NTMG (N-terminal Truncated Mutants Generator for cDNA): an automatic multiplex PCR assays design for generating various N-terminal truncated cDNA mutants

Yung-Fu Chen, Rung-Ching Chen, Lin-Yu Tseng, Elong Lin, Yung-Kuan Chan and Ren-Hao Pan

1Department of Health Services Management, China Medical University, Taichung, Taiwan, ROC, 2Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan, ROC, 3Department of Computer Science, National Chung Hsing University, Taichung, Taiwan, ROC, 4Department of Food Science, Central Taiwan University of Science and Technology, Taichung, Taiwan, ROC and 5Department of Management Information System, National Chung Hsing University, Taichung, Taiwan, ROC

Received January 31, 2007; Revised April 3, 2007; Accepted April 15, 2007

ABSTRACT

The sequential deletion method is generally used to locate the functional domain of a protein. With this method, in order to find the various N-terminal truncated mutants, researchers have to investigate the ATG-like codons, to design various multiplex polymerase chain reaction (PCR) forward primers and to do several PCR experiments. This web server (N-terminal Truncated Mutants Generator for cDNA) will automatically generate groups of forward PCR primers and the corresponding reverse PCR primers that can be used in a single batch of a multiplex PCR experiment to extract the various N-terminal truncated mutants. This server is available at http://oblab.cs.nchu.edu.tw:8080/WebSDL/.

INTRODUCTION

The sequential deletion method is used to locate the functional domain of a protein. With this method, in order to find the various N-terminal truncated mutants, researchers have to investigate the ATG-like codons, to design various multiplex polymerase chain reaction (PCR) forward primers and to do several PCR experiments. This web server (N-terminal Truncated Mutants Generator for cDNA) will automatically generate groups of forward PCR primers and the corresponding reverse PCR primers that can be used in a single batch of a multiplex PCR experiment to extract the various N-terminal truncated mutants. This saves much time and money for those who use the sequential deletion method in their research. This server is available at http://oblab.cs.nchu.edu.tw:8080/WebSDL/.

SYSTEM

In this section, we describe the input to the NTMG, the methodology of the NTMG and the output of the NTMG. Since the primer design and the multiplex PCR primer design are two important parts of the NTMG, some factors concerning the primer design such as the leading to the binding of the 5S rRNA to the nascent ribosomal protein L5. The formation of 5S rRNP complex facilitates the nuclear entry of the protein L5. Lin et al. (1) used an in vitro translation system to investigate how and when 5S rRNA triggers the formation of the eukaryotic 5S rRNP. The L5 and truncated L5 mutant mRNAs were prepared on a large scale for their investigation and a great amount of time was needed to manually modify the in-frame pattern of ATG start codon for conventional PCR and truncated mutant translation experiments.

In order to save time and money needed in the traditional sequential deletion method, this web-based application system NTMG is proposed to automatically do the multiplex PCR assays design in order to generate the various N-terminal truncated mutants. Given a protein cDNA sequence, the NTMG will first find those ATG-like codons that are suitable to act as the starting positions of truncated mutants. Then, the NTMG will design the forward primers for all possible truncated mutants. Finally, with all these primers, the NTMG will choose those primers that can be divided into the least number of groups such that each group constitutes a multiplex PCR assay.
primer length and the melting temperature are input as parameters.

There are also some factors that may affect the multiplex PCR amplification with multiple primers in the same tube (2,3). These factors include the cross-dimerization, the melting temperature, the products co-existence and others. All these factors are also input as parameters. Each input parameter has a default value but users may change that value. We first introduce the input parameters in the following subsection.

**Input Parameters**

Figure 2 gives the input screen of the NTMG. On the top of the screen, the cDNA sequence, the start codon address and the stop codon address are input. Then comes four classes of parameters:

(i) **Primer Criteria (4–7)**

These include the forward primer length (default 20–30 bp), the reverse primer length (default 14–28 bp) and the GC content (default 40-60%).

(ii) **Temperature Criteria (5,8–10)**

These include melting temperature (default 51–60°C), the melting temperature range for each group (default 5°C), the molar concentration of monovalent cation (default 50 mmol/l), the molar concentration of Mg²⁺ (default 1.25 mmol/l) and the dNTPs concentration (default 0.02 mmol/l).

(iii) **Complementary Criteria (4,8,10–12)**

These include the terminal repeated sequence (default 3 bp), the intra self-complementary sequence (default 3 bp), the specificity (default 65%), the cross-dimer distance (default 10 bp), the cross-dimer – total similarity (default 50%) and the cross-dimer – terminal similarity (default 3 bp).

(iv) **Grouping Criteria**

These include the product length difference (default 80 bp) and the maximum number of primers in each group (default 16).

**Methodology**

The flowchart of the NTMG is depicted in Figure 3.

First, a cDNA sequence is input, then the NTMG searches the sequence with in-frame criterion in order to find all the ATG-like codons. An ATG-like codon is a XTG, an AXG or an ATX with X representing A, T, C or G. For each ATG-like codon, the NTMG generates candidate primers using the sliding window. The NTMG also modifies the ATG-like codon into the ATG codon in each candidate primer. Hence, each candidate primer
contains the ATG start codon with in-frame criterion that can translate correctly to truncated mutants.

After that, the NTMG applies the primary criteria and the secondary criteria to check if the candidate primers satisfy all the criteria. Those candidate primers that pass the criteria checking are then divided into classes according to the melting temperature. Each class contains primers whose melting temperatures are within the same range (default 5°C). So the primers chosen from the same class may be used in the same PCR experiment. For candidate primers in the same class, the NTMG does the cross-dimer check and the co-existence check and builds the cross-dimer matrix and the co-existence matrix. These two matrices are then ANDed to produce a new matrix which acts as an adjacency matrix of a graph. Previous studies (13,14) proposed the transformation of minimizing the number of primers into the finding of the maximum clique in a graph. Next, each class is further divided into subclasses according to the positions of the ATG-like codons. The proposed genetic algorithm is then applied for each class in order to choose a set of primers, one from each subclass and the objective is to

---

Figure 3. The flowchart of the NTMG.
divide these primers into as few groups as possible such that each group of primers can be put in a tube in the PCR experiment. That is, the genetic algorithm tries to find a good multiplex PCR assay design. A heuristic maximum clique algorithm is proposed to calculate the fitness of a chromosome in the genetic algorithm. This heuristic algorithm tries to find the maximum clique in the graph previously mentioned and the maximum clique corresponds to the minimum number of grouping. Finally, the NTMG finds the corresponding reverse primer and outputs the results.

Environment
The NTMG is written in Java using Java 2 Platform standard Edition 5.0 Development Kit (J2SDK) and employs the java server page (JSP) on the Apache Tomcat Server (http://tomcat.apache.org/).

Output
The NTMG outputs the number of primer groups, the number of forward primers, the number of reverse primers and the primers in each group (Figure 4 shows a solution report).

CONCLUSION AND FUTURE WORK
A web-based application system called the NTMG is provided for researchers who need to apply the sequential deletion method to locate the functional domain of a protein. After input the cDNA sequence, the NTMG automatically generates groups of primers. Each group of primers can be put in a tube and all tubes can be accommodated in a single batch of the multiplex PCR amplification under the same condition. Thus, time and money can be saved. We conducted a wet laboratory experiment on the multiplex PCR assay design proposed by the NTMG on input HL5 cDNA. The NTMG found 48 forward primers and one reverse primer and it divided them into 8 groups. In the wet PCR experiment, 44 PCR products had been found and the success rate is 91.7% (see Supplementary Data). Hence, the NTMG is of practical use to researchers who need to apply the sequential deletion method.

As a future work, we plan to develop the more general multiplex PCR assay design. Given a set of PCR experiment requirements, we plan to develop a system that can automatically find the primers and try to divide the primers into as few groups as possible such that the primers in each group can be put in a tube and all tubes can be accommodated in a single batch of the multiplex PCR experiment.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
This study was supported in part by China Medical University under grand CMU95-184. This study was also partially supported by the National Science Council of R.O.C. under the contract NSC 95-2221-E-005-078. Funding to pay the Open Access publication charges for this article was provided by National Chung Hsing University, ROC.
Conflict of interest statement. None declared.

REFERENCES

1. Lin,E., Lin,S.W. and Lin,A. (2001) The participation of 5S rRNA in the co-translational formation of eukaryotic 5S ribonucleoprotein complex. *Nucleic Acids Res.*, **29**, 2510–2516.
2. Edwards,M.C. and Gibbs,R.A. (1994) Multiplex PCR: advantages, development, and applications. *PCR Methods Appl.*, **3**, 65–75.
3. Bourque,S.N., Valero,J.R., Mercier,J., Lavoie,M.C. and Levesque,R.C. (1993) Multiplex polymerase chain reaction for detection and differentiation of the microbial insecticide Bacillus thuringiensis. *Appl. Environ. Microbiol.*, **59**, 523–527.
4. Kampke,T., Kieninger,M. and Mecklenburg,M. (2000) Efficient primer design algorithms. *Bioinformatics*, **17**, 214–225.
5. Wu,J.S., Lee,C., Wu,C.C. and Shiue,Y.L. (2004) Primer design using genetic algorithm. *Bioinformatics*, **20**, 1710–1717.
6. Lowe,T., Sharefkin,J., Yang,S.Q. and Dieffenbach,C.W. (1990) A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res.*, **18**, 1757–1761.
7. Li,P., Kupfer,K.C., Davies,C.J., Burbec,D., Evans,L.A. and Garner,H.R. (1997) PRIMO: a primer design program that applies base quality statistics for automated large-scale DNA sequencing. *Genomics*, **40**, 476–485.
8. Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (1990) PCR Protocols: A Guide to Methods and Applications Academic Press, New York.
9. von Ahsen,N., Wittwer,C.T. and Schutz,E. (2001) Oligonucleotide melting temperature under PCR conditions: nearest-neighbor corrections for Mg2+, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin. Chem.*, **47**, 1956–1961.
10. Chen,S.H., Lin,C.Y., Cho,C.S., Lo,C.Z. and Hsiung,C.A. (2003) Primer Design Assistant (PDA): a web-based primer design tool. *Nucleic Acids Res.*, **31**, 3751–3754.
11. Haas,S., Vingron,M., Poustka,A. and Wiemann,S. (1998) Primer design for large scale sequencing. *Nucleic Acids Res.*, **26**, 3006–3012.
12. Kwok,S., Kellogg,D.E., McKinney,N., Spasic,D., Goda,L., Levenson,C. and Sninsky,J.J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human Immunodeficiency Virus 1 model studies. *Nucleic Acids Res.*, **18**, 999–1005.
13. Ouyang,Q., Kaplan,P.D., Liu,S. and Libchaber,A. (1997) DNA solution of the maximal clique problem. *Science*, **278**, 446–449.
14. Linhart,C. and Shamir,R. (2005) The degenerate primer design problem: theory and applications. *J. Comput. Biol.*, **12**, 431–456.