

PcrBase : un logiciel d'aide à la PCR en interface avec les banques de séquences nucléiques

Nédia Kamech, Christian Fondrat, Saïd Akli, Marie-France Szajnert

La PCR (polymerase chain reaction), technique d'amplification de séquences nucléiques, nécessite le choix de deux oligonucléotides de part et d'autre de la région à amplifier. Ces oligonucléotides doivent être choisis selon certains critères nécessaires à un rendement optimal de l'expérience d'amplification. PcrBase est un logiciel conçu pour choisir un ou plusieurs couples d'oligonucléotides pour l'amplification de séquences personnelles ou issues des deux principales banques de séquences nucléiques : GenBank et l'EMBL. Avec PcrBase, l'expérimentateur a la possibilité d'engager une recherche d'oligonucléotides après avoir fixé les paramètres suivants : (1) les bornes supérieures et inférieures de la région à amplifier ainsi que la marge de recherche des oligonucléotides de part et d'autre de cette région; (2) la taille des oligonucléotides et leur pourcentage G+C; (3) la présence ou non d'une succession de dinucléotides G, C à l'extrémité 3' des oligonucléotides; (4) le nombre maximal de bases consécutives appariées dans les « épingles à cheveux » susceptibles de se former dans l'oligonucléotide; (5) le nombre maximal de bases consécutives appariées maximum dans le dimère susceptible de se former entre les deux oligonucléotides. Une fois la sélection établie, le logiciel peut tester la spécificité des oligonucléotides sur l'ensemble des séquences nucléiques de GenBank.

La *polymerase chain reaction* est une technique d'amplification de séquences nucléiques utilisée en routine dans de nombreux laboratoires de biologie (*m/s*, n° 8, vol. 4, p. 515, [1, 2]). Cependant, le choix des oligonucléotides qui vont permettre d'amplifier une séquence est délicat car il nécessite de prendre en compte plusieurs paramètres qu'il devient difficile de gérer « manuellement ». Nous avons donc conçu un logiciel d'aide au choix des oligonucléotides appelé PcrBase. Celui-ci permet la sélection de couples d'oligonucléotides sur la base de cribles successifs définis par l'utilisateur.

Il est également indispensable de tester la spécificité de ces oligonucléotides sur les séquences nucléiques stockées dans les bases de données telles que l'EMBL (*European Molecular Biology Laboratory*) ou GenBank (*Genetic Sequence Data Bank USA*). Nous avons donc interfacé ce logiciel avec le contenu de ces banques : PcrBase est implanté en mode interactif sur le serveur BISANCE [3]. Il permet ainsi de trouver les couples d'oligonucléotides pour l'amplification d'une séquence issue de la banque ou pour l'amplification d'une séquence personnelle stockée dans un fichier informatique.

L'originalité de PcrBase par rapport aux logiciels commerciaux de recherche d'oligonucléotides réside dans son exploitation interactive et

son interfaçage avec les principales banques de séquences nucléiques.

Critères retenus pour la sélection

Les bornes inférieures et supérieures de la région à amplifier sont fixées par l'utilisateur ainsi que la marge de recherche des oligonucléotides de part et d'autre de cette région (en nombres de bases), cela afin d'éviter d'inclure une région non désirée dans le produit d'amplification (région intronique par exemple) ou d'obtenir un nombre trop élevé d'oligonucléotides.

Les oligonucléotides utilisés pour l'expérience d'amplification ont le plus souvent une taille voisine de 20 nucléotides [4]. Il est possible de choisir une taille comprise entre 10 et 25 nucléotides; ce paramètre est fixe, c'est-à-dire que tous les oligonucléotides sélectionnés par le logiciel auront la même taille.

Le pourcentage G+C des oligonucléotides étant le plus souvent une contrainte majeure de l'expérience d'amplification [5], PcrBase permet de fixer les limites inférieures et supérieures de ce pourcentage dans l'intervalle compris entre 0 et 100. Un des facteurs limitants de la réaction PCR est la formation de structures secondaires et de dimères qui diminuent la concentration des oligonucléotides au site de polymérisation. Ces appariements peuvent être internes à l'oligonucléotide et for-

mer des « épingles à cheveux », ou externes et aboutir à la formation de dimères, engageant soit deux copies du même oligonucléotide (homodimères), soit deux oligonucléotides différents (hétérodimères) [4,5]. Il est possible d'éliminer les oligonucléotides susceptibles de former une épingle à cheveux avec un nombre de bases appariées successives supérieur à une valeur seuil fixée par l'expérimentateur. Ce seuil peut être choisi dans l'intervalle compris entre 0 et 3. PcrBase donne aussi la possibilité d'éliminer les oligonucléotides formant des homodimères et des hétérodimères. Pour cela l'utilisateur fixe un seuil correspondant au nombre de bases appariées successives autorisé (entre 0 et 3). Par exemple si le seuil choisi est de 3, l'oligonucléotide ci-dessous est rejeté de la sélection car il peut former un homodimère avec 4 appariements successifs :

```
5' ATTGCCTCCTACATGCTTA 3'
      ::::
3' ATTCGTACATCCTCCTGTTA 5'
```

Certains auteurs pensent que l'étape de polymérisation est facilitée lorsqu'il y a une courte succession de dinucléotides G,C à l'extrémité 3' de l'oligonucléotide ; d'autres suggèrent au contraire qu'une succession de G,C peut engager cette extrémité dans un autoappariement qui bloquerait la polymérisation [4, 6] ; le logiciel donne la possibilité de rechercher ou non des oligonucléotides présentant une succession de G,C en 3' et laisse ainsi le choix à l'utilisateur.

Enfin, pour minimiser le risque d'amplifier une séquence du même génome autre que celle désirée, en particulier dans les cas où la réaction PCR est réalisée à partir d'extrait brut, d'ARN totaux ou d'ADN [7], PcrBase teste la spécificité des oligonucléotides sur toutes les séquences de GenBank ou sur un sous-ensemble de la même espèce.

Algorithmes

Sélection des couples

Dans le logiciel nous appelons oligonucléotide droit le segment d'ADN à droite de la région à amplifier et oligonucléotide gauche le segment d'ADN à gauche de la région à amplifier (dans le sens 5'-3') (figure 1).

La recherche des couples va s'effectuer en trois étapes.

• Étape 1

L'algorithme recherche indépendamment les oligonucléotides gauches puis les oligonucléotides droits. Plusieurs filtres sont appliqués successivement sur un ensemble initial d'oligonucléotides se trouvant dans la zone de recherche autorisée par l'utilisateur (figure 2) : tout d'abord le filtre de la taille qui permettra de ne traiter que les oligonucléotides ayant la taille requise, puis le filtre du pourcentage G+C qui ne retiendra que les oligonucléotides se trouvant dans l'intervalle du pourcentage G+C désiré, puis à partir de ce nouveau sous-ensemble ne seront

retenus que les oligonucléotides ayant une succession de dinucléotides G,C à leur extrémité 3' (si cette option est demandée). Enfin seront éliminés les oligonucléotides susceptibles de former des épingles à cheveux avec un nombre d'appariements successifs supérieur au seuil fixé. Pour ce dernier filtre, l'extrémité 3' de l'oligonucléotide est déplacée vers l'extrémité 5' en décalant à chaque fois d'une base et le nombre de bases consécutives appariées dans chaque conformation est compté : si ce nombre est supérieur au seuil fixé, l'oligonucléotide est aussitôt rejeté.

A l'issue de cette étape, deux sous-ensembles distincts d'oligonucléotides droits et d'oligonucléotides gauches sont retenus et visualisés à l'écran. Pour chaque oligonucléotide sont indiqués la séquence, la position sur le brin 5'-3' , le pourcentage G+C et la température de fusion (Tm) : l'utilisateur peut alors choisir les oligonucléotides qui lui conviennent en indiquant leur numéro (figure 3 A).

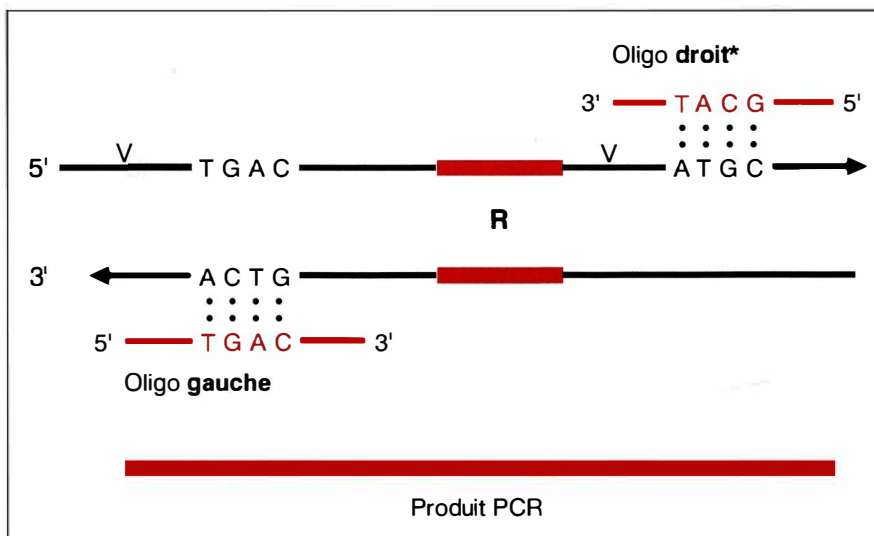


Figure 1. **Table d'orientation.** On appelle oligonucléotide droit le segment d'ADN à droite de la région à amplifier et oligonucléotide gauche le segment d'ADN à gauche de la région à amplifier R (dans le sens 5'-3'). v = positions sur le brin 5'-3' des extrémités 5' des oligonucléotides (oligo). *: inverse complémentaire.

• *Étape 2*

L'algorithme forme toutes les paires possibles d'oligonucléotides sur eux-mêmes (droit-droit ou gauche-gauche) pour éliminer ceux formant des homodimères puis teste ensuite toutes les paires droit-gauche et élimine les couples dont le nombre de bases successives appariées dépasse le seuil fixé. Pour cela il fait coulisser un oligonucléotide sur un autre en décalant d'une base à chaque fois et compte pour chacune des configurations du dimère le nombre de bases appariées. Si ce nombre est supérieur à la valeur seuil, la combinaison oligonucléotide-droit, oligonucléotide-gauche est rejetée. Les couples d'oligonucléotides retenus sont visualisés dans un tableau indiquant leur séquence, leur position sur le brin (5'-3'), leur pourcentage G+C, leur température de fusion, la taille du produit obtenu après amplification (produit PCR) ainsi que sa température de fusion (figure 3 B).

• *Étape 3*

L'ultime étape de la sélection consiste à tester dans une base de séquences nucléiques la spécificité des couples d'oligonucléotides sélectionnés. Cette recherche s'effectue avec l'oligonucléotide dans le sens direct et inverse-complémentaire.

Recherche de la spécificité des oligonucléotides dans la base de données

Afin de rechercher rapidement un motif, c'est-à-dire un court segment d'ADN de 60 caractères maximum dans une banque de séquences nucléiques, nous avons utilisé le système de codification d'une banque de séquences nucléiques mis au point au CITI 2 (Centre Interuniversitaire de Traitement de l'Information n°2) [8]. Ce système a été adapté spécifiquement pour la recherche d'oligonucléotides dans une banque.

• *Principe de la codification des séquences d'une banque nucléique*

Cette codification consiste à mémoriser dans un fichier à accès direct

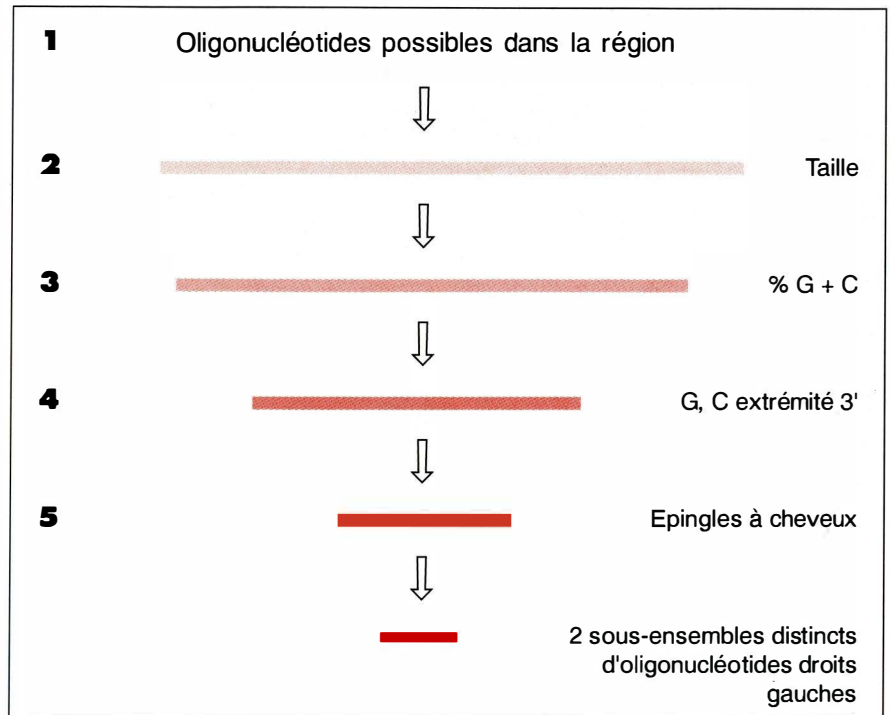


Figure 2. **Filtres successifs utilisés pour la sélection de chaque sous-ensemble d'oligonucléotides.** On détermine à l'avance la taille des oligonucléotides, leur pourcentage G+C, la possibilité d'une extrémité G,C. Sont aussi éliminés les oligonucléotides susceptibles de former des épingles à cheveux avec un nombre d'appariements successifs supérieur au seuil fixé.

Figure 3. **Recherche des oligonucléotides pour l'amplification dans la région 1020-1140 de l'exon 8 de l'hexosaminidase A (la délétion du codon 304-305 est impliquée dans la maladie de Tay-Sachs).** (A). Deux oligonucléotides gauches et deux oligonucléotides droits sont sélectionnés par le logiciel dans un premier temps. (B). Quatre couples d'oligonucléotides sont retenus. * Les oligonucléotides droits sont donnés dans le sens inverse complémentaire (la position indiquée est celle de l'extrémité 3'). (C). Amplification de la région 1039-1131 de l'exon 8 avec le couple n°1 et détection d'un mutant de 90 paires de bases correspondant à la délétion du codon 304. Piste 1 enfant normal, pistes 2-9 malades Tays-Sachs.

les noms des séquences avec la position du premier nucléotide d'un motif de longueur donnée. Une molécule d'ADN étant représentée par les 4 lettres A, C, G, T, il existe 4ⁿ possibilités de coder un motif de longueur n. Chaque séquence de la banque est alors décomposable en motifs chevauchants de même longueur n. Par exemple, une séquence de 20 bases se décomposera en 18

motifs de longueur n=3 bases (figure 4). A chaque motif correspond un code qui est une entrée possible dans ce fichier à accès direct. Ainsi pour un code C représentant un motif m, il existe un enregistrement de numéro C donnant toutes les séquences de la banque contenant ce motif ainsi que sa position dans cette séquence. Dans notre cas, la longueur d'un motif est fixée à 8

concordance d'une base est systématiquement tolérée. Par exemple pour l'oligonucléotide TATTCGACGGGC, le logiciel cherchera dans la banque toutes les séquences possédant exactement ce motif mais également toutes celles possédant le motif approché ayant une base de différente comme

TATNCGACGGGC

ou

TATTNGACGGGC

etc. (N remplaçant A,C,G ou T). Si l'utilisateur désire une recherche avec plus d'une base de différence sur l'oligonucléotide, il lui est alors conseillé d'utiliser un algorithme classique de recherche d'alignements avec une base de données [10].

Calcul des températures de fusion

La température de fusion (T_m) des oligonucléotides est calculée en utilisant les critères énergétiques des dinucléotides adjacents composant la séquence de l'oligonucléotide [11,12]. La concentration de l'oligonucléotide ainsi que la concentration en sels dans la réaction PCR sont des paramètres du calcul de ce T_m et sont donc fixés par l'utilisateur :

$$T_m = \frac{dH}{dS + [R \times \ln(C/4)]} - 273,15 + 16,6 \times \log [Na^+]$$

dH (kcal.mol⁻¹), dS (cal.mol⁻¹, 1 calorie = 4 184 Joules) sont respectivement l'enthalpie et l'entropie de la formation de l'hélice, R est la constante des gaz parfaits 1,987 (cal.k⁻¹.mol⁻¹), C la concentration de l'oligonucléotide (μM) et $[Na^+]$ la concentration en sels (mM).

La température de fusion du produit PCR (T_{mPCR}) est calculée selon la relation suivante applicable à des molécules de plus de 50 nucléotides [6] :

$$T_{mPCR} = 81,5 + 16,6 \times \log[Na^+] + 0,41 \times (GC) - 500 / \text{taille}$$

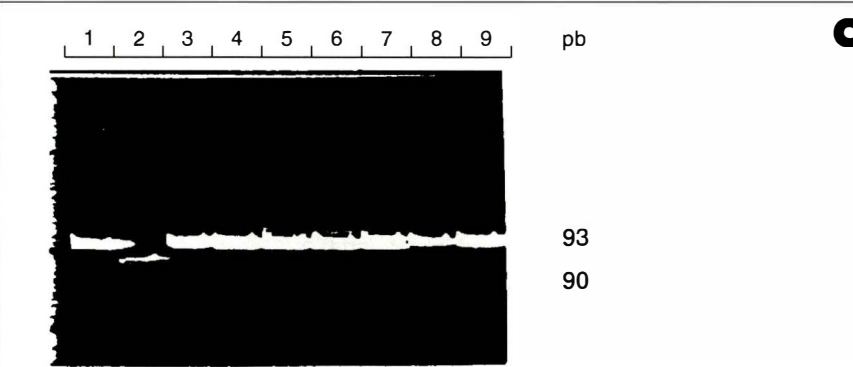
où $[Na^+]$ désigne la concentration en sels, (GC) le pourcentage G+C et « taille » la longueur du produit PCR.

Mnémonique de la séquence : hsbha (EMBL) région 1020-1140 **A**

Oligonucléotides gauches	Position	%GC	Tm
1 GAATCCCAGTCTCAATAATACCT	1038	39	58,9
2 AATCCCAGTCTCAATAATACCTA	1039	35	57,2
Oligonucléotides droits			
1 ATTTTATCTTCATCTTGAGGA	1109	30	59,4
2 TTTTATCTTCATCTTGAGGAG	1110	35	60,1

Couples retenus après le test des dimères : **B**

Oligonucléotides	Position	%GC	Tm	Taille	TmPCR
Couple n°1					
Gauche: AATCCCAGTCTCAATAATACCTA	1039	35	57,2		
Droit*: TCCTCCAAGATGAAGATAAAAAAT	1109	30	59,4	93	66,8
Couple n°2					
Gauche: AATCCCAGTCTCAATAATACCTA	1039	35	57,2		
Droit*: CTCCTCCAAGATGAAGATAAAAA	1110	35	60,1	94	67,2
Couple n°3					
Gauche: GAATCCCAGTCTCAATAATACCT	1038	39	58,9		
Droit*: TCCTCCAAGATGAAGATAAAAAAT	1109	30	59,4	94	66,8
Couple n°4					
Gauche: GAATCCCAGTCTCAATAATACCT	1038	39	58,9		
Droit*: CTCCTCCAAGATGAAGATAAAAA	1110	35	60,1	95	67,1



nucléotides [9]. Nous aurons alors 4⁸ soit 65 536 possibilités de coder un motif et notre fichier aura donc 65536 entrées. Un programme optimisé par une méthode d'intersection de listes triées peut ensuite exploiter ce fichier [9]. Ainsi des motifs de longueur inférieure ou égale à 60 caractères pourront être repérés très rapidement à l'intérieur d'une base de données de plus de

110 millions de paires de bases distribuées en quelque 100 000 séquences nucléiques. La recherche des oligonucléotides, c'est-à-dire des motifs, est ici effectuée sur toute la banque. Les résultats peuvent néanmoins être partiellement filtrés et ne présenter que les séquences d'une même espèce ou d'une extraction personnalisée de la banque. Dans cette recherche, une non-

Exemple d'une application à l'amplification de l'exon 8 du gène de l'hexosaminidase A

L'hexosaminidase A est une enzyme impliquée dans le métabolisme des gangliosides dont le déficit conduit à la maladie de Tays-Sachs [13]. Il s'agit ici d'amplifier une courte région de l'exon 8 (inférieure à 120 paires de bases) afin d'identifier rapidement une délétion du codon 304/305 par simple migration du produit PCR sur un gel de polyacrylamide [13]. On se propose d'amplifier la région 1020-1140 en utilisant PcrBase pour rechercher les oligonucléotides. La marge de recherche est fixée à 50 nucléotides en amont et en aval de la région 1070-1090. Les contraintes fixées pour les oligonucléotides sont : une taille de 23 bases, un pourcentage G+C compris entre 30 et 45, pas de dinucléotides G,C à l'extrémité 3'. Le seuil autorisé pour l'appariement dans les épingles à cheveux est de 3. Deux oligonucléotides gauches et deux oligonucléotides droits sont sélectionnés par le logiciel dans un premier temps (figure 3 A). Le seuil des appariements consécutifs autorisés pour les homodimères et des hétérodimères est fixé à 3. Quatre couples d'oligonucléotides sont alors sélectionnés par PcrBase. Leurs caractéristiques sont affichées ainsi que la taille et la température de fusion (TmPCR) du produit PCR pour chacun des couples (figure 3 B). Leur spécificité testée sur l'ensemble de GenBank ne fait pas apparaître de séquences incluant ces oligonucléotides (à part, bien sûr, la séquence de l'exon 8 de l'hexosaminidase A).

L'amplification est réalisée à partir d'ARN totaux humains selon une méthode précédemment décrite [13] en utilisant le couple n°1. La taille du produit PCR attendu est de 93 paires de bases et nous mettons effectivement en évidence sur le gel une bande de 93 paires de bases après révélation au BET (figure 3 C).

Matériel et langage

Le logiciel est programmé en Pascal et en Fortran 77 pour sa version

implantée sur le Vax du serveur BISANCE. Une version portable sur PC programmée en TurboPascal est en cours d'achèvement.

Conclusion

Il existe un certain nombre de logiciels aidant l'expérimentateur à choisir des oligonucléotides pour la réaction PCR [6, 14, 15]. Cependant PcrBase est un logiciel directement relié aux banques de séquences nucléiques. Il permet de travailler à partir d'une séquence stockée dans une banque (EMBL ou GenBank) en donnant simplement son mnémonique. Une fois les oligonucléotides sélectionnés pour la réaction PCR, PcrBase donne la possibilité de tester leur spécificité sur l'ensemble des séquences de la banque (ou sur un sous-ensemble). PcrBase étant

mis en ligne sur le serveur français BISANCE, il bénéficie de l'assurance d'avoir accès à des bases de données périodiquement mises à jour. Les développements futurs de ce logiciel donneront la possibilité de rechercher des structures secondaires sur la séquence à amplifier, en particulier lorsque l'on désire amplifier une séquence d'ARN [7, 16] ■

Remerciements

Nous remercions MM. Philippe Desse et Claude Bouchex pour leurs remarques et leurs suggestions ainsi que MM. Anders Russel et Yves Gerday pour leur aide au niveau informatique.

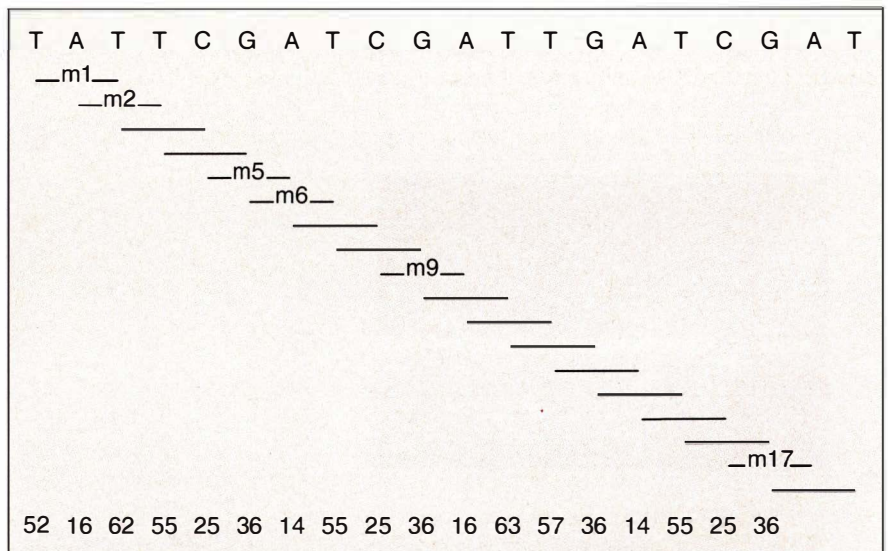


Figure 4. **Transcodage d'une molécule d'ADN-chaîne de caractères en molécule d'ADN-chaîne de motifs codés : exemple sur une séquence de 20 nucléotides.** Le code C_x d'un motif m_x commençant à la position x s'obtient de la manière suivante :

$$C_x = (P_1 \times 4^{(n-1)}) + (P_2 \times 4^{(n-2)}) + \dots + (P_i \times 4^{(n-i)}) + P_n + 1$$

où P a les valeurs 0, 1, 2, 3 pour A, C, G, T respectivement.

Pour $n = 3$, il y a 64 motifs différents (4^3) codés de 1 à 64 par $C_x = (P_1 \times 4^2) + (P_2 \times 4^1) + (P_3 + 1)$.

$$m_2 = ATT \text{ et } C_2 = (0 \times 4^2) + (3 \times 4^1) + (3 + 1) = 16$$

$$m_6 = GAT \text{ et } C_6 = (2 \times 4^2) + (0 \times 4^1) + (3 + 1) = 36$$

$$m_5 = m_9 = m_{17} = CGA \text{ et } C_5 = C_9 = C_{17} = (1 \times 4^2) + (2 \times 4^1) + 0 + 1 = 25$$

RÉFÉRENCES

- Ohta K, Nobukuni Y, Mitsubichi H, Ohta T, Tohma T, Jinno Y, Endo F, Matsuda I. Characterization of the gene encoding human pituitary-specific transcription factor. *Gene* 1992; 122: 387-8.
- Deltour L, Leduque P, Blume N, Madssen O, Dubois P, Jami J, Bucchini D. Differential expression of the 2 nonallelic proinsulin genes in the developing mouse embryo. *Proc Natl Acad Sci USA* 1993; 90: 527-31.
- Dessen P, Fondrat C, Valencien C, Mugnier C. A French service for access to biomolecular sequence databases. *Comp Applic in Biosciences* 1990; 6: 355-6.
- Acuto O, Tosi M. Amplification enzymatique de séquences nucléotidiques par PCR. Ed. INSERM, Société Française d'Immunologie, 1992.
- Taylor GR, Quirke P. Polymerase chain reaction: a practical approach. In: Melherston MJ, ed. Oxford, New York: Oxford University Press, 1991: XXI, 253 p.
- Todd I, Sharefkin J, Shi Qi Yang, Diefenbach CW. A computer program for selection of oligonucleotide primers for polymerase chain reaction. *Nucleic Acids Res* 1989; 18: 1757-61.
- Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. *Science* 1991; 252: 1643-50.
- Fondrat C, Dessen P, Le Beux P. Principle of codification for quick comparisons with the entire biomolecule databanks and associated programs in fortran 77. *Nucleic Acids Res* 1986; 14: 197-203.
- Fondrat C, Chatelus G, Dessen P. Optimization of pattern searching using a specific codification of databases. En préparation, 1993
- Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988; 85: 2444-8.
- Breslauer KJ, Frank R, Blocker H, Markey LA. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* 1986; 83: 3746-50.
- Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA* 1986; 83: 9373-7.
- Akli S, Chelly J, Lacorte JM, Poenaru L, Kahn A. Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of pcr-amplified cDNA fragments. *Genomics* 1991; 11: 124-34.
- Rychlick W, Rhoads R. A computer program for choosing optimal oligonucleotide for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucleic Acids Res* 1989; 17: 8543-50.
- Hillier L, Green P. OSP: a computer program for choosing PCR and DNA sequencing primers. *PCR Met Appl* 1991; 1: 124-8.
- Pallansch L, Beswick H, Talian J, Zelenka P. Use of an RNA folding algorithm to choose regions for amplification by the polymerase chain reaction. *Anal Biochem* 1990; 185: 57-62.

* GLOSSAIRE *

EMBL: European Molecular Biology Laboratory, banque européenne de séquences nucléiques. La version 34 de février 1993 contient près de 99 591 séquences, soit 122 millions de bases.

GenBank: Genetic sequence databank, banque américaine de séquences nucléiques. La version 75 de février 1993 contient près de 106 637 séquences (126 millions de bases).

Inverse complémentaire d'une séquence: sens direct, 5'-GCAT-3'; sens inverse complémentaire, 5'-ATGC-3'.

Mnémonique: nom sous lequel une séquence est répertoriée dans la banque.

Fichier à accès direct: structure de stockage de l'information décomposée en lignes ou enregistrements. Chacune de ces lignes est accessible par son numéro.

Nédia Kamech

Université Paris VI, GRIB (Groupe de Recherche en Imagerie Biomédicale), 45, rue des Saints-Pères, 75270 Paris Cedex 06, France.

Christian Fondrat

CITI 2 (Centre Interuniversitaire de Traitement de l'Information n°2), 45, rue des Saints-Pères, 75270 Paris Cedex 06, France.

Saïd Akli, Marie-France Szajnert

Institut Cochin de Génétique Moléculaire (ICGM), INSERM U.129, 24, rue du Faubourg Saint Jacques, 75014 Paris, France.

Summary

PcrBase: a PCR software interfaced with nucleic sequence databases

PCR, polymerase chain reaction, is a powerful technique of nucleic sequence amplification. It requires to choose two oligonucleotides at each extremity of the segment to be amplified. These oligonucleotides must be carefully chosen according to specific rules that are necessary for an optimal yield of amplification. PcrBase is a software designed to select one or more oligonucleotide couples for the amplification of personal sequences or database sequences (EMBL, GenBank). With PcrBase, the user has the possibility to search oligonucleotides according to the following criteria: (i) the upstream and downstream limits of the region to amplify as well as the range of search from both sides of this area, (ii) the oligonucleotide size and their G+C percentage, (iii) the occurrence or not of GC dinucleotide succession at the 3' end, (iv) the maximum number of consecutive matched bases susceptible to make hairpins in the oligonucleotide, (v) the maximum number of consecutive matched bases that may appear in the dimer made by the two oligonucleotides. Once the selection criteria have been established, the software will test the specificity of oligonucleotides on the whole set of sequences of GenBank.

TIRÉS A PART

N. Kamech.