



impression PDF

## La bioinformatique

1. La bioinformatique : définition, description, démarche et principales étapes
2. Deux types de molécules support de la bioinformatique : les acides nucléiques et les protéines
3. Deux types de bioinformatique : la séquence des nucléotides et la séquence des acides aminés
4. L'obtention des séquences : méthodes de F. Sanger (acide nucléique) et P. Edman (protéine)
5. Le séquençage du génome humain
6. Le stockage de la bioinformatique : les banques de données
7. La structuration de la bioinformatique : fichiers et formats
8. Exemples d'algorithmes et de programmes en bioinformatique
9. Exemple d'analyse bioinformatique d'une enzyme : la glutamate déshydrogénase (GDH)
10. Les domaines d'étude en "ome" ou "omique"
11. Société : champs d'application de la bioinformatique
12. Liens Internet et références bibliographiques

Propositions d'emplois en bioinformatique

### 1. La bioinformatique : définition, description, démarche et principales étapes

#### Définition

La **bioinformatique** est l'information liée aux **molécules biologiques** : leurs structures, leurs fonctions, leurs liens de "parenté", leurs interactions et leur intégration dans la cellule.

Divers domaines d'études permettent d'**obtenir cette bioinformatique** : la génomique structurale, la génomique fonctionnelle, la protéomique, la détermination de la structure spatiale des molécules biologiques, la modélisation moléculaire ...

La **bioinformatique** est l'**analyse de la bioinformatique**.

#### Description

C'est une discipline **récente** (quelques dizaines d'années).

C'est une discipline "**hybride**" (au même titre que la biochimie ou la **biophysique**) : elle est fondée sur des concepts et des formalismes issus de la biologie, de l'informatique, des mathématiques et de la physique.

C'est une discipline qui **utilise** toutes les potentialités de traitement de l'**informatique** : modèles théoriques, algorithmes et programmes, ordinateurs, réseau Internet, bases de données ...

#### Démarche

1. **Compilation et organisation des données biologiques dans des banques de données** : ces banques sont soit généralistes (elles contiennent le plus d'information possible sans expertise particulière de l'information déposée), soit spécialisées dans un domaine autour de thèmes précis.

2. **Traitements systématiques des données** : l'objectif principal est de repérer et de caractériser une fonction et/ou une structure biologique importante. Les résultats de ces traitements constituent de **nouvelles données biologiques** obtenues "*in silico*".

3. **Elaboration de stratégies** :

- le but est d'apporter des connaissances biologiques supplémentaires en combinant les données biologiques initiales et les données biologiques obtenues "*in silico*".
- ces connaissances permettent, à leur tour, de développer de **nouveaux concepts en biologie**.
- concepts qui nécessitent l'élaboration de nouvelles théories et outils en mathématiques et en informatique.

#### Principales étapes

## 2. Deux types de molécules support de la bioinformation : les acides nucléiques et les protéines

### ADN : Acide DésoxyriboNucléique

- macromolécule : chaîne nucléotidique
- constituée par un enchaînement d'unités élémentaires : les **désoxyribonucléotides**
- forme de stockage de l'information génétique. Cette information est représentée par une suite linéaire de gènes
- formée de deux brins complémentaires enroulés en double hélice ce qui lui permet de se dupliquer en deux molécules identiques entre elles et identiques à la molécule mère

#### On distingue :

- l'ADN du génome du noyau
- l'ADN du génome mitochondrial
- l'ADN du génome chloroplastique

### ARN : Acide RiboNucléique

- macromolécule : chaîne nucléotidique
- constitué par un enchaînement d'unités élémentaires : les **ribonucléotides**
- forme qui permet de transférer et de traiter l'information dans la cellule
- le plus souvent formé d'un simple brin

#### On distingue :

- les ARN messagers ou ARNm : ils sont transcrits à partir d'un gène (ADN). Ils sont ensuite traduits en protéines.
- les ARN de transfert
- les ARN ribosomiaux
- les ARN nucléaires
- les ARN cytoplasmiques

### Protéine

- macromolécule : chaîne polypeptidique
- constitué par un enchaînement d'unités élémentaires : les **acides aminés**
- l'ensemble des protéines assurent les principales fonctions cellulaires
- se replie sur elle-même et adopte une conformation ou **structure particulière dans l'espace**. Cette structure tridimensionnelle est à l'origine de la fonction de la protéine et de la spécificité de cette fonction.

Nucléotides - ADN et ARN

## 3. Deux types de bioinformation : la séquence des nucléotides et la séquence des acides aminés

Les chaînes nucléotidiques (ADN, ARN) et les chaînes polypeptidiques (protéines) sont des polymères d'unités élémentaires :

- ADN : 4 désoxyribonucléotides = dCMP, dGMP, dAMP, dTMP
- ARN : 4 ribonucléotides = CMP, GMP, AMP, UMP
- protéines : 20 acides aminés = Ala (A), Cys (C), Asp (D), Glu (E), Phe (F), Gly (G), His (H), Ile (I), Lys (K), Leu (L), Met (M), Asn (N), Pro (P), Gln (Q), Arg (R), Ser (S), Thr (T), Val (V), Trp (W), Tyr (Y)

Elles possèdent 2 extrémités distinctes et sont donc orientées :

- de l'extrémité dite **5'** vers l'extrémité dite **3'** pour les chaînes nucléotidiques
- de l'extrémité dite **N-terminale** vers l'extrémité dite **C-terminale** pour les chaînes polypeptidiques

En conséquence :

- les chaînes nucléotidiques et polypeptidiques sont une succession **ordonnée** et **orientée** d'unités élémentaires
- les **séquences** sont leur **transcription** sous forme d'une succession **ordonnée** et **orientée** de **lettres** qui correspondent à ces unités élémentaires

Exemple de séquence nucléotidique	Exemple de séquence polypeptidique
aattccggca tagaaactca aatcaaagag gaagaaacac cgattctcct tttctctctc taacaacta gatcagatct ctgagtttaa ggaagctttc agcctattcg ataaggatgg cgatggttgc atcacaacca aggagcttgg aactggtatg cgatcattgg gacaaaacc aactgaagca gagcttcagg acatgatcaa	MADQLTDDQI SEFKEAFSLF DKDGDGCITT KELGTVMRSL GQNPTEAELQ DMINEVDADG NGTIDFPEFL NLMARKMKDT DSEELKEAF RVFDKQNGF ISAAELRHVM TNLGEKLTDE EVDEMIREAD VDGDGQINYE EFKVMMMAK

Les séquences constituent l'un des principaux types de bioinformation qu'analyse la bioinformatique.

Exemples d'autres types de bioinformation (directe ou obtenue "in silico")	
Les structures tridimensionnelles des protéines et aussi, malgré leur nombre plus restreint, des acides nucléiques (en particulier les ARN de transfert).	Protein Data Bank
Les données obtenues en protéomique (gels d'électrophorèse bidimensionnel).	SWISS-2DPAGE
Le changement d'un nucléotide dans un gène quelconque ("Single Nucleotide Polymorphism").	SNP
La taxonomie (classification) des organismes.	Taxonomy
Les réseaux d'interactions qu'établissent les molécules biologiques.	BIND
L'ontologie : l'organisation hiérarchique de la connaissance sur un ensemble d'objets par leur regroupement en sous-catégories suivant leurs caractéristiques essentielles.	GO
Les données bibliographiques (diffusion des résultats de la recherche par les articles).	PubMed

## 4. L'obtention des séquences

Séquence des **nucléotides** : méthode de **F. Sanger** (1977)

Séquence des **acides aminés** : méthode de **P. Edman** (1950)

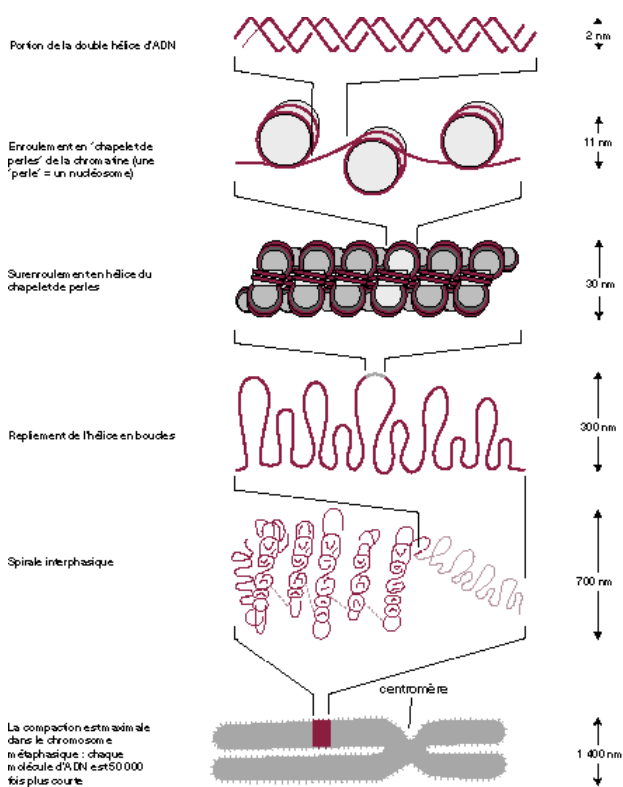
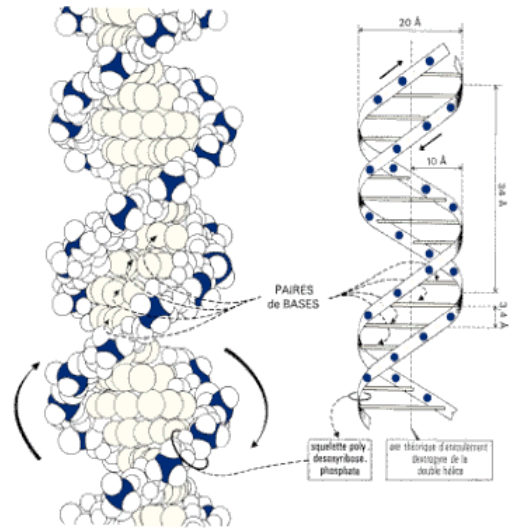
**HAUT PAGE**

**5. Le séquençage du génome humain**

La molécule d'**ADN** est le support biologique de l'information héréditaire.  
 Cette information est transmise par la copie conforme de cette molécule.

C'est une **macromolécule** formée par l'enchaînement de milliers, de millions ou de **milliards** (dans le cas de l'homme) de nucléotides.

L'extension complète de l'ADN humain **formerait un fil de plus d'1,2 m.**



Ci-contre, schéma du super-enroulement de l'ADN dans les **chromosomes** qui explique son extrême compacité.

Source figure : **Alberts et al.** - "Access Excellence"

Ci-dessous, une seule molécule d'ADN (long brin jaune) d'une bactérie *Escherichia coli* éclatée par un choc osmotique.

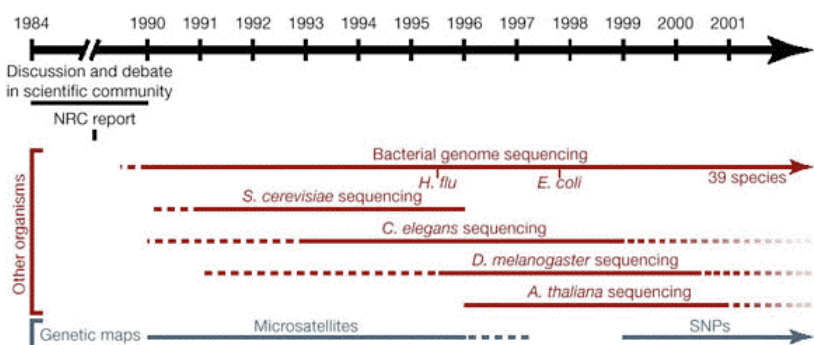
Echelle : un fragment de la molécule d'ADN d'environ la moitié de la longueur de la bactérie correspond à environ 5000 paires de bases.

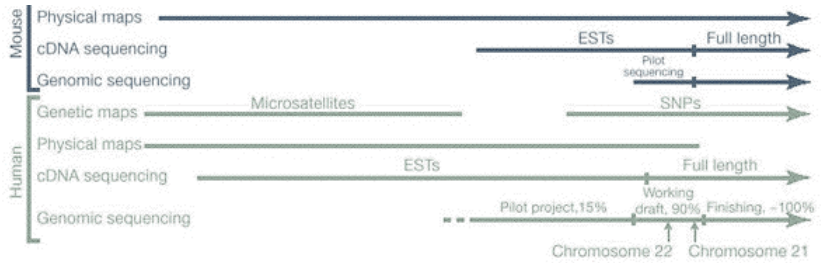


Source figure : "L'information biologique"

Ci-contre, un aperçu du déroulement du séquençage du génome humain.

Source figure : *Nature* 409, 860 - 921





1984	Les pré-projets.	
à	<ul style="list-style-type: none"> <li>L'Initiative du Génome Humain est annoncée et quelques projets technologiques commencent.</li> <li>Création de la fondation HUGO ("Human Genome Organization") pour coordonner le séquençage au niveau mondial (éviter les doublons). Echec du projet (coût trop élevé).</li> </ul>	
1990	<p>Le NIH ("National Institute of Health" - USA) et le DOE ("Department of Energy Office of science" - USA) présentent au Congrès américain le projet "Génome Humain" ("Human Genome Project" - HGP).</p> <p>HGP est un consortium regroupant des laboratoires de différents pays (États-Unis, Royaume-Uni, Japon, Allemagne, Chine et France - Génoscope). Il est financé par des fonds publics et caritatifs.</p> <p>Pour éviter les problèmes liés au dépôt de brevet, les résultats du séquençage sont accessibles à tous sur internet dans les 24 heures.</p>	<p>Les objectifs du HGP étaient de :</p> <ul style="list-style-type: none"> <li>séquencer les 3 milliards de paires de bases du génome humain avec un taux d'erreur minimal</li> <li>identifier tous les gènes.</li> <li>développer des méthodes plus rapides et efficaces pour le séquençage de l'ADN et l'analyse des séquences</li> <li>transférer ces technologies à l'industrie</li> </ul> <p>Le Consortium public n'a pas accordé d'importance au choix des multiples donneurs d'ADN.</p>
1992	Première phase du projet : première carte physique génétique complète (basse résolution).	Voir une belle animation décrivant le principe du séquençage (Jussieu - Génoscope).
1993	Le Généthon fournit des mega-YACs au HGP.	<p>Le YAC ("Yeast Artificial Chromosome") est un vecteur utilisé pour cloner des fragments d'ADN jusqu'à une taille de 400 kb.</p> <p>Le BAC ("Bacterial Artificial Chromosome") est aussi un vecteur utilisé pour cloner des fragments d'ADN jusqu'à une taille de 300 kb.</p>
1997	Séquençage complet du génome de la bactérie <i>Escherichia Coli</i> . Publication de cartes physiques génétiques à haute résolution des chromosomes humains 7 et X.	Le décryptage du génome pose la question de la brevetabilité du vivant, l'UNESCO le 11 novembre 1997 à déclaré que le génome humain est un patrimoine de l'humanité, or un patrimoine de l'humanité ne peut pas être la propriété d'un individu. Donc, une séquence d'ADN ne peut pas être brevetée.
1998	Création de la société "Celera Genomics" (USA) par Craig Venter dans le but de séquencer le génome humain en compétition avec l'HGP. Cette société a fait le choix de séquencer l'ADN de cinq personnes d'origine : africaine, asiatique, caucasienne et latino-américaine.	Une vraie course au séquençage est lancée : l'enjeu est la propriété publique ou privée du génome humain dans le but d'une exploitation commerciale des tests et des médicaments (brevets).
1999	Première séquence complète du chromosome 22 établie par HGP.	
2000	HGP annonce 90 % du séquençage du génome humain. "Celera Genomics" propose les premiers résultats du séquençage total du génome d'une personne. Publication du génome du chromosome 21. Publication du génome complet de la mouche <i>Drosophila melanogaster</i> .	<p>Communiqué commun de Tony Blair et Bill Clinton (14 mai 2000) qui annoncent leur souhait que les résultats du séquençage soient en accès libre et que les brevets soient limités à leur exploitation industrielle et commerciale.</p> <p>L'entreprise "Celera Genomics" va donc devoir rendre ses résultats publics de manière trimestrielle.</p>
Février 2001	La même semaine, publication du brouillon initial des travaux de séquençage du génome humain complet par : <ul style="list-style-type: none"> <li>Le HGP dans la revue : <i>Nature</i> 409, 860 - 921</li> <li>"Celera Genomics" dans la revue : <i>Science</i> 291, 1304-1351</li> </ul> <p>Le génome humain contient 3,2 milliards de nucléotides et il contiendrait environ 30.000 à 40.000 gènes (chiffre nettement inférieur à ce qui était prévu).</p> <p>Cependant, la structure des gènes est très complexe (introns, exons, épissage alternatif) : un même gène peut coder différentes protéines.</p> <p>On ignore le nombre total de protéines différentes que peut fabriquer l'organisme humain. Pour le moment, on estime ce nombre entre 50 000 et 150 000.</p>	<p>1. L'ADN humain est extrêmement hétérogène.</p> <p>Les gènes ne sont pas répartis uniformément sur le génome. Il existe des zones qui n'en contiennent aucun, mais possèdent des séquences répétitives. On ne connaît pas encore les fonctions de ces parties de l'ADN.</p> <p>2. Les résultats de la société "Celera Genomics" montrent qu'il y a plus de différences entre l'ADN des deux Caucasiens qu'entre celui d'un Africain et d'un Caucasien.</p> <p>Tous les êtres humains sont différents et cette différence résulte des variations entre l'ADN des individus. Ces variations correspondent au changement d'un nucléotide dans un gène quelconque que l'on appelle un "Single Nucleotide Polymorphism" (SNP).</p>

Les SNPs représentent **0,1% de différence entre deux génomes** (plus de 1,4 millions de SNP ont été identifiés). Ils sont particulièrement intéressants pour la médecine et l'industrie pharmaceutique (détermination de l'origine de nombreuses maladies, développement de tests de prédisposition aux maladies, synthèse de médicaments en fonction de la sensibilité génétique). Voir "Société : champs d'application de la bioinformatique et des domaines afférant".

3. Le gène ZNF217 est identifié dans le chromosome 20. Ce gène apparaît en nombre croissant de copies dans beaucoup de tumeurs. Il jouerait un rôle dans le cancer du sein.

Résultats du séquençage du génome humain

HAUT PAGE

## 6. Le stockage de la bioinformation : les banques de données

Les fichiers contenant l'information biologique sous la forme de séquences est l'élément central autour duquel les **banques de données** se sont constituées. Il existe un grand nombre de bases de données d'intérêt biologique. On peut distinguer :

- les banques de données ou bases de données **généralistes** : elles correspondent à une collecte des données la plus exhaustive possible et qui offrent un ensemble plutôt hétérogène d'informations
- les bases de données ou bases de données **spécialisées** : elles correspondent à des données plus homogènes établies autour d'une thématique et qui offrent une valeur ajoutée

Exemple de grandes banques généralistes :

- **EMBL** : Banque européenne créée en 1980 et financée par l'EMBO (*European Molecular Biology Organisation*). Elle est aujourd'hui diffusée par l'EBI (*European Bioinformatics Institute*, Cambridge)
- **Genbank** : Créée en 1982 par la société IntelliGenetics et diffusée maintenant par le NCBI (*National Center for Biotechnology Information*, Los Alamos)
- **DDBJ** (*DNA Data Bank of Japan*) : Créée en 1986 et diffusée par le NIG (*National Institute of Genetics*, Japon).

Ces trois banques s'échangent systématiquement leur contenu depuis 1987 et adoptent un système de conventions communes (*The DDBJ/EMBL/GenBank Feature Table Definition*).

- **PIR-NBRF** (*Protein Identification Resource*) : banque de protéines créée sous l'influence du NBRF (*National Biomedical Research Foundation*) à Washington. Elle diffuse maintenant des données issues du MIPS (*Martinsried Institute for Protein Sequences*), de la base Japonaise JIPID (*Japan International Protein Information Database*) et des données propres de la NBRF.
- **Swissprot** : elle a été constituée à l'Université de Genève à partir de 1986 et regroupe entre autres des séquences annotées de la PIR-NBRF ainsi que des séquences codantes traduites de l'EMBL.

Exemple d'une banque spécialisée :

- **Disulfide Bridge DataBase** (DBDB - J.-M. Richer, G. Hunault & E. Jaspard) : Cette base de données contient un grand nombre d'informations structurales sur les cystéines de plus de 400 protéines cristallisées. Elle a aussi pour but de servir à la mise au point d'un logiciel de prédiction des cystéines impliquées dans la formation de pont disulfure.

Les multiples liens entre les groupes de données dans les banques généralistes sont d'une **complexité** étonnante. Voir les exemples de **Genbank** ou **ExpASY**.

Développement sur les banques généralistes et spécialisées

## Projet "Embrace" : standardiser l'accès aux données bioinformatiques en Europe

Projet de réseau bioinformatique européen (début février 2005), doté par la commission européenne de 8,28 millions d'euros pour cinq ans, il rassemble 16 partenaires de 11 pays, dont 3 équipes du CNRS.

Depuis la première ébauche du séquençage du génome humain, en 2001, les biologistes produisent toujours plus de données bioinformatiques. Pour les stocker, ils **multiplient les bases de données**, mais **aucune standardisation n'existe** encore dans ce domaine. En conséquence, les données ne sont pas toutes présentées sous la même forme. De plus, les protocoles d'accès sont différents, de même que les algorithmes qui traitent les données.

Le but du projet "Embrace" est de standardiser l'accès aux innombrables données issues des projets de génomique, et leurs méthodes d'étude, afin que les chercheurs puissent les consulter et les exploiter facilement. **Embrace** utilisera la "technologie de grille" (*grid*), où un grand nombre d'ordinateurs géographiquement éloignés les uns des autres travaillent en réseau afin d'offrir une importante capacité de stockage et de calcul.

HAUT PAGE

## 7. La structuration de la bioinformation: fichiers et formats

Les séquences sont stockées en général sous forme de **fichiers** texte qui peuvent être soit des fichiers personnels (présents dans un espace

personnel), soit des fichiers publics (séquences des banques) accessibles par des programmes interfaces (tels que SRS, GCG, Entrez).

Le **format** correspond à l'ensemble des règles (contraintes) de présentation auxquelles sont soumises la ou les séquences dans un fichier donné. Le format permet :

- une mise en forme automatisée
- le stockage homogène de l'information
- le traitement informatique ultérieur de l'information.

Pour lire et traiter les séquences, les logiciels d'analyse autorisent un ou plusieurs formats des données.

Exemples de formats

### 8. Exemples d'algorithmes et de programmes en bioinformatique

La bioinformatique utilise des programmes **spécifiques** écrits dans des langages qui peuvent lui être spécifiques. En effet, si la recherche de motifs dans les séquences est bien traitée par les algorithmes d'analyse de texte ("*combinatorial pattern matching*"), la séquence seule ne suffit pas pour déterminer la fonction de certaines macromolécules comme les ARN, car il faut tenir compte de leur structure tridimensionnelle. Dans ce cas, l'analyse bioinformatique nécessite de nouvelles **méthodes** :

- le développement de structures de données et d'**algorithmes** (arbres et tableaux de suffixes, automates...)
- la construction automatique (inférence) de structures d'ARN à partir d'alignements de séquences d'ARN ou d'un ensemble de séquences de même type et issues de différents organismes.

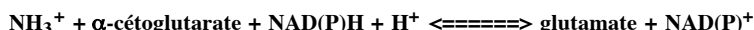
Il est difficile de définir en informatique les objets que manipulent les biologistes. Depuis des années, les bioinformaticiens utilisent des concepts aussi divers que :

- le recuit simulé
- les chaînes de Markov
- les statistiques bayésiennes
- les réseaux de neurones
- approche par modélisation ("*model-driven approach*") : les objets (par exemple, les séquences, les structures, les motifs, ...) ne sont plus étudiés (alignés, classés, ...) par comparaison directe (deux à deux ou multiple) mais à travers la construction d'un modèle qui tente d'en capturer les propriétés communes

Exemples d'algorithmes et programmes

### 9. Exemple d'analyse bioinformatique d'une enzyme : la glutamate déshydrogénase (GDH)

On peut considérer la première réaction d'assimilation de l'azote (sous forme d'ammoniac) par la glutamate déshydrogénase (GDH) comme un point d'entrée dans le métabolisme azoté. L'atome d'azote est à l'origine de la fonction α-aminée des acides aminés selon la réaction :



Il existe trois isoformes de GDH :

- la GDH EC 1.4.1.2 qui catalyse la réaction dans le sens de la désamination essentiellement
- la GDH EC 1.4.1.3 qui catalyse la réaction dans les deux sens
- la GDH EC 1.4.1.4 (**GDH4**) qui catalyse la réaction dans le sens de formation du glutamate

La **GDH4** joue peut-être un rôle clé dans l'assimilation de l'azote. Or ce rôle n'a pas encore été démontré, notamment chez les plantes. Par ailleurs, on ne dispose d'aucune information concernant la structure de la GDH4.

La bioinformatique permet l'étude prospective de la **relation structure - fonction** de la GDH.

Supports pour le cycle de cours - atelier CC30



### 10. Les domaines d'étude en "ome" ou "omique"

Le plein essor de la bioinformatique (telle qu'on la conçoit maintenant) a commencé avec la **génomique** (début des années 80) marquée par des événements phares comme le séquençage du génome humain.

Puis la quantité et la multiplicité d'informations biologiques obtenues de même que les avancées technologiques ont fait apparaître **au fur et à mesure** d'autres domaines d'étude en "ome" ou "omique".

Leur définition évolue ce qui explique qu'elle puisse différer selon les sources.

Domaine d'étude	source	définition
Génomique	Les Actions Concertées Incitatives Le programme Génomique	La génomique est une nouvelle discipline de la biologie qui vise à l'analyse et physiologique complète du matériel héréditaire des organismes vivants et à déduire les fonctions des gènes et leurs interactions à partir de leurs séquences. [...] facilite l'intégration de la génomique dans la physiologie. [...]

	Ministère de la Recherche	La génomique repose sur des techniques qui évoluent à une vitesse prodigieuse et l'ensemble des bases de données obtenues, qui représente des dizaines de milliards de caractères répartis en dizaines de milliers de fichiers, pose déjà aux informaticiens des problèmes considérables pour classer et interpréter cette énorme masse de données.
	Rapport sur Génomique et informatique L'impact sur les thérapies et sur l'industrie pharmaceutique F. Sérusclat - Sénateur (1999)	<ul style="list-style-type: none"> <li>• La génomique est l'étude exhaustive des <b>génomés</b> et en particulier de l'ensemble des gènes, de leur disposition sur les chromosomes, de leur séquence, de leur fonction et de leur r<sup>TM</sup>le.</li> <li>• La cartographie physique est le positionnement de repères sur le génome.</li> <li>• Le séquençage : pour connaître les "instructions " que renferme un fragment d'ADN, on lit la succession des bases puriques et pyrimidiques (A, T, G, C) de l'enchaînement. Cette lecture est appelée séquençage.</li> </ul>
Génomique structurale	Infobiogen	La génomique structurale (qu'il serait plus judicieux et exact de nommer "protéomique structurale") s'attèle à déterminer la structure tridimensionnelle des protéines codées par le génome. On s'aperçoit de l'ambiguïté du nom de génomique structurale.
Génomique fonctionnelle ou post-génomique		La génomique fonctionnelle ou post-génomique étudie le <b>transcriptome</b> (ensemble des ARN messagers transcrits à partir du génome). Le but est de déterminer la fonction des gènes à partir de leurs produits d'expression (ARN et protéines) et d'étudier leur mode de régulation et leurs interactions.
Protéomique		La protéomique étudie le <b>protéome</b> : ensemble des protéines identifiées à partir d'un génome. Toutes les cellules de l'organisme possèdent le même génome, mais ont un protéome différent selon l'organe et le moment du développement de l'individu. La protéomique s'attache à déterminer la localisation, la structure et la fonction de ces protéines. Elle analyse leurs interactions et leurs modifications au cours du temps.
Les autres "omes" et "omiques"		<p><b>Métabolome</b> - métabolomique : description de la population des différents métabolites d'une cellule dans un stade physiologique donné.</p> <p><b>Interactome</b> - interactomique : description des interactions entre toutes les macromolécules d'une cellule.</p> <p>On trouve aussi dans la <b>littérature</b> : lipidome, protéasome, régulome, spliceosome, prédicome, réactome</p>

11. Société : champs d'application de la bioinformatique et des domaines afférant

Santé (source : Rapport sénatorial sur "Génomique et informatique" - 1999)	
La génomique devrait permettre une évolution fondamentale de notre approche en soins de santé, beaucoup plus axée sur la prévention que sur le traitement.	
Le diagnostic moléculaire (tests ciblant le patrimoine génétique).	Diagnostic de maladies infectieuses et génétiques (connaissances des prédispositions génétiques).
Tests de dépistage de maladies.	On peut supposer qu'il sera possible de déceler et de diagnostiquer les maladies beaucoup plus tôt, des années souvent avant l'apparition de symptômes. D'où des traitements précoces donc plus efficace et à moindre coût.
La pharmacogénomique : adaptation des traitements aux malades en fonction de leur profil génétique.	La séquence d'ADN d'un individu pourrait devenir un élément fondement de son traitement. Les patients pourront subir un test et voir si un médicament leur convient.
Détection de cibles d'intérêt pour la mise au point de nouveaux médicaments.	La génomique jouera un rôle primordial dans la mise au point des médicaments.
Maîtrise de la toxicité des médicaments.	En utilisant la génomique pour présélectionner les sujets, les fabricants de médicaments pourront cibler leurs essais.
Thérapie génique : réparation d'un gène ou apport <i>in situ</i> d'un gène fonctionnel	Utilisation de gènes normaux pour remplacer un gène défectueux, y suppléer ou stimuler une résistance totale à une maladie. Elle pourrait permettre de soigner (voire de guérir) des maladies génétiques et acquises.
Mise au point de nouveaux vaccins.	Vaccins à base d'ADN et vaccins "traditionnels" découverts grâce à la connaissance du génome des bactéries.
Les protéines thérapeutiques.	L'utilisation des techniques du génie génétique pour la production de protéines par des bactéries ou des levures et, plus récemment par des animaux génétiquement modifiés.
Agro-alimentaire	<ul style="list-style-type: none"> <li>• Plantes résistantes aux pesticides, aux pathogènes, aux insectes, à la sécheresse permettant de nourrir les populations.</li> <li>• Tolérance au stress environnementaux.</li> <li>• Rendements et qualité.</li> </ul>
Environnement	<ul style="list-style-type: none"> <li>• Lutte contre les pollutions chimiques ou biologiques (traitements de l'eau, des sols, des déchets)</li> <li>• Dégradation de pesticides par les bactéries</li> <li>• Préservation de la biodiversité</li> <li>• <b>Gestion du milieu marin</b> : étude de la structure et du fonctionnement des écosystèmes marins et apport des modèles marins à la compréhension de l'évolution des fonctions biologiques</li> <li>• Nouvelles sources d'énergie.             <ol style="list-style-type: none"> <li>1. Dans un avenir immédiat, la génomique peut servir à mieux comprendre comment les plantes peuvent retirer plus de carbone de l'atmosphère et en retenir davantage lorsqu'elles meurent et se décomposent.</li> <li>2. Des solutions biotechnologiques au problème du <b>changement climatique</b> seront envisageables quand la biologie moléculaire des plantes et des micro-organismes comme les bactéries et les algues sera</li> </ol> </li> </ul>

- élucidée.
3. Par exemple, les cyanobactéries transforment le CO<sub>2</sub> de l'atmosphère en composés organiques, sans produire de gaz à effet de serre. L'environnement peut tirer profit de la pleine compréhension du mode de fonctionnement de tels organismes *via* la génomique.

## 12. Liens Internet et références bibliographiques

Informations générales	
<p>La Bioinformatique (Infobiogen - Génopole d'Evry - France)</p> <p>Cours : "Analyse bioinformatique des séquences"</p> <p>Site : "L'information génétique" - Simple et didactique</p>	<p><a href="#">Infobiogen</a></p> <p><a href="#">GeNET- TOURS</a></p> <p><a href="#">Aller au site</a></p>
<p>Séquençage du génome humain : "The Human Genome Project"</p> <p>Actualités concernant la génomique et les biotechnologies</p>	<p><a href="#">HGP</a></p> <p><a href="#">Génopôle Evry</a></p>
<p>Revue scientifique : "Bioinformatics"</p> <p>Revue scientifique : "Nucleic Acids Research"</p> <p>Revue scientifique : "OMICS: A Journal of Integrative Biology"</p>	<p><a href="#">Bioinformatics</a></p> <p><a href="#">NAR</a></p> <p><a href="#">OMICS</a></p>
<p>Fields &amp; Johnston (2002) "Genomics : A Crisis in Postgenomic Nomenclature" <i>Science</i> 296, 671 - 672</p> <p>Kanehisa &amp; Bork (2003) "Bioinformatics in the post-sequence era" <i>Nature Genetics</i> 33, 305 - 310</p>	<p><a href="#">Article</a></p> <p><a href="#">Article</a></p>
Algorithmes et programmes en bioinformatique	
<p>Needleman, S.B. &amp; Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequence of two proteins" <i>J. Mol. Biol.</i> 48, 443 - 453</p> <p>Smith, T. &amp; Waterman M. (1981) "Identification of common molecular subsequences" <i>J. Mol. Biol.</i> 147, 195 - 197</p>	
<p>Clustal W : Thompson, J. D., Higgins, D. G. &amp; Gibson, T. J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice" <i>Nuc. Acids Res.</i> 22, 4673 - 4680</p>	<p><a href="#">ClustalW</a></p> <p><a href="#">Article</a></p>
<p>FASTA : Pearson, W. R. &amp; Lipman, D. J. (1988) "Improved tools for biological sequence comparison" <i>Proc. Natl. Acad. Sci. USA</i> 85, 2444 - 2448</p>	
<p>BLAST : Altschul, S. F., Gish, W., Miller, W., Myers, E. W. &amp; Lipman, D. J. (1990) "Basic local alignment search tool" <i>J Mol Biol</i> 215, 403 - 410</p>	<p><a href="#">BLAST</a></p> <p><a href="#">Article</a></p>
<p>Multalin : Corpet, F. (1988) "Multiple sequence alignment with hierarchical clustering" <i>Nuc. Acids Res.</i> 16, 10881 - 10890</p>	<p><a href="#">Multalin</a></p>
<p>ESpript : Gouet, P., Courcelle, E., Stuart, D. I. &amp; Metz, F. (1999) "ESpript : analysis of multiple sequence alignments in PostScript" <i>Bioinformatics</i> 15, 305 - 308</p>	<p><a href="#">ESpript</a></p>
<p>ESyPred3D : Lambert, C., Leonard, N., De Bolle, X. &amp; Depiereux, E. (2002) "ESyPred3D : Prediction of proteins 3D structures" <i>Bioinformatics</i> 18, 1250 - 1256</p>	<p><a href="#">ESyPred3D</a></p>
<p>Swiss-PdbViewer : Guex, N. &amp; Peitsch M. C. (1997) "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling" <i>Electrophoresis</i> 18, 2714 - 2723</p>	<p><a href="#">DeepView</a></p>

**HAUT PAGE**