

Le plant

- PRINCIPLE DE BASE DE L'ALIGNEMENT DE SÉQUENCES
- ALIGNEMENT ET DÉTERMINATION DU SCORE
- MATRICES DE SCORE
- LES TYPES D'ALIGNEMENT

PRINCIPE DE BASE DE L'ALIGNEMENT DE SÉQUENCES

La diversité génétique est due à des **mutations ponctuelles** et à des **insertions /délétion** apparues au cours de l'évolution .

Comparer des séquences par alignement permet de déterminer et estimer leur degré de similarité et d'émettre une hypothèse quant à leur parenté évolutive (homologie).

Il faut bien distinguer les termes identité, similitude et homologie :

-Identité : estimation de la fraction de résidus identiques entre deux séquences (fondées sur un alignement)

-Similitude : estimation de la fraction de résidus Similaires entre deux séquences. Deux résidus sont considérés comme similaires lorsque leur score de substitution est supérieur à 0. La mesure de similarité dépend d'un alignement et d'une matrice de substitution.

-Homologie : désigne une parenté évolutive entre deux séquences, c'est-à-dire dérivant d'un ancêtre commun.

Pourquoi avons-nous besoin de comparer des séquences?

La comparaison des séquences est un aspect fondamental de la bioinformatique et très souvent la première étape de l'analyse de séquences. Il est nécessaire pour:

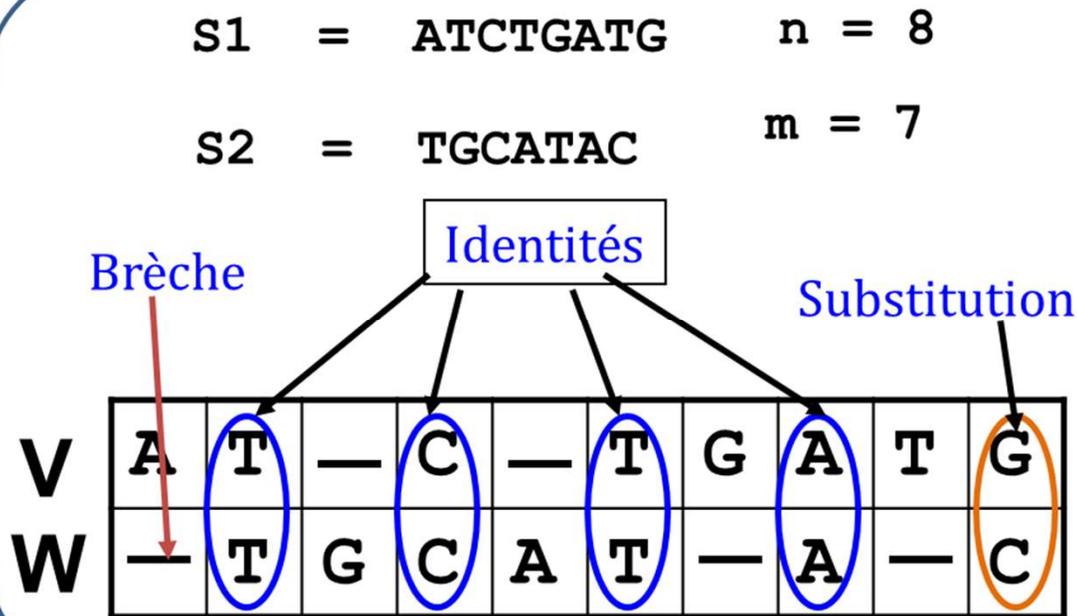
- La recherche de fonctions biologiques similaires.
- Construction d'arbres phylogénétiques.
- Identification de mutations dans des gènes.
- Prédiction des sites d'épissage dans les séquences eucaryotes.
- Détection du transfert de gène

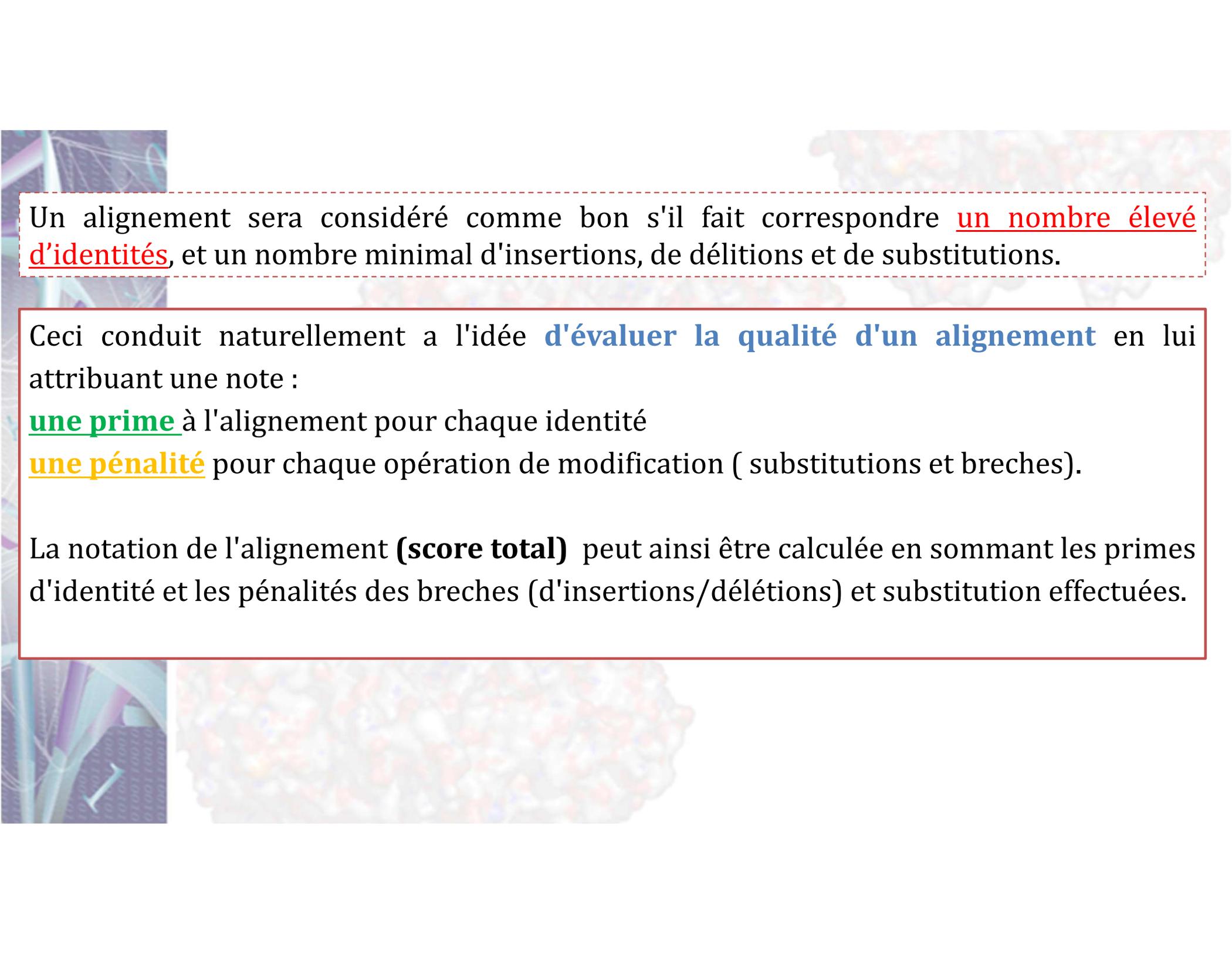
Que faisons-nous de comparer?

Les méthodes de comparaison et le type de similitude observée dépendent avant tout de la nature des séquences : ADN (codage ou non codante), ARN, protéines

ALIGNEMENT ET DÉTERMINATION DU SCORE :

Aligner deux séquences, c'est rechercher **le maximum d'appariement** entre les lettres qui les composent (nucléotides ou résidus d'acides aminés) **avec le minimum de mésappariement et des brèches** (gaps) (voir schéma).





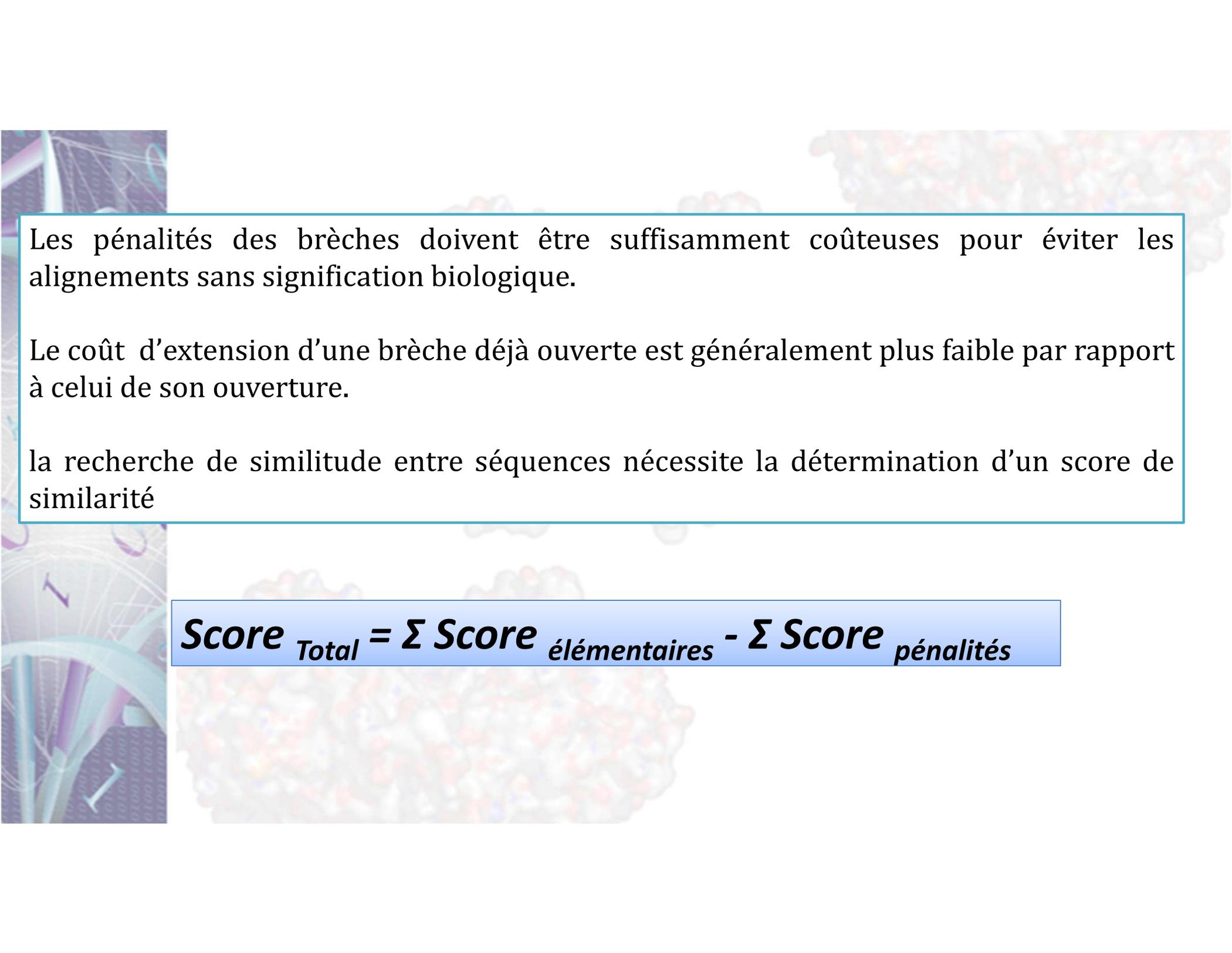
Un alignement sera considéré comme bon s'il fait correspondre un nombre élevé d'identités, et un nombre minimal d'insertions, de délétions et de substitutions.

Ceci conduit naturellement à l'idée **d'évaluer la qualité d'un alignement** en lui attribuant une note :

une prime à l'alignement pour chaque identité

une pénalité pour chaque opération de modification (substitutions et breches).

La notation de l'alignement (**score total**) peut ainsi être calculée en sommant les primes d'identité et les pénalités des breches (d'insertions/délétions) et substitution effectuées.



Les pénalités des brèches doivent être suffisamment coûteuses pour éviter les alignements sans signification biologique.

Le coût d'extension d'une brèche déjà ouverte est généralement plus faible par rapport à celui de son ouverture.

la recherche de similitude entre séquences nécessite la détermination d'un score de similarité

$$\text{Score}_{\text{Total}} = \sum \text{Score}_{\text{élémentaires}} - \sum \text{Score}_{\text{pénalités}}$$

Exemple de détermination de score avec la matrice unitaire (l'appariement vaut +1, le mésappariement vaut 0 et une brèche vaut -1)

Alignement sans brèches	Alignement avec brèches
Séquence 1 ATGACTGGGCCACT Séquence 2 ATACTGGGACAACT	
Séquence 1 ATGACTGGG CCACT Séquence 2 ATACTGGG ACAACT	Séquence 1 ATGACTGGG CC- ACT Séquence 2 AT- ACTGGGACAACT
8 appariements (match) et 6 mésappariement (mismatch).	12 appariements, 1 mésappariement et 2 brèches.
Score $8 - 0 = 8$	Score $12 - 2 = 10$

Les matrices **BLOSUM** (de Steve **Henikoff 1950**) (**BLO**cks **SU**bstitution **M**atrix) sont déduites d'alignements de fragments (Blocks) de protéines très éloignées.

Par exemple: BLOSUM62 est déduite à partir d'un alignement de séquences ayant 62% de similitude. Ces matrices sont bien adaptées aux recherches de séquences dans les banques de données (Blast, FASTA).

Chaque score donne le coût de remplacement d'un résidu par un autre. On note que :

- Les acides aminés rares ont un score élevés (Trp, Cys, His)
- les acides aminés communs ont des scores faibles (Ala, Leu, Ile,.....)
- les substitutions conservative entre acides aminés similaire sont peu pénalisantes. Ces substitutions peuvent se produire sans affecter l'activités de la protéine (ex : Lys↔Arg) .

Autre matrice : **PAM** (de Margaret Oakley **Dayhoff 1925-1983**) (**P**oint **A**ccepted **M**utation) déduites d'alignements globaux de famille de protéines très proches (exemple : cytochromes, hémoglobines

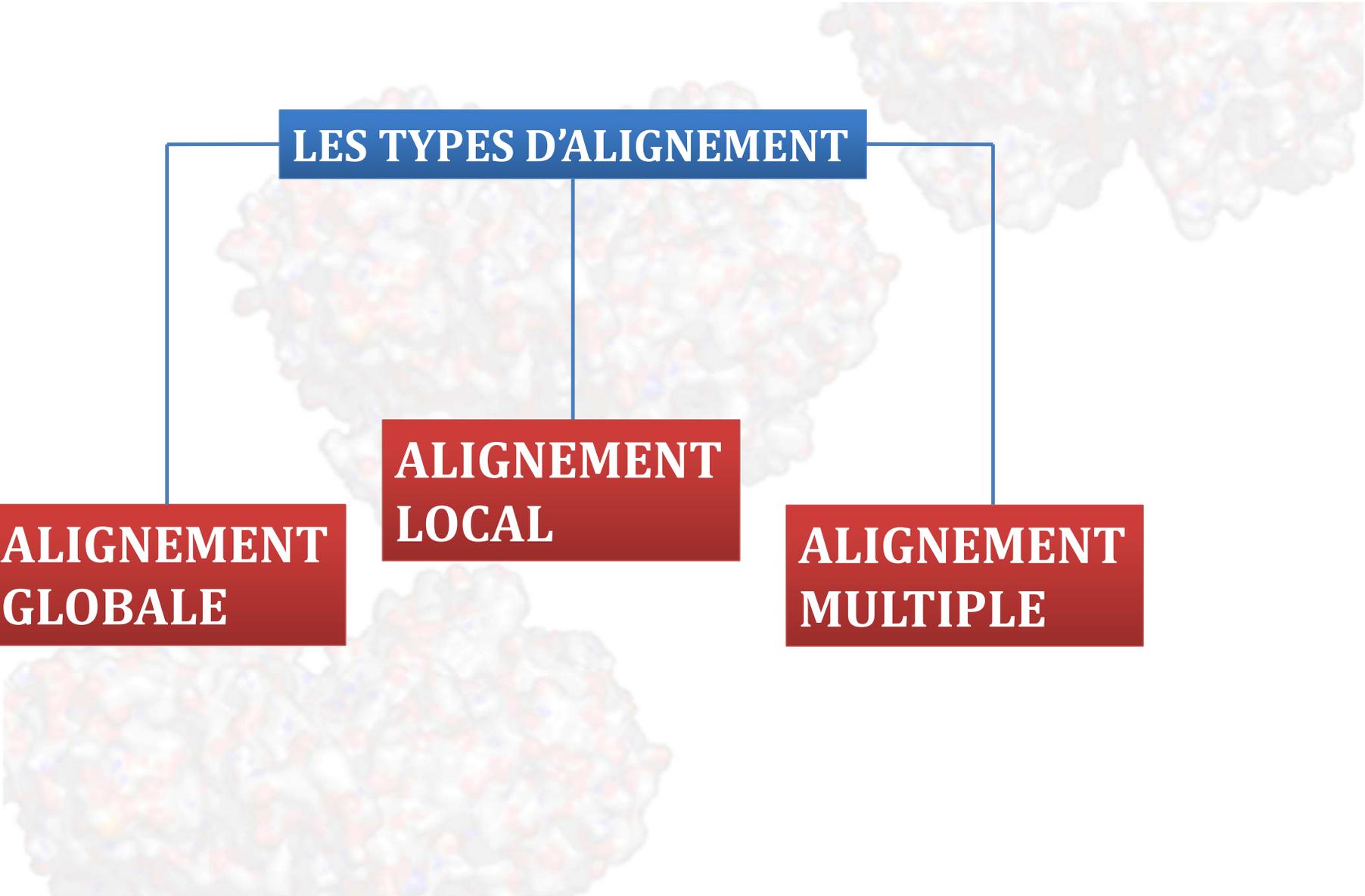


LES TYPES D'ALIGNEMENT

**ALIGNEMENT
GLOBALE**

**ALIGNEMENT
LOCAL**

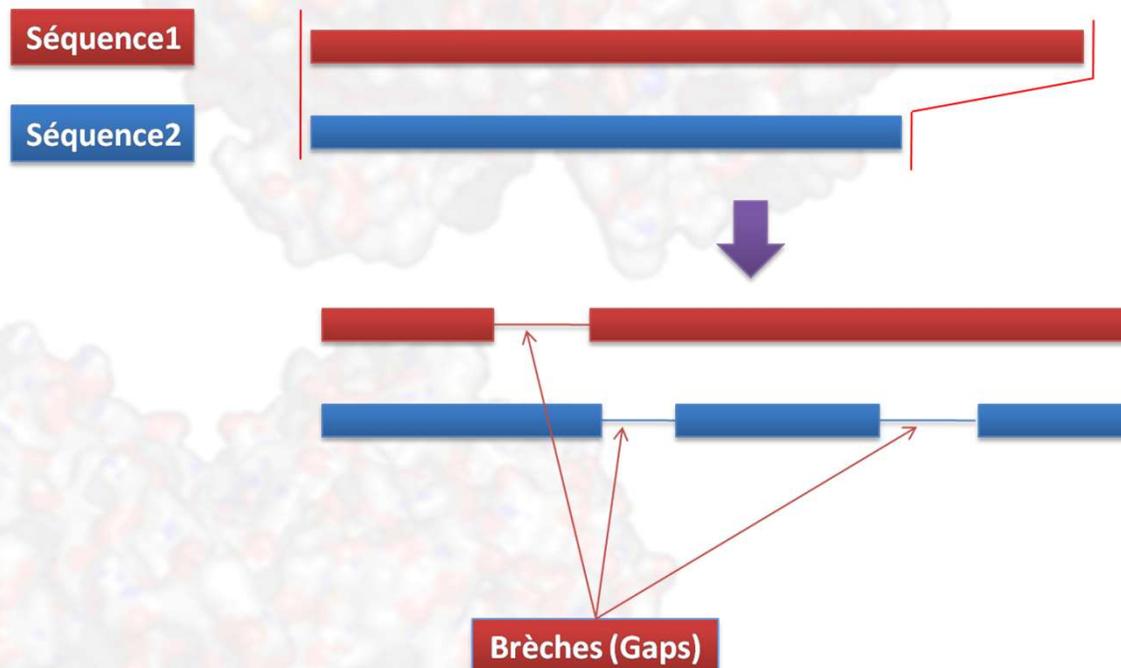
**ALIGNEMENT
MULTIPLE**



Alignement globale :

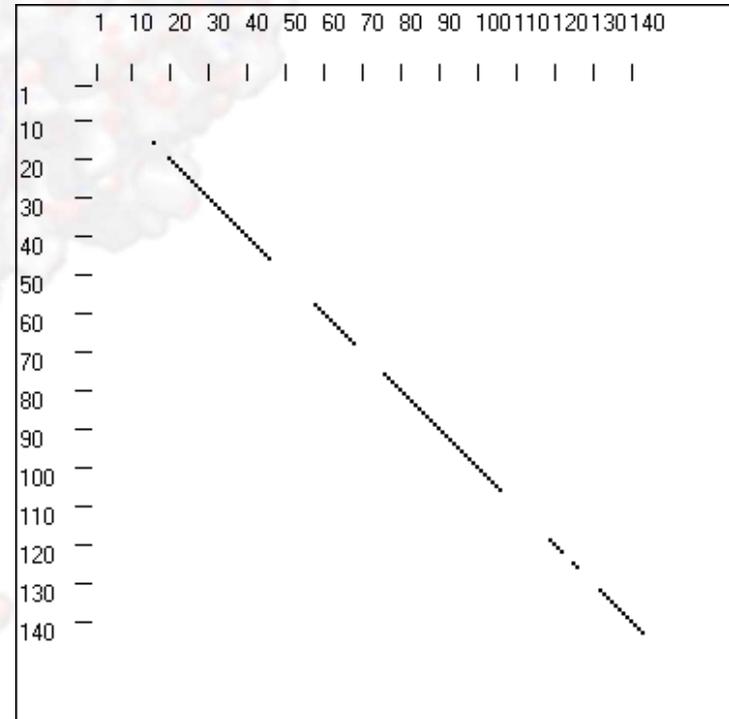
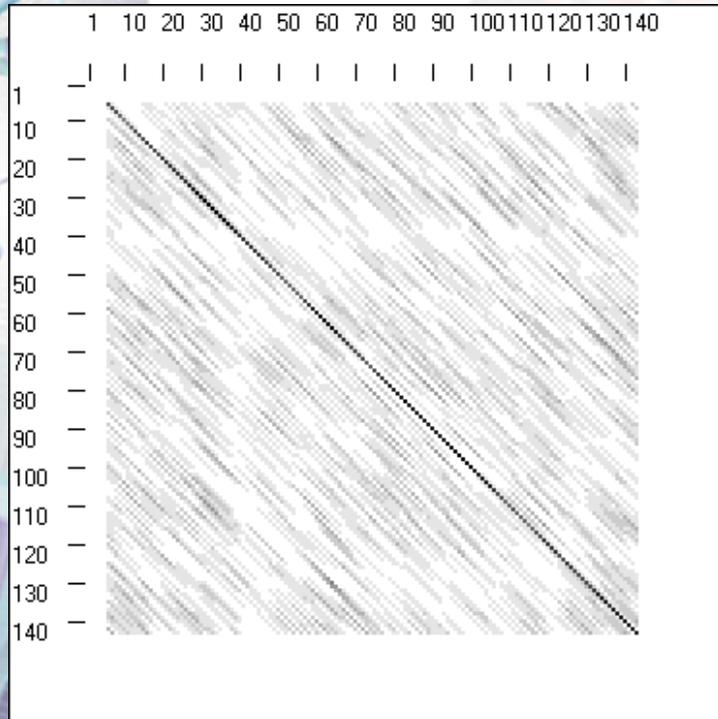
Alignement de deux séquences sur la totalité de leur longueur en tenant compte de tous les résidus. Si les longueurs des séquences sont différentes des insertions / délétions sont introduites pour aligner les deux extrémités des deux séquences.

Il permet de mesurer le degré de similitude entre deux séquences connues.



Avantage et inconvénients : simple et intuitif. Mais des problèmes de bruit de fond se posent pour les longues séquences.

Cela nécessite l'utilisation d'un filtrage : on ne met un point que si **n** caractères sont identiques dans une fenêtre donnée, pour éliminer les segments de similitudes courtes.



Needle (Algorithme de Needleman et Wunsch, 1970)

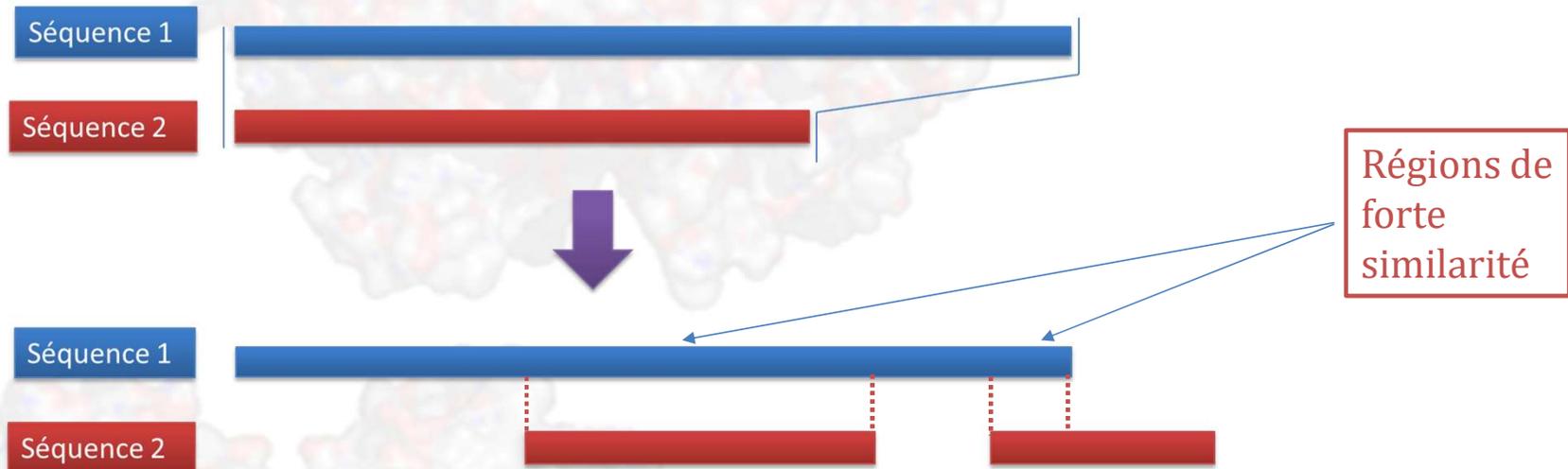
Principe : Needle permet de déterminer le meilleur alignement évalué par un score. Pour cela il construit le chemin optimal allant du coin inférieur droit au coin supérieur gauche d'un tableau à deux dimensions.

Chaque case contient le score cumulé à partir des scores précédents autour de cette case. L'alignement optimal est obtenu en remontant à travers la matrice de la position finale à la position initiale.

	A	C	T	G	A	T	T	C	A
0	-2	-4	-6	-8	-10	-12	-14	-16	-18
A	-2	2	0	-2	-4	-6	-8	-10	-14
C	-4	0	4	2	0	-2	-4	-6	-10
G	-6	-2	2	1	4	2	0	-2	-6
C	-8	-4	0	-1	2	1	-1	-3	0
A	-10	-6	-2	-3	0	4	2	0	2
T	-12	-8	-4	0	-2	2	6	4	2
C	-14	-10	-6	-2	-4	0	4	2	6
A	-16	-12	-8	-4	-5	-2	2	1	4

Alignement local :

Alignement de deux séquences portant sur des régions isolées et permettant de trouver des segments qui ont un haut degré de similarité. Outil efficace et rapide de recherche dans les bases de données en comparant une séquence inconnue à celles de la banque. (BLAST, FASTA)



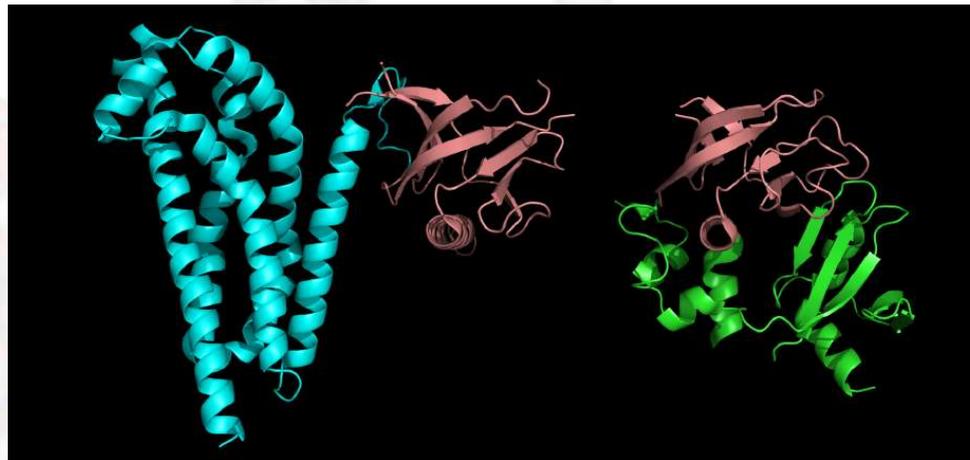
Application :

Le fait d'ignorer segments d'ADN non-codant:

- Régions **non codantes** sont plus susceptibles d'être soumises à des mutations que les régions codantes.
- Alignement local entre deux séquences est susceptible d'être entre deux exons.

Localisation domaines protéiques:

- Les protéines de type différent et de différentes espèces présentent souvent des similitudes locales.
- Similitudes locales peuvent indiquer "sous-unités fonctionnelles»





QUELQUES LOGICIÉLS



BLAST (Basic Local Alignment Search Tool) :

- . Programme pour la recherche de similarités dans les bases de données
- . Utilise un algorithme heuristique linéaire pour l'alignement local
- . Séquences nucléiques et protéiques
- . Disponible sur le Web
- . Connecte aux principales banques de données

FASTA (FAST All):

L'algorithme est basé sur l'identification rapide des zones d'identité entre la séquence recherchée et les séquences de la banque de données. Cette reconnaissance est essentielle car elle permet de considérer uniquement les **séquences présentant une région de forte similitude** avec la séquence recherchée.

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

Search Betacoronavirus Database

We have created a new BLAST database focused on the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. For further detail please visit

[NCBI GenBank.](#)

Mon, 03 Feb 2020 10:00:00 EST

[More BLAST news...](#)

Web BLAST



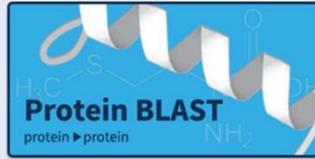
Nucleotide BLAST
nucleotide ▶ nucleotide



blastx
translated nucleotide ▶ protein

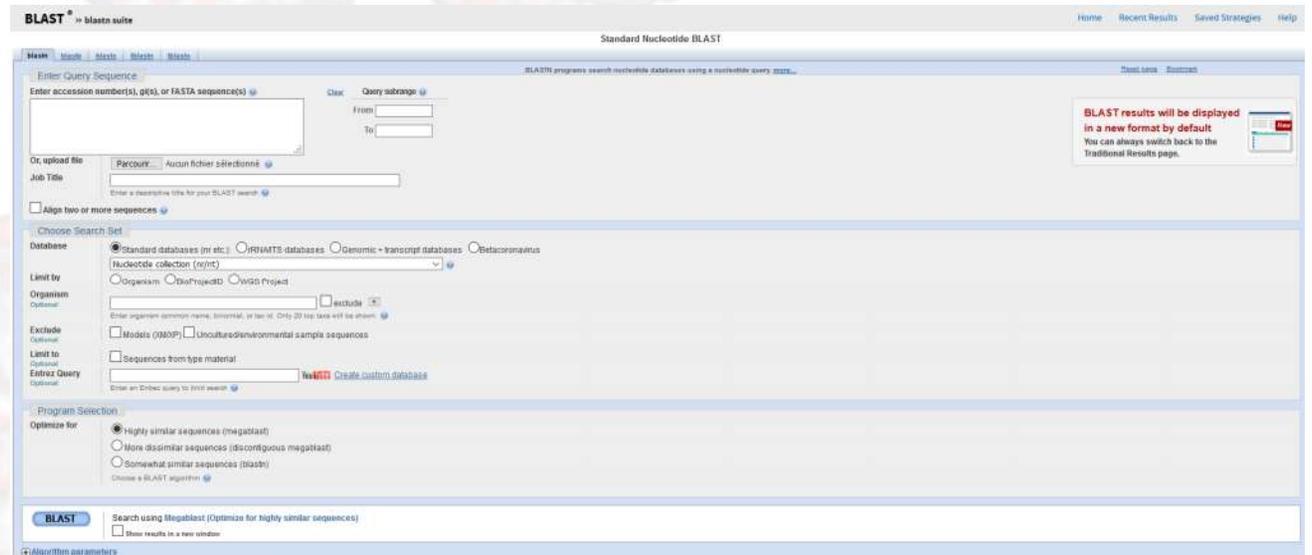


tblastn
protein ▶ translated nucleotide



Protein BLAST
protein ▶ protein

Interface de l'algorithme



The screenshot shows the BLAST web interface. At the top, it says "BLAST® - blastn suite" and "Standard Nucleotide BLAST". There are navigation links for "Home", "Recent Results", "Saved Strategies", and "Help". The main area is titled "Enter Query Sequence" and includes a text input field for "Enter accession number(s), g(e)n, or FASTA sequence(s)", a "Query subrange" section with "From" and "To" fields, and an "Or, upload file" section with a "Parcourir..." button and "Aucun fichier sélectionné" text. Below this is a "Job Title" field. There are checkboxes for "Align two or more sequences" and "Choose Search Set". The "Choose Search Set" section includes "Database" (Standard databases (nr etc.), RefSeq databases, Genomic + transcript databases, Betacoronavirus), "Limit by Organism" (Organism, TaxProject, WGS Project), "Exclude" (Models (3DOP), Uncultured/environmental sample sequences), and "Limit to" (Sequences from type material). There is also an "Enter Query" field and a "Create custom database" button. At the bottom, there is a "Program Selection" section with "Optimize for" options: "Highly similar sequences (megablast)", "More dissimilar sequences (discontiguous megablast)", and "Somewhat similar sequences (blastn)". A "BLAST" button is present, along with "Search using Megablast (Optimize for highly similar sequences)" and "Show results in a new window" checkboxes. A red box on the right says "BLAST results will be displayed in a new format by default. You can always switch back to the Traditional Results page."

L'alignement multiple

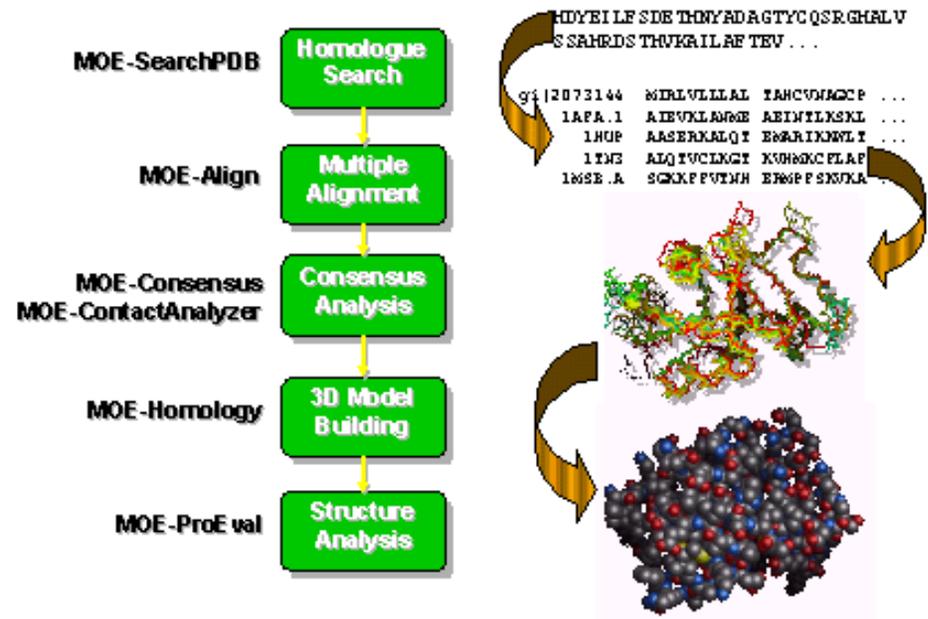
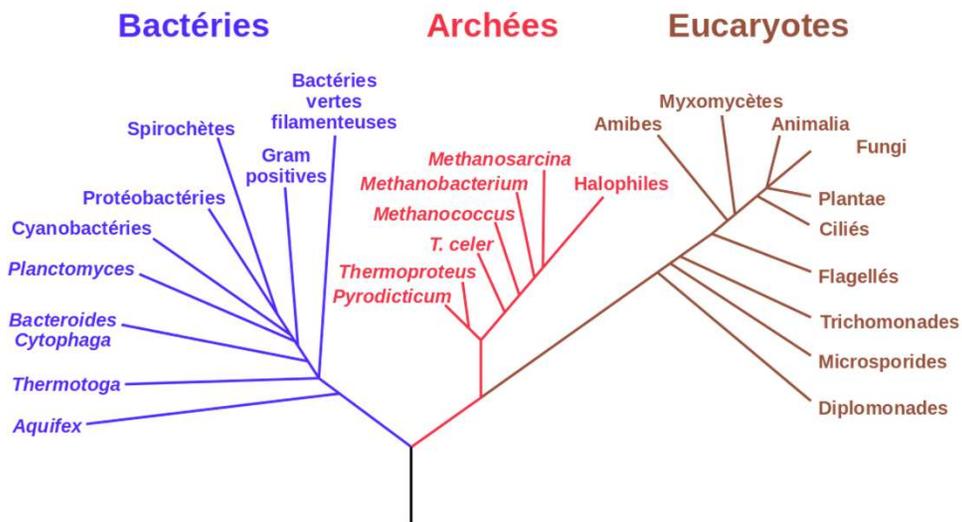
Principe : alignement portant sur plusieurs séquences à la fois et dans leur intégralité. Il permet de mettre en évidence les relations entre séquences que l'on ne peut pas visualiser en comparant les séquences 2 à 2.



Application

- caractérisation de régions conservées des protéines (motifs et domaines conservés).
- prédiction des structures 2D ou 3D par comparaison avec des séquences et structures connues
- construction des arbres phylogénétiques des séquences homologues.

Arbre phylogénétique de la vie



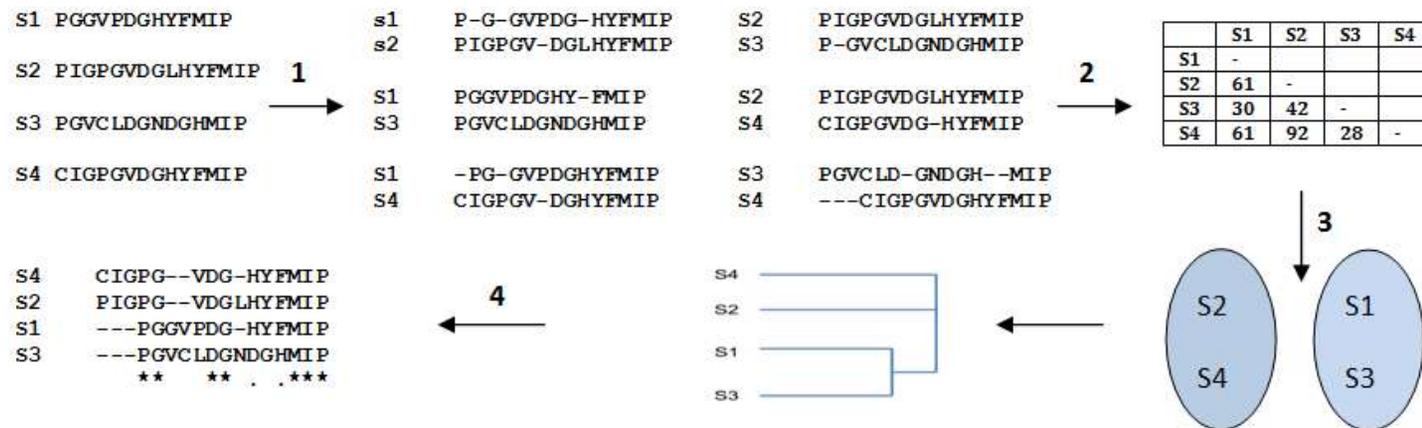


QUELQUES PROGRAMMES



CLUSTALW (Cluster Alignment). (Thompson, Higgins et Gibson, 1995)

Principe : CLUSTALW est fondé sur l'utilisation d'un algorithme d'alignement progressif. Les séquences les plus similaires sont alignées en premier puis l'alignement progresse vers les séquences les plus distantes. C'est également un programme de construction d'arbre phylogénétique.



Étape schématique de l'alignement multiple avec CLUSTALW avec 4 séquences d'acides aminés :

- 1-Alignement de toutes les séquences 2 à 2 et détermination des scores des alignements
- 2-Construction d'une matrice de score (BLOSUM62) pour l'ensemble des séquences
- 3-construction d'un arbre guide à partir de la matrice traduisant les relations globales entre les séquences
- 4-Alignement progressif à partir de l'alignement des 2 séquences les plus proches. les séquences voisines sont alignées de proche en proche jusqu'à l'alignement multiple final.

Légende : « * » : résidus conservés

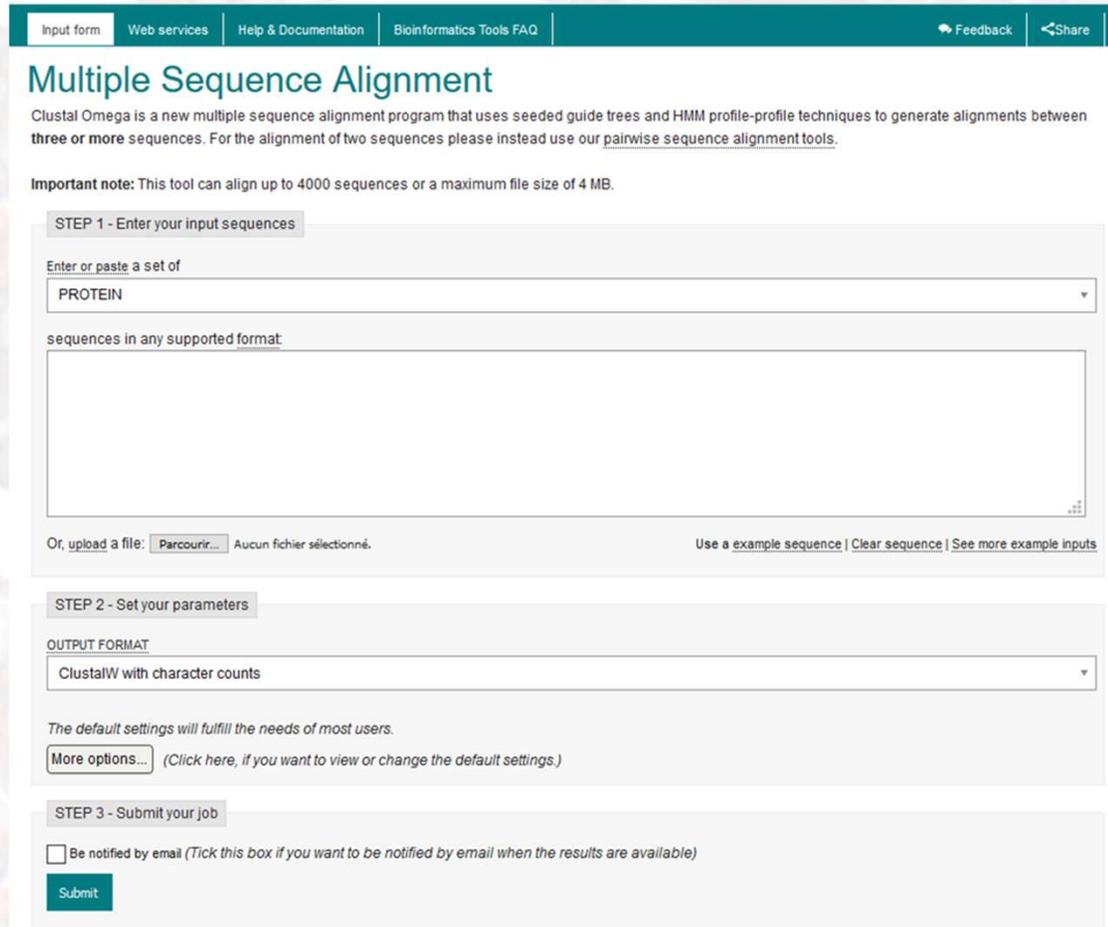
« . » substitution conservatives

CLUSTAL sur internet

sur internet vous trouvez la version
recente CLUSTAL Omiga
Comme montré ci-dessous suivant ce
lien :

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

Il est aussi intégré dans plusieurs logiciels
d'analyse de séquences comme MEGA X,
Bioedit, UGENE packages...



The screenshot shows the Clustal Omega web interface. At the top, there is a navigation bar with links for 'Input form', 'Web services', 'Help & Documentation', 'Bioinformatics Tools FAQ', 'Feedback', and 'Share'. The main heading is 'Multiple Sequence Alignment'. Below this, a brief description states: 'Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).' An 'Important note' specifies: 'This tool can align up to 4000 sequences or a maximum file size of 4 MB.' The interface is divided into three steps: 'STEP 1 - Enter your input sequences', 'STEP 2 - Set your parameters', and 'STEP 3 - Submit your job'. In Step 1, there is a dropdown menu for 'Enter or paste a set of' with 'PROTEIN' selected, and a large text area for 'sequences in any supported format'. Below this is an 'Or, upload a file:' section with a 'Parcourir...' button and the text 'Aucun fichier sélectionné.' There are also links for 'Use a example sequence', 'Clear sequence', and 'See more example inputs'. In Step 2, there is a dropdown for 'OUTPUT FORMAT' with 'ClustalW with character counts' selected. A note says 'The default settings will fulfill the needs of most users.' and there is a 'More options...' button with a link '(Click here, if you want to view or change the default settings.)'. In Step 3, there is a checkbox for 'Be notified by email (Tick this box if you want to be notified by email when the results are available)' and a 'Submit' button.



T-COFFEE (Noterdame, Higgins, Hering, 2000)

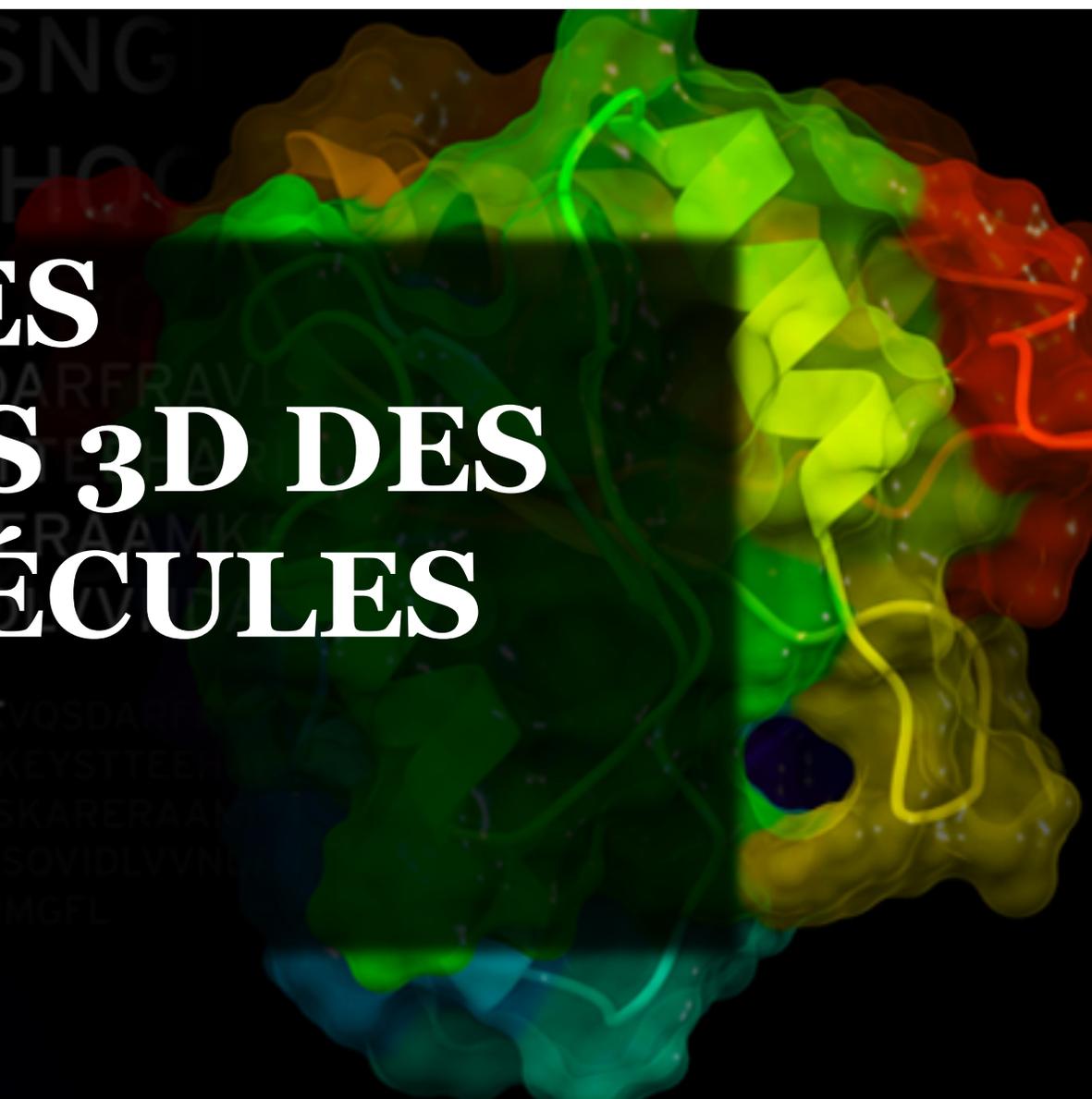
T-COFFEE est fondé également sur un alignement progressif. En plus de réaliser un alignement global entre chacune des paires de séquences, il procède à un alignement local afin d'optimiser l'alignement entre les séquences très divergentes.

Séquence logo

Représentation graphique de séquences alignées. La hauteur des caractères est proportionnelle à leur fréquence dans les séquences.

RCSB **PDB**
PROTEIN DATA BANK

**BANQUES DES
STRUCTURES 3D DES
MACROMOLÉCULES
'PDB'**



LA PDB
&
RCSB PDB

INTRODUCTION

La Protein Data Bank (PDB) est une archive des structures tridimensionnelles déterminées expérimentalement des macromolécules biologiques, y compris des protéines et des acides nucléiques.

Ce sont les molécules de vie que l'on retrouve dans tous les organismes, notamment **les bactéries, les levures, les plantes, les être humains**, et autres organismes.

- L'archive de PDB est disponible sans frais pour les utilisateurs (étudiants, chercheur, enseignants).
- Cette archive est mis à jour chaque semaine.

- La PDB a été établi en 1971 au Brookhaven National Laboratory, sous la direction de Walter Hamilton et contenait à l'origine 7 structures.
- le centre de données américain « The **Research Collaboratory for Structural Bioinformatics**” (**RCSB**) est **devenu responsable de la gestion** de la PDB.

En 2003

- le wwPDB a été formé afin de maintenir l'archive de PDB de données structurelles macromoléculaires qui est librement accessible au public et à la communauté mondiale.

La wwPDB comprend les quatre banques : PDBe, PDBj (Japon), BMRB « Biological Magnetic Resonance Data Bank” (USA) et RCSB-PDB .
un effort de collaboration entre ces banques pour assurer que l'archive de PDB soit uniforme et accessible au communauté scientifique.



Interface web de la PDB-RCSB

RCSB PDB Deposit Search Visualize Analyze Download Learn More

MyPDB Login



An Information Portal to
115031 Biological
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations



Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

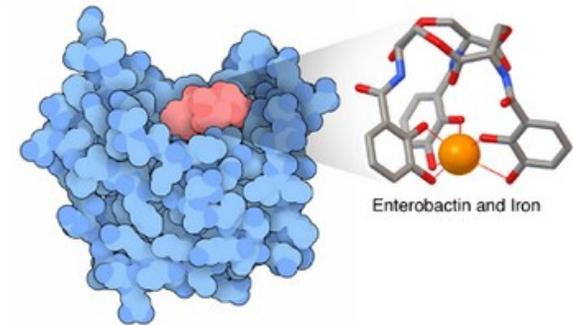
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Take an Interactive Tour of the PDB



December Molecule of the Month



Enterobactin and Iron

Siderocalin

Contact Us



STATISTIQUES SUR LA RCSB-PDB

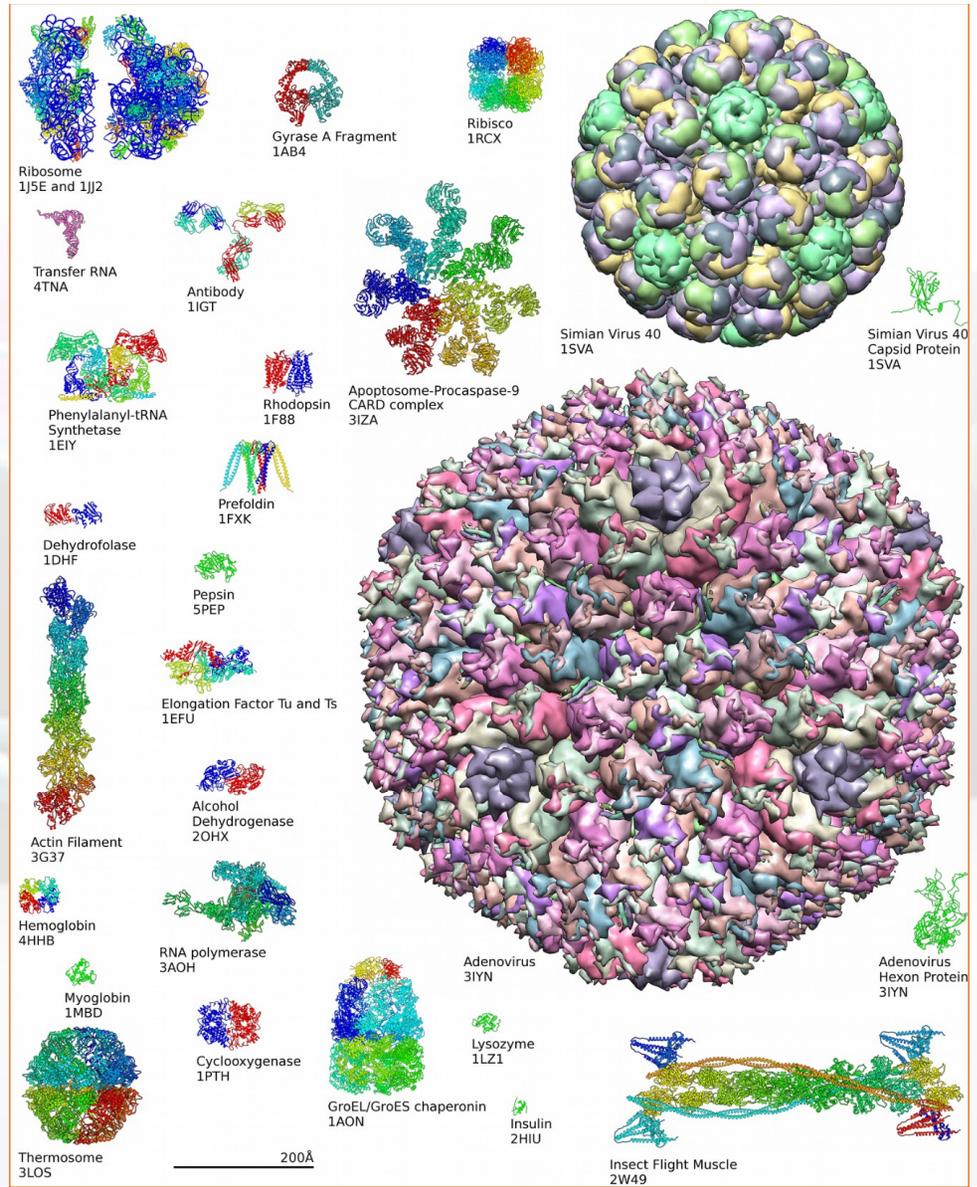
Tableau (2016) affiche les données moléculaires (différents type de macromolécules) contenues dans la Protein Data Bank énumérés par méthodes expérimentales.

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	<u>106453</u>	<u>1817</u>	<u>5467</u>	<u>4</u>	<u>113741</u>
NMR	<u>10287</u>	<u>1190</u>	<u>241</u>	<u>8</u>	<u>11726</u>
ELECTRON MICROSCOPY	<u>1011</u>	<u>30</u>	<u>367</u>	<u>0</u>	<u>1408</u>
HYBRID	<u>99</u>	<u>3</u>	<u>2</u>	<u>1</u>	<u>105</u>
other	<u>181</u>	<u>4</u>	<u>6</u>	<u>13</u>	<u>204</u>
Total	<u>118031</u>	<u>3044</u>	<u>6083</u>	<u>26</u>	<u>127184</u>

Tableau (2017) affiche les données moléculaires (différents type de macromolécules) actuelles contenues dans la Protein Data Bank énumérés par méthodes expérimentales.

Experimental Method	Proteins	Nucleic Acids	Protein/N A Complex	Other	Total
X-ray	114904	1905	5862	10	122681
NMR	10615	1232	247	8	12102
Electron Microscopy	1400	30	494	0	1924
Other	204	4	6	13	227
Multimethod	103	3	2	1	109
Total	127226	3174	6611	32	137043

Exemples de structures protéiques issues de la PDB.





Les entrées **PDB-RCSB**

Une entrée ou fichier PDB contiennent les informations qui vont permettre à des logiciels de visualisation moléculaire (ex : Pymol, RasTop ou Jmol) d'afficher les structures atomiques des molécules .

En générale il y a deux formats essentiels d'entrée PDB

1 format PDB

2 format mmCIF

format ou fichier PDB

```
!"#$%&'()*+,-./  
0123456789:;<=>?  
@ABCDEFGHIJKLMNO  
PQRSTUVWXYZ[\]^_  
`abcdefghijklmno  
pqrstuvwxyz{|}~
```

Un fichier au format PDB est un fichier texte composé de caractères **ASCII (American Standard Code for Information Interchange)**, caractérisant les structures tridimensionnelles des macromolécules disponibles dans la Protein Data Bank (PDB).

Les Structures des protéine déterminées se trouvent en coalition avec d'autres molécules tels que les acides nucléiques, de l'eau, des ions, des molécules de médicaments et ainsi de suite, qui peuvent donc être décrites dans le format PDB et ont été déposés dans la base de données PDB.

Il est possible d'accéder à l'information brute contenue dans ces fichiers en les ouvrant avec un éditeur de texte comme **Notepad++**.

Le fichier PDB est composé de lignes de 80 colonnes au maximum, soit 80 caractères.

Chaque colonne possède sa signification, par exemple les 6 premières colonnes, c'est-à-dire les 6 premiers caractères pour une ligne donnée, déterminent le champ.



Type de molécule, date de publication dans la banque, ainsi le code

Titre de fichier d'entrée de la molécule étudiée

Les différentes chaînes et molécules dans le fichier

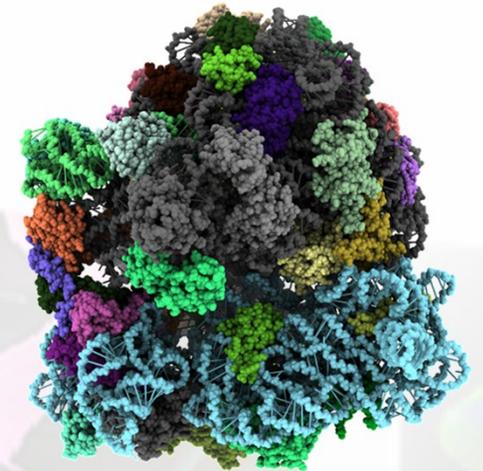
```
HEADER      HYDROLASE/DNA/RNA                25-JUL-12  4B3Q
TITLE       STRUCTURES OF HIV-1 RT AND RNA-DNA COMPLEX REVEAL A UNIQUE RT
TITLE       2 CONFORMATION AND SUBSTRATE INTERFACE
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: REVERSE TRANSCRIPTASE/RIBONUCLEASE H;
COMPND      3 CHAIN: A;
COMPND      4 SYNONYM: EXORIBONUCLEASE H, P66 RT, REVERSE TRANSCRIPTASE P66
COMPND      5 SUBUNIT;
COMPND      6 EC: 2.7.7.49, 2.7.7.7, 3.1.26.13, 3.1.13.2, 3.4.23.16;
COMPND      7 ENGINEERED: YES;
COMPND      8 MUTATION: YES;
COMPND      9 MOL_ID: 2;
COMPND     10 MOLECULE: P51 RT;
COMPND     11 CHAIN: B;
COMPND     12 SYNONYM: REVERSE TRANSCRIPTASE P51 SUBUNIT;
COMPND     13 ENGINEERED: YES;
COMPND     14 MUTATION: YES;
COMPND     15 MOL_ID: 3;
COMPND     16 MOLECULE: PRIMER DNA;
COMPND     17 CHAIN: D;
COMPND     18 MOL_ID: 4;
COMPND     19 MOLECULE: TEMPLATE RNA;
COMPND     20 CHAIN: R
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HUMAN IMMUNODEFICIENCY VIRUS 1;
SOURCE      3 ORGANISM_TAXID: 11676;
SOURCE      4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      5 EXPRESSION_SYSTEM_TAXID: 1007065;
SOURCE      6 EXPRESSION_SYSTEM_STRAIN: M15;
SOURCE      7 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE      8 EXPRESSION_SYSTEM_VECTOR: PDMI.1;
```

Format ou fichier CIF

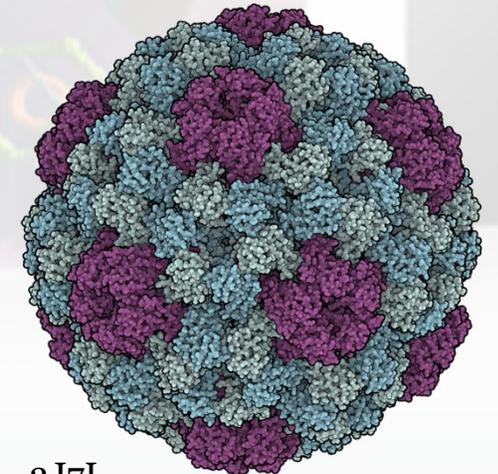
C'est un nouveau format textuel ajouté a coté de fichier PDB, CIF est l'extension de Fichier de l'information cristallographique ou (CIF) Crystallographic Information File.

Il été créé pour décrire les structures cristallographiques des molécules de faible poids moléculaire, et après pour décrire les structures des macromolécules plus compliquées sous forme d'un fichier **mmCIF**. (**m**acrom**m**olecular Crystallographic Information File).

En titre d'exemple les structure **des ribosomes (PDB code 5MDZ**
), des virus...



Structure of the 70S ribosome



3J7L
Full structure of bromo mosaic
virus



LE CODE PDB

PDB ID, de 4 caractères, identifiant unique, de chaque entrée dans la Protein Data Bank

Les 4 caractères PDB ID est attribué à chaque nouvelle structure au moment du dépôt dans la banque.

- Les identifiants sont automatiquement assignés et ne pas avoir un sens.
- Commence toujours par un chiffre Exp: **4HHB**
- C'est un identificateur unique, permanent de chaque entrée dans la Protein Data Bank., il est utilisé dans les littératures scientifiques (par exemple dans des articles de journaux et dans d'autres bases de données).

- Par conséquent, si le PDB ID d'une entrée dans la Protein Data Bank est connu, il est **le moyen le plus direct** pour récupérer la structure à partir de la banque de données.
- Un ou plusieurs ID PDB peuvent être saisis ou copiés et collés dans la boîte **de recherche avancée** et peuvent être séparés par des virgules ou des espaces blancs, y compris les sauts de ligne.

Choose a Query Type:

Quick Search

All/Experimental Type/Molecule Type

ID(s) and Keywords

- PDB ID(s)
- Entity ID(s)
- Chain ID(s)
- PubMed ID(s)
- UniProtKB Accession Number(s)
- Text Search
- mmCIF Keyword Search (Classification)
- Pfam Accession Number(s)
- UniProt Gene Name
- Sequence Cluster Name

Structure Annotation

- Structure Title
- Structure Description

Choose a Query Type:

Result Count

Add Search Criteria +

Retrieve only representatives at 90% sequence identity ?
Match all of the above conditions.

Clear All Parameters Submit Query

Download Learn More

MyPDB Login

Go

Advanced Search | Browse by Annotations | Search History (8) | Previous Results (0)

Journal of Biological Chemistry
Worldwide Protein Data Bank Foundation

f t y

Contact Us

Exemples:

Entrez **4HHB** dans la zone de texte à côté de « PDB ID (s) » et appuyez sur «GO». La page Résumé de Structure pour 4HHB va charger.

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB PROTEIN DATA BANK An Information Portal to 115031 Biological Macromolecular Structures

4HHB Go

Advanced Search | Browse by Annotations

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank NUCLEIC ACID DATABASE StructuralBiology Knowledgebase

Welcome Deposit

Search Options
Drill Down by Categories
Advanced Search
Sequences
Ligands

Explore the PDB Archive

Organism UniProt Molecule Name
Taxonomy Experimental

Select a Organism category below:
Homo sapiens (30470)

Structure Summary

3D View

Annotations

Sequence

Sequence Similarity

Structure Similarity

Experiment

Literature

Biological Assembly 1



View in 3D: JSmol or PV (in Browser)

Standalone Viewers

Simple Viewer Protein Workshop Ligand Explorer Kiosk Viewer

Protein Symmetry: Cyclic - C2 (View in 3D)

4HHB

THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION

DOI: 10.2210/pdb4hhb/pdb Entry 4HHB supersedes 1HHB

Classification: [OXYGEN TRANSPORT](#)

Deposited: 1984-03-07 Released: 1984-07-17

Deposition author(s): [Fermi, G.](#), [Perutz, M.F.](#)

Organism: [Homo sapiens](#)

Structural Biology Knowledgebase: 4HHB (2 models >17 annotations) [SBKB.org](#)

Display Files

Download Files

Experimental Data Snapshot

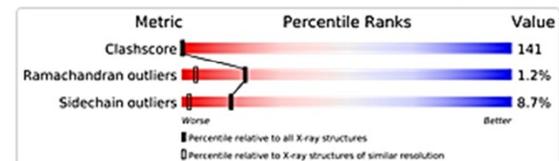
Method: X-RAY DIFFRACTION

Resolution: 1.74 Å

R-Value Work: 0.135

wwPDB Validation

Full Report



Literature

Download Primary Citation

The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution

Contact Us

The background features several semi-transparent panels displaying various molecular models. On the left, there is a ribbon diagram of a protein structure in orange and yellow. Below it, a space-filling model shows atoms in cyan, brown, and purple. In the center, a light blue ribbon structure is visible. On the right, a pink ribbon structure is shown, and below it, a green and yellow space-filling model. The overall theme is molecular biology and biochemistry.

TELECHARGER ET VISUALISER UN FICHER PDB

Telechargement d'un fichier PDB

Dans le coin en haut à droite du site il y a une barre de recherche similaire à l'image ci-dessous. Tapez le code de fichier par exemple "1HBB" et cliquez sur le bouton « GO ».

RCSB
PDB
PROTEIN DATA BANK
An Information Portal to
115031 Biological
Macromolecular Structures



[Advanced Search](#) | [Browse by Annotations](#) | [Search History \(7\)](#) | [Previous Results \(16\)](#)

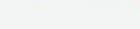
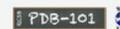


Cela devrait vous amènera à la page de « 1HBB.pdb ». Juste en dessous de la boîte de recherche sur le droit vous trouvez deux

Possibilités la 1^{ère} est **Download Files** et la 2^{ème} **Display files**

Cliquez sur « Download Files » et vous verrez un menu élargi similaire à l'image montrée sur la droite

RCSB
PDB
PROTEIN DATA BANK
An Information Portal to
115031 Biological
Macromolecular Structures

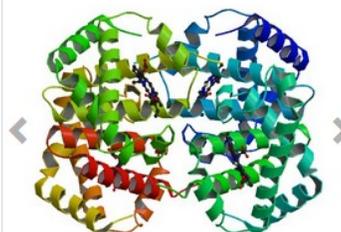


[Advanced Search](#) | [Browse by Annotations](#) | [Search History \(7\)](#) | [Previous Results \(16\)](#)



[Structure Summary](#) | [3D View](#) | [Annotations](#) | [Sequence](#) | [Sequence Similarity](#) | [Structure Similarity](#) | [Experiment](#) | [Literature](#)

Biological Assembly 1



1HBB

HIGH-RES

ARG: A MU

DOI: 10.2210/

Classification

Deposited: 1

Deposition a

Organism: [H](#)

Structural Bi

Experimenta

Method: X-R

Resolution:

R-Value Obs

[Display Files](#) | [Download Files](#)

FASTA Sequence

PDB File (Text)

PDB File (gz)

PDBx/mmCIF File

PDBx/mmCIF File (gz)

PDBML/XML File

PDBML/XML File (gz)

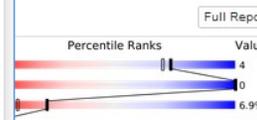
Structure Factor (Text)

Structure Factor (gz)

Biological Assembly (gz) (A+S)

N ROTHSCHILD 37BETA TRP->
ORIDE-BINDING SITE

[PDB.org](#)



Full Report

Fichier FASTA

>1HBB:A | PDBID | CHAIN | SEQUENCE

VLSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL
SALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR

>1HBB:B | PDBID | CHAIN | SEQUENCE

VHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDN
LKGT FATLSELHCDKLHVDPENFRLLGNVLCVLAHHEFGKEFTPPVQAAYQKVVAGVANALAHKYH

>1HBB:C | PDBID | CHAIN | SEQUENCE

VLSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL
SALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR

>1HBB:D | PDBID | CHAIN | SEQUENCE

VHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDN
LKGT FATLSELHCDKLHVDPENFRLLGNVLCVLAHHEFGKEFTPPVQAAYQKVVAGVANALAHKYH

Pour voir la structure 3D de PDB il y a deux possibilité

➤ Utilisé le logiciel de visualisation intégré dans la banque comme **Jmol** comme suit

Dans la page de la structure obtenue cliquez sur l'Onglet **3D View** il vous donne ce résultat suivant

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

Structure Summary **3D View** Annotations Sequence Sequence Similarity Structure Similarity Experiment Literature

1HBB
HIGH-RESOLUTION X-RAY STUDY OF DEOXYHEMOGLOBIN ROTHSCHILD 37BETA TRP-> ARG: A MUTATION THAT CREATES AN INTERSUBUNIT CHLORIDE-BINDING SITE

Display Files Download Files

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure. Help

Structure Details

Structure Biological Assembly
Symmetry Type Global Symmetry
Protein Symmetry C2
Protein Stoichiometry A2B2

Select Orientation
Front C2 axis

Select Display Mode
Secondary Structure Subunit
Symmetry

Display Options

Style Cartoon
Color Secondary Structure
Surface None

H-Bonds SS Bonds
 Rotation Black Background
 Polyhedron Axes

Fenêtre de visualisation de structure 3D

Fenêtres des paramètres pour changer les présentations de la structure les couleurs des structures secondaires.....

- On peut aussi faire des alignement entre des structure 3D ,Utilisant l'option d'analyse **Sequence & Structure Alignment** trouvée au niveau de l'onglet **Analyze** dans la page de structure

RCSB PDB Deposit Search Visualize **Analyze** Download Learn More MyPDB Login

PDB ID, author, macromolecule, sequence, or ligands Go

Search | Browse by Annotations | Search History (7) | Previous Results (16)

Sequence & Structure Alignment
Protein Symmetry
Structure Quality
Map Genomic Position to Protein
Third Party Tools

Analyze Options

Sequence & Structure Alignment

RCSB PDB's Comparison Tool calculates pairwise sequence (blast2seq, Needleman-Wunsch, and Smith-Waterman) and structure alignments (FATCAT, CE, Mammoth, TM-Align, TopMatch).

Comparisons can be made for any protein in the PDB archive and for [customized](#) or [local files](#) not in the PDB. Special features include support for both rigid-body and flexible alignments and detection of circular permutations.

Enter la 1^{ere} chaine ou structure

Enter la 2^{eme} chaine ou structure

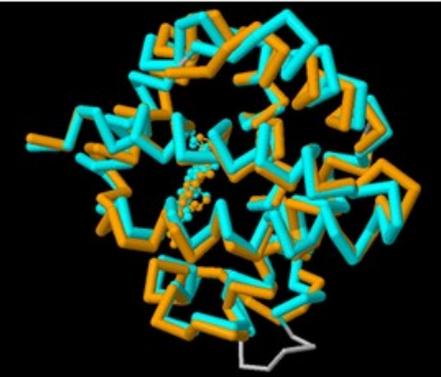
Enter First PDB ID ↔ Enter Second PDB ID

Select Associated Chain ID

...

Choisir le méthode de comparaison

- Select Comparison Method - Align More Options



Pour voir la structure 3D de PDB il y a deux possibilité

- Utilisé le logiciel de visualisation intégré dans la banque comme **Jmol** comme suit
- Ou utilisé un logiciel libre comme le Pymol, chimera, swisspdb.....

