W66–W70 *Nucleic Acids Research, 2007, Vol. 35, Web Server issue*
doi:10.1093/nar/gkm305

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[1], Rung-Ching Chen[2], Lin-Yu Tseng[3,*], Elong Lin[4], Yung-Kuan Chan[5] and Ren-Hao Pan[3]

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

INTRODUCTION

The sequential deletion method and other biological and biochemical experiments are generally used to locate the functional domain of a protein. For example, a previous study (1) used the sequential deletion method accompanied by manual PCR primers design to generate the N-terminal truncated mutants of different lengths (Figure 1), which in turn were used in further biological experiments to decipher the functional domain of the 5S RNA-protein complex (5S rRNP). The 5S rRNP is believed to be formed by a co-translation event 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.

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

*To whom correspondence should be addressed. Tel: +886-4-22874020; Fax: +886-4-22853869; Email: lytseng@cs.nchu.edu.tw

The authors wish it to be known that the first two authors should be regarded as joint First Authors

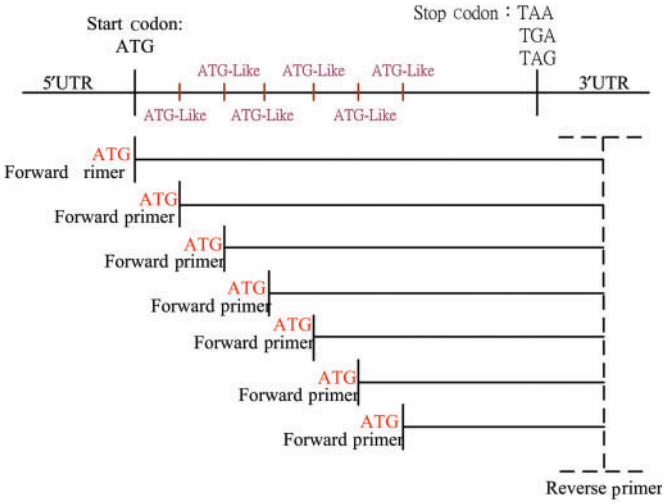

Figure 1. The illustration of the produced N-terminal truncated mutants.

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 Mg^{2+} (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).

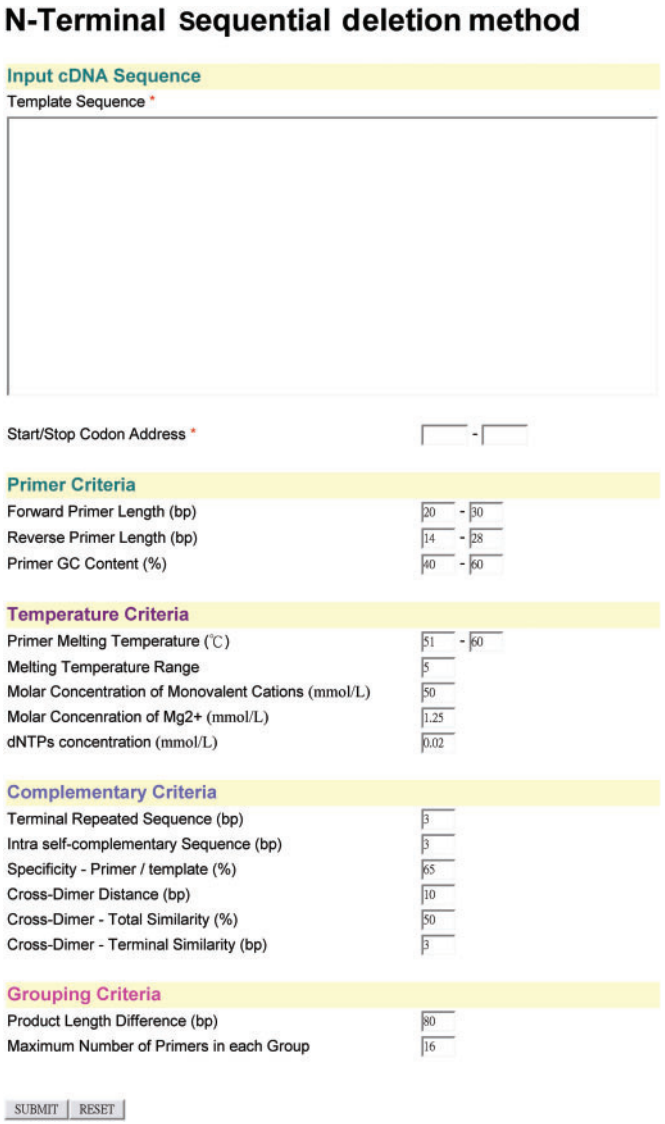

Figure 2. The input screen of the NTMG.

(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

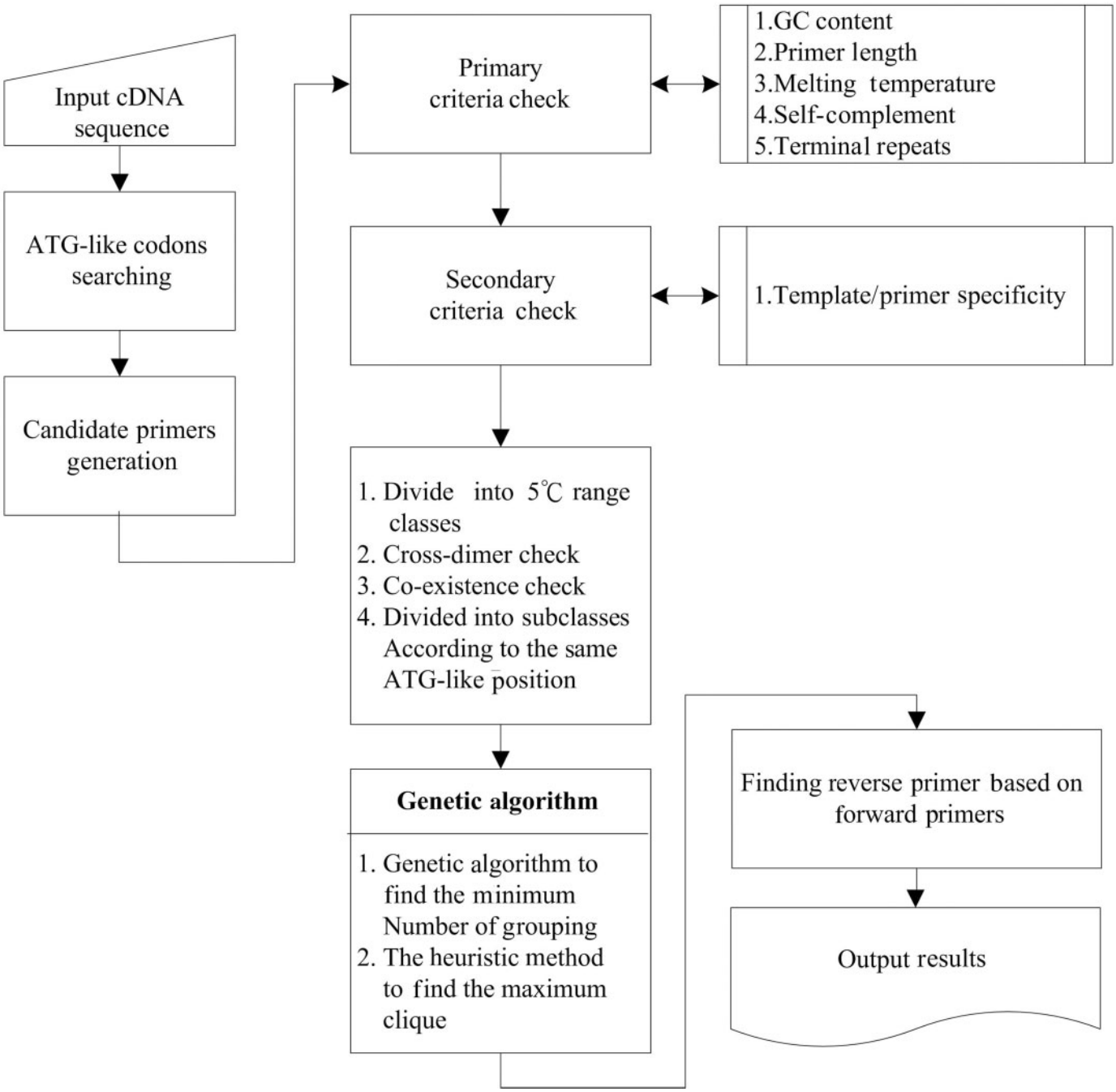

Figure 3. The flowchart of the NTMG.

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

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 researches 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.

N-Terminal Sequential Deletion Method
Genetic Algorithm & Clique Classification

Melting Temperture :51 ~ 55

Total Primer Groups : 8
Total Primer Numbers : 48
Forward Primer Numbers : 47
Reverse Primer Numbers : 1

Group : 0

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	113	27	113	54	ATGAAATTTAGAAGACGACGAGAGGGT	113	969	857
F	2	198	25	212	54	CACCCAAATACAGGATGATAGTTCG	198	969	772
F	3	278	21	281	53	ATGATGGTCTGCGCAGCGTAT	278	969	692
F	4	379	24	383	55	CCGCATGCTTCTCAATAGGTTTGG	379	969	591
F	5	461	21	464	55	AGCATGGATGGTCAGCCAGGT	461	969	509
F	6	545	21	548	55	GCCATGAAGGGAGCTGTGGAT	545	969	425
F	7	648	21	650	55	GGATGCACATCATGGGCCAGA	648	969	322
F	8	740	22	740	51	ATGAAGAACAGCGTAACTCCAG	740	969	230
F	9	847	25	848	55	AATGAAGAGGTGGAACCGTCCCAAA	847	969	123
R	0	947	Cross Dimer : 0/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 1

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	89	27	89	54	ATGGCCTACTTTAAGAGATACCAAGTG	89	969	881
F	2	213	23	215	53	TGATGGTTCGTGTGACAAACAGA	213	969	757
F	3	304	27	308	54	CGAAATGCCAAAATATGGTGTGAAGGT	304	969	666
F	4	404	26	404	54	ATGGACAAGATCTATGAAGGCCAAGT	404	969	566
F	5	496	22	497	55	TATGGATGCAGGCCTTGCCAGA	496	969	474
F	6	577	23	578	53	TATGCCTCACAGTACCAAACGAT	577	969	393
F	7	665	23	683	53	CAGAATGTTGCAGATTACATGCG	665	969	305
F	8	771	27	776	54	AGGAGATGTATAAGAAAGCTCATGCTG	771	969	199
F	9	851	22	851	53	ATGAGGTGGAACCGTCCCAAAA	851	969	119
R	0	947	Cross Dimer : 1/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 2

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	53	20	62	53	TTCCGCAGGATGGGGTTTGT	53	969	917
F	2	238	27	239	54	TATGATTTGTCAGATTGCTTATGCCCG	238	969	732
F	3	321	22	323	51	GTATGAAGGTTGGCCTGACAAA	321	969	649
F	4	410	24	410	54	ATGATCTATGAAGGCCAAGTGGAG	410	969	560
F	5	548	21	551	55	CTGATGGGAGCTGTGGATGGA	548	969	422
F	6	653	21	656	51	CACATGATGGGCCAGAATGTT	653	969	317
F	7	738	24	743	52	ACATAATGAACAGCGTAACTCCAG	738	969	232
F	8	851	20	854	51	AAGATGTGGAACCGTCCCAA	851	969	119
R	0	947	Cross Dimer : 2/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 3

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	240	24	242	54	TCATGTGTCAGATTGCTTATGCCC	240	969	730
F	2	321	26	326	54	GTGTGATGGTTGGCCTGACAAATTAT	321	969	649
F	3	412	24	413	55	GATGTATGAAGGCCAAGTGGAGGT	412	969	558
F	4	559	24	560	55	TATGGATGGAGGCTTGTCTATCCC	559	969	411
F	5	648	22	659	55	GGAAGCACATCATGGGCCAGAA	648	969	322
F	6	753	21	764	51	TAACTCCAGACATGATGGAGG	753	969	217
F	7	867	25	872	54	CCAAAATGTCCCTTGCTCAGAAGAA	867	969	103
R	0	947	Cross Dimer : 3/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 4

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	275	25	278	55	GATATGATAGTCTGCGCAGCGTATG	275	969	695
F	2	374	22	374	55	ATGGCCCGCAGGCTTCTCAATA	374	969	596
F	3	569	21	572	53	GGCATGTCTATCCCTCACAGT	569	969	401
F	4	675	25	695	52	CAGATTACATGCGCTACTTAATGGA	675	969	295
F	5	884	20	887	53	CAGATGAAGGATCGGGTAGC	884	969	86
R	0	947	Cross Dimer : 4/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 5

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	250	21	251	51	GATGGCTTATGCCCGTATAGA	250	969	720
F	2	426	25	428	52	AAATGGAGGTGACTGGTGATGAATA	426	969	544
F	3	753	21	767	51	TAACTCCAGACATGATGGAGG	753	969	217
F	4	887	25	890	54	AAGATGGATCGGGTAGCTCAAAAGA	887	969	83
R	0	947	Cross Dimer : 5/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 6

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	450	27	455	54	ACAATATGGAAAGCATTGATGGTCAGC	450	969	520
F	2	798	22	800	51	CTATGCGAGAGAATCCAGTCTA	798	969	172
F	3	905	24	908	54	CAAATGAAGGCAAGCTTCCTCAGA	905	969	65
R	0	947	Cross Dimer : 6/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Group : 7

Type	ID	Start Addr	Length	ATG Addr	Tm	Primer Sequence	Product Start	Product End	Product Length
F	1	430	26	434	54	GGAGATGACTGGTGATGAATACAATG	430	969	540
F	2	909	20	911	51	AGATGGCAAGCTTCCTCAGA	909	969	61
R	0	947	Cross Dimer : 7/0		53	GAAAATTGCTGGGTTTAGCTCTC			

Figure 4. The output solution of the NTMG.

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., Burbee,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.