiency Virus 1 model studies. Nucleic Acids Res. 18: 999-1005.`