

Summary

A computer program for research of optimal primers for PCR

This paper describes a new computer program to choose oligonucleotides for polymerase chain reaction (PCR). This program, OLIGOTEST, runs on IBM PC XT, AT or IBM compatible computers. The user can set all the parameters used by the program: maximum number of complementarity in contiguous bp between both primers, oligonucleotides length, maximum and minimum GC content. The user also has the possibility to test the maximum bp which are complementary in a partial 3' sequence of the oligonucleotides. This is important in order to avoid a match between the two primers at the 3' end which would facilitate primer-dimer formation. In addition, the program indicates the maximum number of complementary bases between the two primers. Other options will display the possible hairpin-loop structure for each primer or reveal the degenerate oligonucleotide corresponding to a given peptidic sequence. An application of the OLIGOTEST program is presented: research of primers for the amplification of the inserts of a subtracted cDNA library in pUEX1.

ADRESSES

C. Beroud : interne en pharmacie. C. Antignac : chargée de recherche au Cnrs. C. Jeanpierre : chargée de recherche à l'Inserm. C. Junien : professeur de génétique, praticien hospitalier. Inserm U. 73, château de Longchamp, bois de Boulogne, 75016 Paris, France.

Un programme informatique pour la recherche d'amorces pour l'amplification par PCR

Christophe Beroud, Corinne Antignac, Cécile Jeanpierre, Claudine Junien.

L'efficacité des réactions d'amplification par PCR dépend de nombreux paramètres, et en particulier de la façon dont sont choisis les oligonucléotides servant d'amorces. Bien que ce choix reste quelque peu empirique, certains paramètres sont particulièrement importants : distribution au hasard des bases, pourcentage des GC d'environ 50 %, absence de structures secondaires surtout à l'extrémité 3', absence de complémentarité entre les deux amorces pour éviter la formation de dimères.

Afin de pallier la complexité d'une recherche par voie manuelle du couple d'amorces répondant le mieux à ces critères, l'utilisation d'un programme informatique s'imposait. Deux voies d'approche pouvaient alors être envisagées, de façon à obtenir des temps de calcul réduits tout en conservant une souplesse dans les critères de choix. La première a été utilisée par Lowe *et al.* [1] ainsi que par Ricklik *et al.* [2]. Elle consiste à définir *a priori* certains des critères de sélection (séquence de type GC à l'extrémité 3', taille de 18 à 22 nucléotides, maximum de complémentarité de 4 bases contiguës entre les 2 amorces [1] ou maximum de stabilité de l'oligonucléotide [2]) et

à laisser à l'utilisateur quelques critères de choix (pourcentage de GC, taille de la région à amplifier, et température de fusion (T_m) du produit d'amplification [1] ou taille de l'oligonucléotide et maximum de complémentarité entre les deux amorces [2]). L'utilisation d'une telle approche permet de tester un grand nombre de solutions et d'utiliser directement des fichiers de séquences extraits des banques de données. Cependant l'utilisateur n'a qu'une liberté de choix limitée.

La seconde voie d'approche qui a pour but de laisser la totale maîtrise des critères de sélection à l'utilisateur a été choisie pour l'établissement de notre programme OLIGOTEST. Ce programme permet non seulement de déterminer les caractéristiques d'un couple d'oligonucléotides, mais aussi de rechercher, à partir de séquences de longueur variable et parmi toutes les possibilités, les couples d'oligonucléotides qui vont répondre à différents critères de sélection.

Présentation du programme

Cinq options principales peuvent être choisies.

1. *Research of optimal pairs of primers :*

la recherche des couples d'amorces s'effectue à partir de deux séquences connues d'une taille maximale de 256 bases situées de part et d'autre de la séquence à amplifier. Ces séquences peuvent être saisies au clavier ou provenir de fichiers ASCII. Les critères suivants sont alors choisis :

- (a) taille des oligonucléotides (non limitée si ce n'est par la taille de la séquence) ;
- (b) maximum de complémentarité entre bases contiguës des deux oligonucléotides (non limité, les valeurs usuelles sont 2 ou 3) ;
- (c) maximum et minimum de pourcentage en bases GC (cette option est facultative) ;
- (d) utilisation du critère b sur une partie seulement de l'oligonucléotide (cette option est facultative).

Cette dernière option est particulièrement utile dans le cas où aucune solution ne répond à des critères très stricts. Sachant que la polymérase de *Thermophilus aquaticus* peut rapidement initier une amplification à la partie 3' de doubles brins même instables, il est important de conserver des critères stricts au moins pour la partie 3' des oligonucléotides.

L'algorithme utilisé consiste à rejeter les solutions qui ne correspondent pas aux critères de sélection (voir encadré). Pour chaque paire d'amorces répondant aux critères choisis, le programme calcule la température de fusion (Tm), le pourcentage de GC et le coefficient d'extinction molaire (K) utilisé pour le calcul de la concentration des oligonucléotides [3]. De plus, il indique le taux maximal de complémentarité entre les deux oligonucléotides, sur toute leur longueur ainsi qu'entre paires de bases contiguës. Quand plusieurs types d'appariement entre les deux oligonucléotides sont possibles, le programme n'indique que celui où l'appariement est situé le plus en 3' (situation la plus défavorable).

Pour des oligonucléotides de plus de 24 bases, le Tm est calculé suivant la relation [4] :

$$T_m (\text{°C}) = 81,5 + [6,6 \times (\log \text{ concentration saline})] + [0,41 \times (\% \text{ de G} + \text{C})] - [500/(\text{longueur de l'oligonucléotide})] - [0,6 \times (\% \text{ formamide})] - [1,5 \times (\% \text{ misappariement})].$$

5' ACGGTTTCCATATGGGGATTGGTGGCGACGACTCETGGAGCCCGTCAGTATCGGCG	
3'	
5' CTAGAGCCGGATCGATCCGGTCAATCAATCAGCAAGCTTGCTGCAGGT 3'	
Max complementarity in contiguous bp = 2	
Oligonucleotide size = 19 mer	
CG% max requested = 65	
CG% min requested = 45	
5 solutions	
SOLUTION 5	
5' TGG AGC CCG TCA GTA TCG 3'	theoretical Tm = 62°C K = 188600 GC % = 63
5' GGT CAA TCA ATC AGC AAG 3'	theoretical Tm = 56°C K = 219850 GC % = 47
Max complementarity in contiguous bp = 2	
5' TGG AGC CCG TCA GTA TCG G 3'	
	* *
3' CGA ACG ACT AAC TAA CTG G 5'	
Max complementarity = 5	
5' TGG AGC CCG TCA GTA TCG G 3'	
	* * * * *
3' CGA ACG ACT AAC TAA CTG G 5'	

Figure 1a. Recherche d'amorces de chaque côté du site de clonage de **puEX1**. Affichage des données concernant l'option « Research of optimal pairs of primers » : séquences d'ADN dans lesquelles les amorces ont été recherchées, critères de sélection et solutions proposées. Seule la solution que nous avons choisie est présentée ici.

sequence : 5'- TGG AGC CCG TCA GTA TCG - 3' Secondary structure:	sequence : 5'- GGT CAA TCA ATC AGC AAG - 3' Secondary structure:
<pre> ┌ A-C-T ┐ ┌ C-G-A-G-G-T -5' │ │ │ │ │ │ G-C-C │ │ │ * * * └ T-A-T ┘ └ -3' C-G-G </pre>	<pre> ┌ T-G ┐ │ │ │ │ G │ │ * └ A-T ┘ C │ └ A-A-T-C-A-G-C-A-A-G-C -3' </pre>

Figure 1b. Affichage des données concernant l'option : Checking for self complementarity.

Pour les 14-24 mers, le programme utilise la relation simplifiée [5] : $T_m (\text{°C}) = [2 \times (A + T)] + [4 \times (G + C)]$.

2. *Checking for complementarity of two oligonucleotides* analyse le taux de complémentarité entre deux oligonucléotides, entre bases contiguës ainsi que sur toute la longueur des amorces.

3. *Checking for complementarity with a DNA sequence* recherche les homologies

entre un oligonucléotide donné et une séquence d'ADN (séquence à amplifier ou autre séquence). La longueur de la séquence n'est pas limitée.

4. *Checking for self complementarity* recherche la présence de séquences pouvant former des structures en épingle à cheveux.

5. *Degenerated oligonucleotides coding for a peptidic sequence* fournit l'oligonucléotide dégénéré correspondant à la

séquence peptidique saisie. Pour certains acides aminés, deux types de séquences dégénérées sont possibles et l'utilisateur pourra choisir celle qu'il désire.

Le programme OLIGOTEST est utilisable sur les micro-ordinateurs de type IBM PC XT, AT et compatibles équipés d'une version MS-DOS 2.0 ou supérieure. Il est disponible sur demande.

Applications

Dans l'exemple présenté (figure 1), nous avons utilisé le programme OLIGOTEST pour rechercher des amorces dans les séquences flanquant le site de clonage du plasmide pUEX1. Ces amorces ont été utilisées pour amplifier par PCR les insertions d'une banque d'ADNc soustraite, au lieu d'utiliser la technique classique des « minipreps » [6] ■

RÉFÉRENCES

1. Lowe T, Sharefkin J, Yang SQ, Diefenbach CW. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res* 1990 ; 18 : 1757-61.
2. Richlik W, Rhoads RE. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucleic Acids Res* 1989 ; 17 : 8543-50.
3. Thein SL, Wallace RB. The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. In : Davies KE, ed. *Human Genetic Diseases, a Practical Approach*. Oxford : IRL, 1986 : 33-50.
4. Anderson MLM, Young BD. Quantitative filter hybridisation. In : Hames BD, Higgins SJ, eds. *Nucleic Acid Hybridisation, a practical approach*. Oxford : IRL, 1985 : 73-111.
5. Meinkot HJ, Wahl G. Hybridization of nucleic acids immobilized on solid supports. *Ann Biochem* 1984 ; 138 : 267-84.
6. Güssow D, Clakson T. Direct clone characterization from plaques and colonies by polymerase chain reaction. *Nucleic Acids Res* 1989 ; 17 : 4000.

m/s n° 9, vol. 6, novembre 90

ALGORITHME DU PROGRAMME OLIGOTEST

Si l'on considère deux séquences A et B respectivement de taille a et b, t la taille des oligonucléotides recherchés et m le maximum de complémentarité en bases contiguës accepté, la séquence des événements de calcul sera la suivante : dans un premier temps, le programme va décomposer A et B en i et j oligonucléotides $i = (a - t) + 1$, $j = (b - t) + 1$.

Chaque séquence A_i va ensuite être testée vis-à-vis des séquences B_j . A cette étape, les séquences A_i et B_j sont fractionnées en k sous-séquences de taille m + 1. Si une complémentarité est détectée entre deux séquences A_{ik} et B_{jk} , la solution est rejetée et le programme effectue un saut à la solution $B_j + s$ où s représente la position de B_{jk} dans B_j . Cette étape permet de rejeter toutes les solutions possédant la même séquence de complémentarité et augmente grandement les performances du programme. Les solutions n'ayant pas été rejetées sont alors testées pour leur pourcentage en GC. A titre d'exemple, le temps de calcul pour deux séquences de 251 bases avec sélection d'oligonucléotides de 20 bases et un maximum de complémentarité de 2 bases est d'environ 2 min 30 s. Ce test a été réalisé sur un micro-ordinateur IBM compatible équipé d'un microprocesseur 80386 de 20 Mhz.

Remerciements

Les auteurs remercient C. de la Salle, T. Soussi et S. Amsellem d'avoir accepté de tester les différentes options du programme OLIGOTEST, et de leurs conseils avisés. Ce travail a été financé par l'Inserm, la Ligue nationale contre le cancer et l'Arc.

TIRÉS A PART

C. Junien.