**CHAPITRE VII : LA STRUCTURATION DE LA BIOINFORMATION : FICHIERS ET FORMATS**

1. **Traitement de l’information biologique en bioinformatique**

Le traitement bioinformatique des séquences biologiques peut être :

• **Simple** ...

> Composition

> Calcul de la Tm

> Traduction

> Carte de restriction

> Recherche de cadres de lecture ouverts (ORFs)

> ...

• **Complexe** ...

> Alignements

> Recherche et optimisation d’amorces et de sondes

> Prédiction de structures secondaires et tertiaires

> Recherche de motifs

> Construction d’arbres phylogénétiques

> Analyses de liaison (bi-points ou multi-points)

> ...

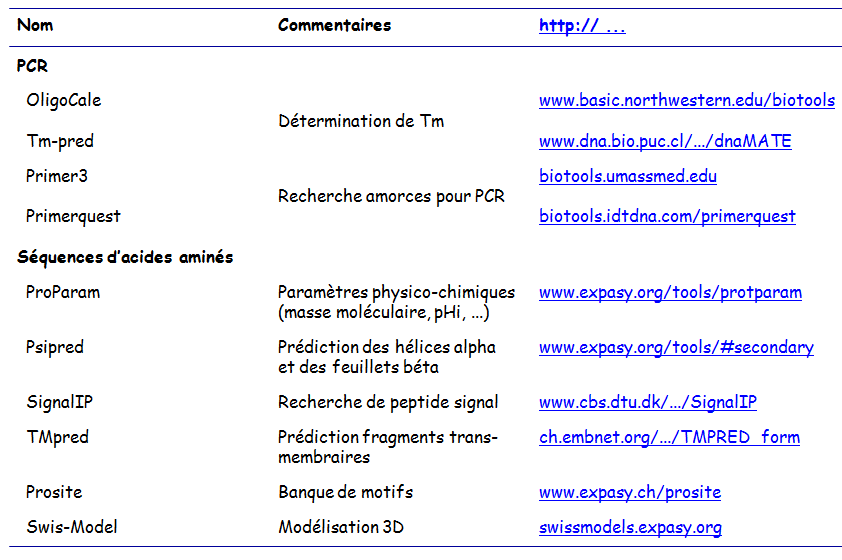
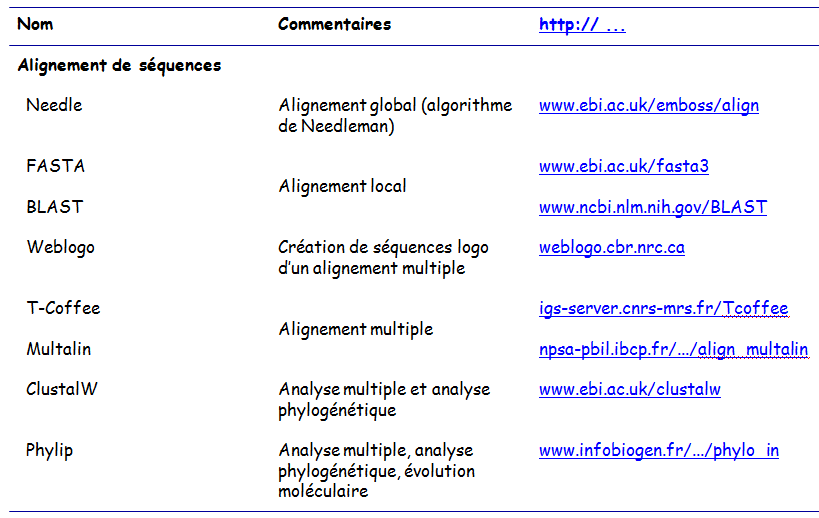
1. **Organisation de l’information bilogique**   
   - Chaque séquence (entrée) dans une base de données est sous forme de fichiers texte: Fichier plat (Flat file).  
   - Chaque fichier contient 2 parties principales:  
   1. Des informations relatives à la séquence (annotation)  
   2. La séquence elle-même

**III. Les différents fichiers en bioinformatique**

Les bases de données comportent des outils associés (logiciels) nécessaires pour l’accession à la banque.

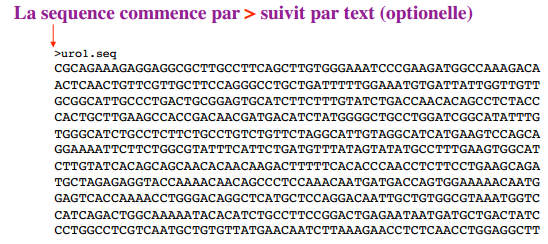
Chaque base de séquences possède son propre système d’interrogation.





**IV. Format des séquences**

La séquence est écrite dans un format FASTA, qui est universelle pour toutes les bases de donées et les logiciels pour l’analyse de séquences ADN et proteines.

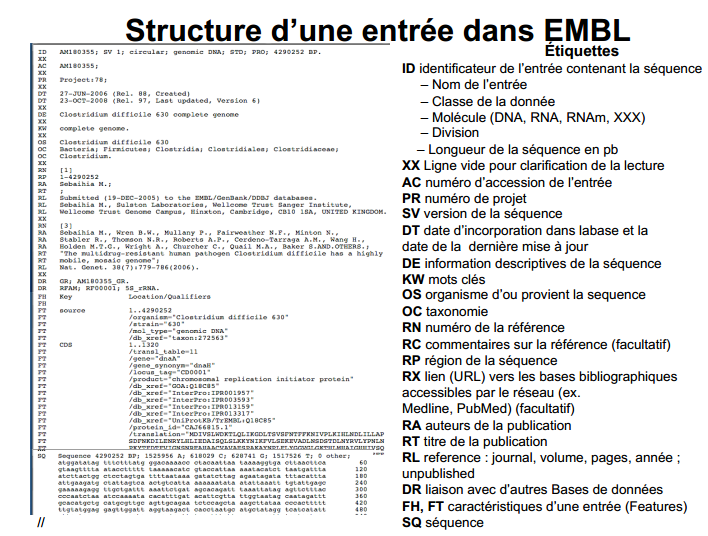


**Figure 28: Ecriture conventionnelle d’une séquence par format FASTA**

**V. Structure d’une entrée dans une Base de données**



**Figure 29: Structure d’une entrée dans une Base de données**





**Figure 30: Structure d’une entrée dans les Bases EMBL et GenBank**