

# SÉQUENÇAGE DES ACIDES NUCLÉIQUES

Coordonnateurs

L. MAILLET-BARON



T. SOUSSI

la science  
TEC  
DOC

EM  
inter





COLLECTION

## GÉNIE GÉNÉTIQUE G2

Sous la direction  
de G. Lucotte.

*Cette collection a pour objectif de fournir une panoplie complète des aspects technologiques récents utilisés dans le domaine en plein essor du génie génétique.*

*Une vingtaine de volumes sont prévus, ils ont été confiés à des spécialistes du monde de la recherche dans l'objectif de concevoir des outils pratiques présentant les nombreuses applications de cette nouvelle science, tout particulièrement dans le secteur médical.*

La technologie du séquençage nucléaire fait appel à un grand nombre de techniques qui, individuellement, sont assez bien maîtrisées dans la plupart des laboratoires. Cependant, l'utilisateur, confronté au choix des techniques disponibles, manque souvent de clés pour bâtir une **stratégie de séquençage efficace**.

Cet ouvrage, écrit par une équipe de chercheurs et d'enseignants, propose une approche véritablement novatrice qui insiste sur **la compréhension des techniques**. Le lecteur y puisera les données permettant de définir la démarche la plus adaptée à son objectif (recherche de mutations, vérification d'un clonage, séquençage *de novo*...).

Ouvrage de réflexion, il offre également toutes les qualités d'un **manuel pratique** tel que le préconise la collection Génie Génétique. Ainsi, il dresse un panorama complet des techniques et problèmes concernant le séquençage, propose des solutions pour contourner les problèmes d'interprétation liés aux artefacts, décrit le fonctionnement des séquenceurs et expose les méthodes permettant le traitement informatique des données recueillies. De nombreuses illustrations, schémas et planches couleurs contribuent à la clarté de l'exposé.

Conçu dans un langage simple et didactique, cet ouvrage sera indispensable aux laboratoires de recherche et d'analyse en biologie, biotechnologie, pharmacologie et médecine. Il sera également un outil de formation très efficace pour les enseignants, les étudiants de deuxième cycle et les jeunes chercheurs désirant une information concrète sur les avancées actuelles de la biologie moléculaire.

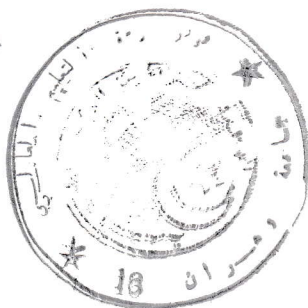
2-85206-855-9



9 782852 068855



# TABLE DES MATIÈRES



## INTRODUCTION

|                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------|----|
| <b>1 L'HÉRÉDITÉ</b>                                                                                                         | 19 |
| <b>2 LES ACIDES NUCLÉIQUES,<br/>SUPPORTS DE L'INFORMATION GÉNÉTIQUE</b>                                                     | 20 |
| <b>3 LE SÉQUENÇAGE<br/>DES ACIDES NUCLÉIQUES</b>                                                                            | 25 |
| 3.1 Les techniques de séquençage chimique                                                                                   | 26 |
| 3.2 La première technique enzymatique de séquençage de l'ADN :<br>la méthode « plus et moins » de Sanger                    | 27 |
| 3.3 Le séquençage enzymatique par incorporation de didésoxy-<br>nucléotides terminateurs de chaîne : la technique de Sanger | 30 |
| 3.4 Une grande contribution : les vecteurs M13mp de Messing                                                                 | 33 |
| 3.5 Les autres développements technologiques                                                                                | 33 |

|                                                                      |    |
|----------------------------------------------------------------------|----|
| <b>4 INTÉRÊTS ET LIMITES DU SÉQUENÇAGE<br/>DES ACIDES NUCLÉIQUES</b> | 35 |
|----------------------------------------------------------------------|----|

## CHAPITRE 1

# LES OUTILS DU SÉQUENÇAGE ENZYMATIQUE DE L'ADN

|                                              |    |
|----------------------------------------------|----|
| <b>1 LES VECTEURS</b>                        | 41 |
| 1.1 Première génération - Le plasmide pBR322 | 41 |
| 1.2 Le développement des vecteurs M13        | 44 |
| 1.3 Les vecteurs pUC                         | 50 |
| 1.4 Les phagemides                           | 52 |
| <b>2 L'ISOTOPE UTILISÉ</b>                   | 55 |
| <b>3 LES DIFFÉRENTES ADN POLYMÉRASES</b>     | 57 |
| 3.1 L'ADN polymérase de Klenow               | 58 |
| 3.2 La reverse transcriptase aviaire         | 59 |
| 3.3 L'ADN polymérase du bactériophage T7     | 59 |
| 3.4 Les ADN polymérases thermostables        | 59 |
| <b>4 LES TECHNIQUES DE MARQUAGE</b>          | 61 |
| <b>5 LE SÉQUENÇAGE D'ADN DOUBLE BRIN</b>     | 65 |
| <b>6 LE CHOIX DES AMORCES</b>                | 66 |
| <b>7 LES TECHNIQUES ÉLECTROPHORÉTIQUES</b>   | 67 |
| <b>8 LES ANALOGUES STRUCTURAUX</b>           | 69 |

## PROBLÈMES RENCONTRÉS : ORIGINE ET RÉOLUTION

- 1 BRUIT DE FOND SUR L'AUTORADIOGRAPHIE** - 75
- 2 MARQUAGE DE FAIBLE INTENSITÉ  
TOUT AU LONG DU GEL** \_\_\_\_\_ 76
- 3 DOUBLES BANDES ET BANDES ARTÉFACTUELLES  
D'INTENSITÉS CORRECTES** \_\_\_\_\_ 77
- 4 DOUBLES BANDES ET BANDES ARTÉFACTUELLES  
DE FAIBLES INTENSITÉS** \_\_\_\_\_ 78
- 5 COMPRESSIONS**
  - 5.1 Conditions d'électrophorèse..... 79
  - 5.2 Conditions de réaction ..... 80
- 6 PAUSES OU BANDES PARALLÈLES  
DANS LES 4 PISTES** \_\_\_\_\_ 83
- 7 SÉQUENÇAGE DE RÉGIONS  
HOMOPOLYMÉRIQUES OU DE FRAGMENTS  
CONTENANT UN DÉSÉQUILIBRE  
NUCLÉOTIDIQUE** \_\_\_\_\_ 85

## STRATÉGIES DE SÉQUENÇAGE

- 1 NÉCESSITÉ D'UNE STRATÉGIE** \_\_\_\_\_ 89
- 2 MÉTHODES DIRIGÉES** \_\_\_\_\_ 90

|          |                                                                               |     |
|----------|-------------------------------------------------------------------------------|-----|
| 2.1      | Utilisation de la carte de restriction .....                                  | 90  |
| 2.2      | Méthodes impliquant des délétions ordonnées.....                              | 92  |
| 2.3      | Utilisation des oligonucléotides de synthèse.....                             | 95  |
| <b>3</b> | <b>MÉTHODES NON ORDONNÉES<br/>OU ALÉATOIRES</b> .....                         | 97  |
| <b>4</b> | <b>NOUVELLES APPROCHES</b> .....                                              | 99  |
| 4.1      | Délétions progressives induites <i>in vivo</i> par un transposon.....         | 99  |
| 4.2      | Méthode Exometh™.....                                                         | 100 |
| 4.3      | Méthode Multiplex™.....                                                       | 103 |
| 4.4      | Séquençage et carte à puce .....                                              | 106 |
| 4.5      | Séquençage des génomes d'organismes présentant un intérêt<br>biologique ..... | 107 |

## CHAPITRE 4

# PCR ET SÉQUENÇAGE

|          |                                                                                                                              |     |
|----------|------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>1</b> | <b>LA TECHNIQUE DE PCR</b> .....                                                                                             | 113 |
| <b>2</b> | <b>LES ADN POLYMÉRASES THERMOSTABLES</b> .....                                                                               | 116 |
| <b>3</b> | <b>LE SÉQUENÇAGE D'UN PRODUIT DE PCR</b> .....                                                                               | 117 |
| 3.1      | Le séquençage direct d'un produit de PCR.....                                                                                | 117 |
| 3.2      | Le séquençage sur support solide.....                                                                                        | 118 |
| 3.3      | Séquençage après sous clonage.....                                                                                           | 119 |
| 3.4      | Amplification et séquençage de génome diploïde ou séquençage<br>de produits de PCR à partir d'échantillons hétérogènes ..... | 120 |
| <b>4</b> | <b>SÉQUENÇAGE PAR AMPLIFICATION LINÉAIRE</b> .....                                                                           | 122 |
| <b>5</b> | <b>CONCLUSION</b> .....                                                                                                      | 123 |



# LE SÉQUENÇAGE AUTOMATISÉ

|                                                                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>1 INTRODUCTION</b>                                                                                                                                            | 127 |
| <b>2 LE MARQUAGE DES ACIDES NUCLÉIQUES<br/>PAR DES MOLÉCULES FLUORESCENTES</b>                                                                                   | 128 |
| <b>3 LA GÉNÉRATION DES FRAGMENTS MARQUÉS</b>                                                                                                                     | 130 |
| 3.1 Génération de fragments de clivage.....                                                                                                                      | 130 |
| 3.1.1 Obention des fragments • 3.1.2 Marquage des fragments                                                                                                      |     |
| 3.2 Synthèse de fragments complémentaires .....                                                                                                                  | 131 |
| 3.2.1. Synthèse des fragments • 3.2.2 Marquage des fragments: a) Le marquage<br>de l'amorce; b) le marquage des extrémités 3'; c) l'utilisation de dXTP marqués. |     |
| <b>4 LA SÉPARATION ET LA DÉTECTION<br/>DES FRAGMENTS</b>                                                                                                         | 134 |
| <b>5 LE MODÈLE 373A D'APPLIED BIOSYSTEMS</b>                                                                                                                     | 137 |
| <b>6 L'A.L.F. DNA SEQUENCER DE PHARMACIA-LKB</b>                                                                                                                 | 139 |
| <b>7 PERSPECTIVES D'AVENIR</b>                                                                                                                                   | 140 |

# SYNTHÈSE D'OLIGONUCLÉOTIDES

|                                               |     |
|-----------------------------------------------|-----|
| <b>1 HISTORIQUE ET INTRODUCTION</b>           | 147 |
| <b>2 FONCTIONNALISATION DU SUPPORT</b>        | 148 |
| <b>3 ASSEMBLAGE DE CHAINES NUCLÉOTIDIQUES</b> | 149 |
| <b>4 CHOIX DE LA MÉTHODE</b>                  | 150 |

|           |                                                                                          |     |
|-----------|------------------------------------------------------------------------------------------|-----|
| <b>5</b>  | <b>MÉTHODES CHIMIQUES DE SYNTHÈSE</b>                                                    | 151 |
| 5.1       | Chimie des phosphotriesters .....                                                        | 151 |
| 5.2       | Chimie des phosphites.....                                                               | 152 |
| 5.3       | Chimie des phosphonates .....                                                            | 154 |
| <b>6</b>  | <b>PURIFICATION</b>                                                                      | 156 |
| 6.1       | Techniques chromatographiques.....                                                       | 156 |
| 6.1.1     | <i>Chromatographie d'échanges d'ions</i> • 6.1.2 <i>Chromatographie en phase reverse</i> |     |
| 6.2       | Technique d'électrophorèse .....                                                         | 158 |
| <b>7</b>  | <b>SÉQUENÇAGE DE L'OLIGONUCLÉOTIDE</b>                                                   | 159 |
| 7.1       | Technique du Wandering Spot.....                                                         | 159 |
| 7.2       | Technique de Maxam-Gilbert .....                                                         | 160 |
| <b>8</b>  | <b>SYNTHÈSES « EXOTIQUES »<br/>NOUVELLE GÉNÉRATION</b>                                   | 161 |
| <b>9</b>  | <b>SYNTHÉTISEURS D'ADN</b>                                                               | 162 |
| <b>10</b> | <b>EN CONCLUSION</b>                                                                     | 163 |

## CHAPITRE 7

# LE SÉQUENÇAGE DES ARN

|          |                                                                                                                  |     |
|----------|------------------------------------------------------------------------------------------------------------------|-----|
| <b>1</b> | <b>INTRODUCTION</b>                                                                                              | 167 |
| <b>2</b> | <b>SÉQUENÇAGE D'ARN† NON RADIOACTIF<br/>PAR HYDROLYSES ENZYMATIQUES</b>                                          | 177 |
| <b>3</b> | <b>SÉQUENÇAGE PAR HYDROLYSES<br/>ENZYMATIQUES SPÉCIFIQUES D'ARN RADIO-<br/>ACTIF UNIFORMÉMENT MARQUÉ IN VIVO</b> | 188 |



## **4 SÉQUENÇAGE D'ARN APRÈS POSTMARQUAGE IN VITRO** \_\_\_\_\_ 191

4.1 Séquençage par coupures enzymatiques spécifiques de l'ARN  
et postmarquage des oligonucléotides..... 192

4.2 Séquençage direct d'ARN postmarqué à une extrémité..... 196

4.2.1 *Séquençage d'ARN postmarqué par coupures enzymatiques spécifiques*

4.2.2 *Séquençage d'ARN postmarqué par coupures chimiques spécifiques*

4.3 Séquençage d'ARN par coupure au hasard suivie de  
postmarquage..... 199

4.4 Séquençage d'ARN par extension/terminaison de chaînes  
(méthode dite de Sanger modifiée) ..... 206

## **5 CONCLUSION** \_\_\_\_\_ 209

## **6 HISTORIQUE** \_\_\_\_\_ 210

### **CHAPITRE 8**

## **SÉQUENÇAGE : ASPECTS INFORMATIQUES**

### **1 INTRODUCTION DÉFINITION DU PROBLÈME** \_\_\_\_\_ 213

### **2 LES MÉTHODES INFORMATIQUES D'ASSEMBLAGE DE FRAGMENTS D'ADN** \_\_\_\_\_ 215

2.1 La méthode de Staden..... 215

2.2 Autres méthodes..... 215

### **3 EXEMPLES D'IMPLANTATION** \_\_\_\_\_ 217

3.1 Logiciel SAP : méthode automatique ..... 219

3.2 Logiciel SAP : méthode manuelle..... 219

3.3 Commentaires ..... 220

|          |                                                                                                |            |
|----------|------------------------------------------------------------------------------------------------|------------|
| 3.4      | Logiciel GCG.....                                                                              | 220        |
| 3.5      | Logiciels MICRO .....                                                                          | 222        |
| <b>4</b> | <b>ENTRÉE MANUELLE DES DONNÉES<br/>ET SÉQUENÇAGE AUTOMATIQUE</b> .....                         | <b>223</b> |
| <b>5</b> | <b>VALIDATION ET ANALYSE<br/>DES SÉQUENCES OBTENUES</b> .....                                  | <b>224</b> |
| 5.1      | Evaluation des erreurs de séquençage .....                                                     | 224        |
| 5.2      | Recherche des phases codantes.....                                                             | 225        |
| 5.3      | Analyse des séquences génomiques Eucaryotes .....                                              | 226        |
| 5.4      | Similitudes avec les banques de données.....                                                   | 227        |
| <b>6</b> | <b>UTILISATION DE RESSOURCES<br/>DE CALCUL ACCESSIBLES<br/>SUR RÉSEAUX INFORMATIQUES</b> ..... | <b>229</b> |
| 6.1      | Centres serveurs.....                                                                          | 229        |
| 6.2      | Ressources spécifiques.....                                                                    | 229        |
| <b>7</b> | <b>PERSPECTIVES</b> .....                                                                      | <b>232</b> |
|          | <b>CONCLUSION</b> .....                                                                        | <b>236</b> |
|          | <b>INDEX</b> .....                                                                             | <b>238</b> |

PCR<sup>TM</sup> est une marque déposée de Cetus corp., propriété actuelle de la société Hoffmann-La Roche.

Sequenase<sup>TM</sup> est une marque déposée, propriété de la United States Biochemicals Corporation.

Multiplex<sup>TM</sup> et Plex sont des marques déposées, propriété de Millipore Corporation.

Exometh<sup>TM</sup> est une marque déposée, propriété de la société Stratagene.

A.L.F.<sup>TM</sup> DNA sequencer est une marque déposée Pharmacia.

Klenow est une abréviation pour : fragment de Klenow de l'ADN polymérase I d'*E. coli*

Les cartes de plasmides ont été fournies par la société Pharmacia P.L. Biochemicals.