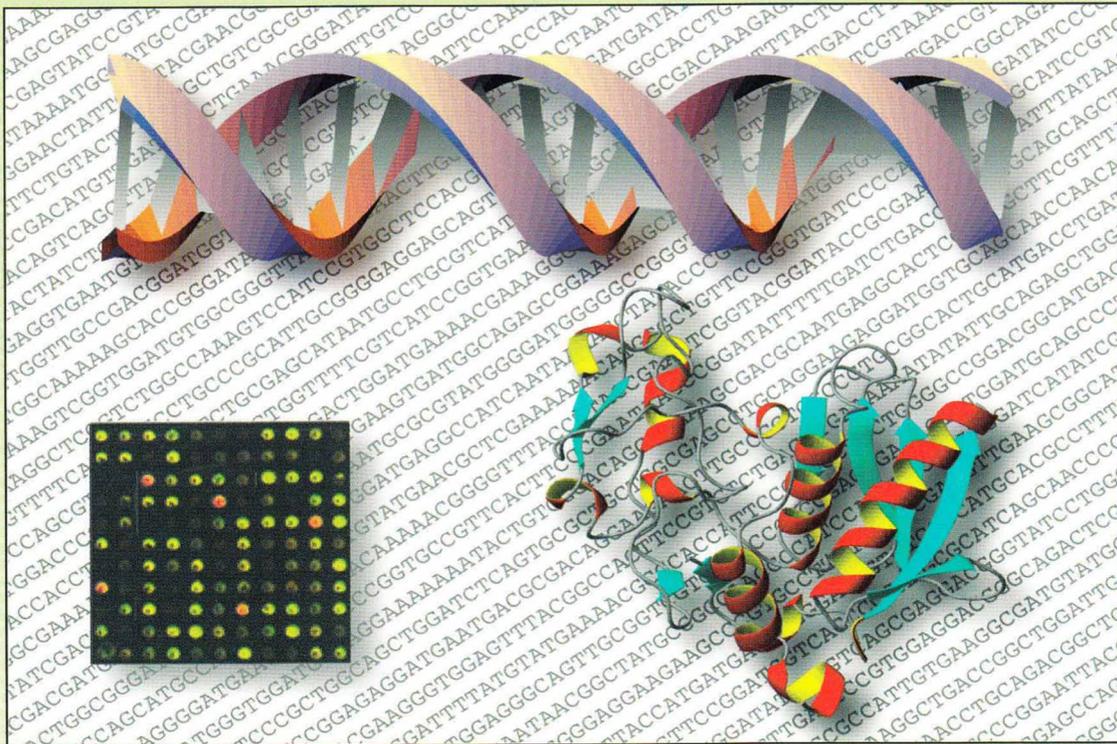


Frédéric Dardel & François Képès

---

# Bioinformatique

## Génomique et post-génomique



# Table des matières

Introduction : l'ère de la génomique	7
<b>1 Séquençage et Génome</b>	<b>9</b>
1.1 Le séquençage automatique . . . . .	9
1.2 Stratégies de séquençage . . . . .	12
1.3 Stratégies de fragmentation . . . . .	15
1.4 Assemblage de séquence . . . . .	20
1.5 Comblement des « trous » . . . . .	23
1.6 Obstacles à la reconstruction . . . . .	25
1.7 Utilisation d'une banque complémentaire de « grands » clones	26
1.8 Le génome de <i>Haemophilus influenzae</i> . . . . .	28
1.9 ADNc et EST . . . . .	29
<b>2 Comparaisons de séquences</b>	<b>35</b>
2.1 Introduction : la comparaison comme méthode de prédiction . .	35
2.2 Notre molécule test : Le récepteur humain de l'androstérone . .	36
2.3 Homologies de séquence – homologies fonctionnelles . . . . .	36
2.4 Matrices de comparaison . . . . .	39,
2.5 Le problème des insertions et délétions . . . . .	44
2.6 Alignement optimal : la méthode par programmation dynamique	45
2.7 Les méthodes heuristiques rapides . . . . .	51
2.8 Sensibilité, sélectivité et niveau de confiance . . . . .	59
2.9 Alignements multiples . . . . .	63
2.10 Domaines et motifs associés . . . . .	71
<b>3 Génomique comparative</b>	<b>75</b>
3.1 Propriétés d'ensemble du génome . . . . .	75
3.1.1 Taille et structure des génomes eucaryotes . . . . .	76

3.1.2	Diversité et plasticité de la structure des génomes bactériens . . . . .	77
3.1.3	Biais, isochores, îlots CpG . . . . .	80
3.2	Comparaison de génomes . . . . .	82
3.2.1	Gènes orthologues et gènes paralogues . . . . .	82
3.2.2	Synténie . . . . .	85
3.2.3	Jeu minimal de gènes . . . . .	87
3.2.4	Îlots de pathogénicité . . . . .	89
3.2.5	Cibles thérapeutiques . . . . .	90
3.3	Évolution des génomes et phylogénie ; applications à l'annotation	90
3.3.1	Évolution des génomes . . . . .	90
3.3.2	Prédiction de fonctions par le contexte génomique . . . . .	95
3.3.3	L'arbre génomique de la vie . . . . .	97
<b>4</b>	<b>Information génétique et Séquences biologiques</b>	<b>101</b>
4.1	Introduction : différents niveaux de codage . . . . .	101
4.2	Gènes - Code génétique . . . . .	102
4.3	Signaux d'expression . . . . .	103
4.4	Sites spécifiques . . . . .	108
4.5	Sites présents sur l'ADN . . . . .	108
4.6	Sites présents sur l'ARN . . . . .	113
4.7	Méthodes de recherche de motif . . . . .	115
<b>5</b>	<b>Statistiques et séquences</b>	<b>127</b>
5.1	Introduction . . . . .	127
5.2	Distribution des bases et des acides aminés . . . . .	127
5.3	Bases biologiques des biais statistiques . . . . .	133
5.4	Utilisation prédictive des biais statistiques . . . . .	135
5.5	Modéliser les séquences d'ADN . . . . .	138
5.6	Modèles complexes . . . . .	142
5.7	Erreurs de séquençage, Modèles de Markov cachés . . . . .	146
5.8	Les processus de Markov cachés : un outil général d'analyse des séquences . . . . .	150
5.9	La quête des gènes, un art difficile . . . . .	151
<b>6</b>	<b>Prédictions de structure</b>	<b>153</b>
6.1	Structure de l'ARN . . . . .	153
6.2	Propriétés de la molécule d'ARN . . . . .	155
6.3	Structures secondaires de l'ARN . . . . .	157

6.4	Stabilité thermodynamique des structures d'ARN . . . . .	161
6.5	Recherche de la structure la plus stable . . . . .	168
6.6	Validation des structures secondaires prédites . . . . .	173
6.7	Interactions à longue distance et prédiction de structure 3D . .	177
6.8	Structure des protéines . . . . .	181
6.9	Prédictions de structure secondaire . . . . .	183
6.10	Modélisation 3D à partir de la structure d'une protéine homologue	187
6.11	Prédiction de repliement . . . . .	193
<b>7</b>	<b>Transcriptome et Protéome. Réseaux macromoléculaires</b>	<b>195</b>
7.1	Introduction . . . . .	195
7.2	Méthodes de la post-génomique . . . . .	196
7.2.1	Protéomique . . . . .	196
7.2.2	Transcriptomique . . . . .	203
7.3	Réseaux macromoléculaires . . . . .	210
7.3.1	Interactions entre protéines . . . . .	211
7.3.2	Interactions entre enzymes et substrats . . . . .	211
7.3.3	Interactions entre protéines régulatrices et régions régula- toires de l'ADN . . . . .	219
7.4	Topologie des réseaux macromoléculaires . . . . .	221
7.4.1	Analyse de la topologie . . . . .	223
7.4.2	Interactome . . . . .	226
7.4.3	Métabolome . . . . .	227
7.4.4	Réseau génétique . . . . .	228
7.5	Transcriptomique et inférence de réseau génétique . . . . .	230
7.5.1	Les données . . . . .	230
7.5.2	Le modèle booléen . . . . .	231
7.5.3	Les autres modèles . . . . .	231
<b>8</b>	<b>Simulation de processus biologiques en génomique</b>	<b>235</b>
8.1	Génomique et simulation . . . . .	236
8.2	Prédiction et explication . . . . .	239
8.3	État de l'art . . . . .	239
8.4	Problèmes et prospective . . . . .	241
8.4.1	Objectifs . . . . .	241
8.4.2	Pauvreté scientifique . . . . .	241
8.4.3	Obstacles ontogéniques . . . . .	242
8.4.4	Obstacles épistémologiques . . . . .	244