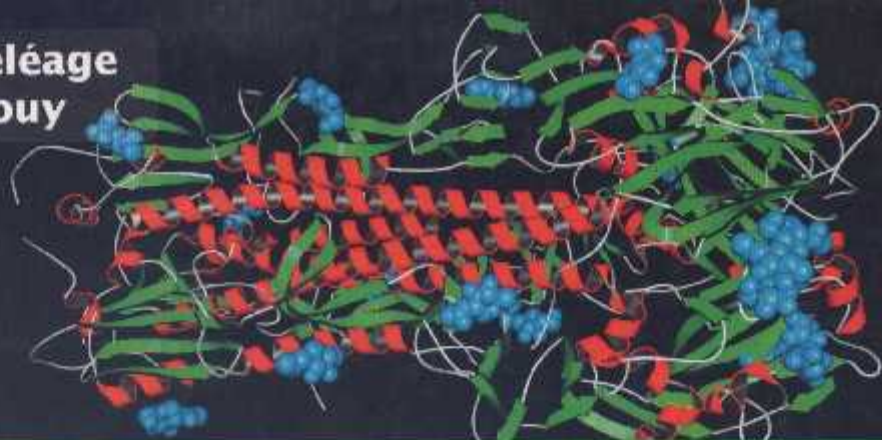


**Gilbert Deléage  
Manolo Gouy**



# Bioinformatique

Cours et cas pratique

Avec ce livre, des  
bonus sur le web



Licence 3  
Master  
Écoles d'ingénieurs

**DUNOD**

Gilbert Deléage  
Manolo Gouy

BL 555

059144

(2)

# Bioinformatique

Cours et cas pratique

DUNOD

# TABLE DES MATIÈRES

<b>Comment utiliser cet ouvrage</b>	VI
<b>Avant-propos</b>	IX
<b>Chapitre 1 • La composition en acides aminés</b>	1
1.1 Acides aminés et séquence	1
1.2 Informations déduites de la composition en acides aminés	4
<b>Chapitre 2 • Bases de données pour données de bases</b>	7
2.1 Les banques de données généralistes	7
2.2 Une entrée SWISS-PROT	14
2.3 Les interrogations Entrez, ACNUC, SRS	17
<b>Chapitre 3 • La comparaison de deux séquences</b>	21
3.1 Matrice de points	21
3.2 Matrice de substitution	26
<b>Chapitre 4 • Recherche dans les banques</b>	33
4.1 Score de similitude entre séquences	33
4.2 Recherche globale ou locale	36
4.3 FASTA	37
4.4 BLAST	41
<b>Chapitre 5 • Alignement de séquences</b>	47
5.1 Introduction	47
5.2 Comparaison de protéines homologues (algorithme global)	49
5.3 Meilleur chevauchement entre séquences (algorithme local)	52
5.4 Alignements multiples	54
5.5 Représentation « logo »	57
<b>Chapitre 6 • Bases théoriques de la phylogénie moléculaire</b>	59
6.1 Arbres phylogénétiques	59
6.1.1 Arbres racinés et arbres non racinés	61
6.1.2 Le format Newick d'arbres phylogénétiques	62
6.2 Arbre des espèces - arbres de gènes	63
6.2.1 Nombre d'arbres binaires possibles	64
6.3 Modèle markovien de l'évolution moléculaire	65

## Table des matières

6.3.1	Matrice de transition	67
6.3.2	Quelques modèles nucléotidiques de Markov	68
6.3.3	Longueur d'une branche	70
6.3.4	Modélisation de la variation des taux d'évolution entre sites	70
6.4	Choix des sites	72
6.5	Matrices de taux de substitution entre séquences protéiques	73
6.6	Distances évolutives entre paires de séquences	74
<b>Chapitre 7 • Algorithmes pour la phylogénie moléculaire</b>		<b>77</b>
7.1	Parcimonie	78
7.1.1	Algorithme	78
7.1.2	Heuristiques	82
7.1.3	Propriétés	82
7.1.4	Implémentations	83
7.1.5	Longueurs de branches des arbres de parcimonie	83
7.1.6	Traitement des indels	84
7.2	Méthodes de distances	85
7.2.1	Méthode d'évolution minimale	86
7.2.2	Méthode Neighbor-Joining	86
7.3	Maximum de vraisemblance	89
7.3.1	Modèle probabiliste utilisé au maximum de vraisemblance	90
7.3.2	Calcul de la vraisemblance	90
7.3.3	L'algorithme de Felsenstein	91
7.3.4	Prise en compte de la variabilité des vitesses d'évolution entre sites	92
7.3.5	Optimisation de la vraisemblance	93
7.3.6	Implémentation	94
7.4	Estimation de la fiabilité d'un arbre par bootstrap	94
7.5	Choix des méthodes de calcul d'arbres	97
<b>Chapitre 8 • Recherche de fonctions</b>		<b>99</b>
8.1	Définitions	99
8.2	Détection de signatures de séquence (PROSITE)	100
8.3	Recherche de fonction avec pondération par la fréquence	103
8.4	Méthodes à base de profils	106
<b>Chapitre 9 • Profils physico-chimiques</b>		<b>111</b>
9.1	Pourquoi les profils physico-chimiques ?	111
9.2	Hydrophobie-paramètres-construction du profil - interprétation	111
9.3	Amphiphilie	114
9.4	Accessibilité au solvant	115
<b>Chapitre 10 • Prédiction de structures secondaires</b>		<b>117</b>
10.1	Méthode « statistique empirique »	120
10.2	Méthode information directionnelle (GOR)	123
10.3	Méthode de recherche des plus proches voisins (NNM)	127

10.4 Méthode auto-optimisée (SOPM)	130
10.5 Méthode auto-optimisée avec alignements (SOPMA)	131
10.6 Méthodes neuronales	132
10.7 Autres méthodes	134
10.7.1 Méthode statistique discriminante (DSC)	134
10.7.2 Méthode neuronale (PREDATOR)	134
10.7.3 Méthode hiérarchisée réseaux de neurones (HNN)	134
10.7.4 Méthodes utilisant les chaînes de Markov	135
10.7.5 Combinaison de méthodes	135
10.8 Critères de qualité prédictive	137
<b>Chapitre 11 • Prédiction de structures 3D</b>	<b>139</b>
11.1 Principe des méthodes de détermination expérimentale	139
11.2 Le format PDB	140
11.3 Les différents modes de représentations	142
11.4 Classification de structures 3D	145
11.5 Comparaison de structures 3D	146
11.6 Énergétique moléculaire	148
11.7 Optimisation de structures 3D	151
11.8 Modélisation de structures 3D	152
11.8.1 Les méthodes d'enfilage des repliements ( <i>threading</i> )	153
11.8.2 Modélisation par homologie	154
11.8.3 Les alphabets structuraux	163
11.8.4 Les méthodes de novo	166
<b>Chapitre 12 • Détection de sites 3D dans les protéines</b>	<b>169</b>
12.1 Problématique	169
12.2 Méthode SuMO	170
<b>Cas pratique d'analyse de séquences</b>	<b>173</b>
<b>Conclusion</b>	<b>183</b>
<b>Glossaire</b>	<b>185</b>
<b>Bibliographie</b>	<b>187</b>
<b>Index</b>	<b>193</b>

Gilbert Deléage  
Manolo Gouy

## Bioinformatique

La bioinformatique a pour but d'intégrer des données d'origines très diverses pour modéliser les systèmes vivants afin de comprendre et prédire leurs comportements (analyse du génome, modélisation de l'évolution d'une population animale, modélisation moléculaire, reconstruction d'arbres phylogénétiques...).

Ce livre est axé sur la **bioinformatique des protéines**. Il aborde de manière simple les tâches courantes qu'un biologiste ou un biochimiste doit savoir traiter sans avoir recours au spécialiste. Conçu de manière à faciliter la compréhension des approches, méthodes, algorithmes et implémentations les plus courants, il leur permettra d'éviter les pièges classiques et de répondre aux questions usuelles :

- Comment extraire des informations pertinentes dans les banques de données biologiques ?
- Deux séquences correspondent-elles à deux gènes homologues ?
- Quelle peut être la fonction d'une protéine ?
- Comment construire un modèle 3D de protéine ?...
- Comment construire un arbre phylogénétique à partir d'un ensemble de séquences protéiques ?

En fin d'ouvrage, un **cas pratique détaillé** permet de mettre directement en application le cours.

### Public :

- Étudiants en Licence 3 ou Master de biologie ou de biochimie
- Étudiants en IUT de génie biologique
- Élèves ingénieurs



9 782100 587513

6973184  
ISBN 978-2-10-058751-3



- MATHÉMATIQUES
- PHYSIQUE
- CHIMIE
- SCIENCES DE L'INGÉNIEUR
- INFORMATIQUE
- SCIENCES DE LA VIE
- SCIENCES DE LA TERRE

### Gilbert Deléage

est professeur de biochimie à l'université Claude Bernard Lyon 1 et directeur de l'Institut de Biologie et de Chimie des Protéines à Lyon.

### Manolo Gouy

est directeur de recherche au laboratoire de Biométrie et Biologie évolutive (CNRS, université Claude Bernard Lyon 1).