Les motifs

Le motif est un terme utiliser en bioinformatique pour designer un ensemble de résidus soit 20 nucléotides ou 7 aminoacides qui sont retrouvés dans un ensemble des séquences alignées préalablement et présentent la similarité entre ces séquences. C'est un court segment continu ou non permettant de caractériser un ensemble de séquences nucléotidiques ou protéiques, non ambiguës et peu dégénérées.
Le motif spécifie généralement une fonction conserves au cours de l'évolution. Pour déterminer un motif dans un ensemble des séquences nucléiques ou protéiques on doit réaliser ce qu'on l'appelle un alignement multiple [Gérard C et al ,2006]. Un ensemble de différents motifs sépares par des régions variables compose un pattern. Une table de fréquence et une matrice de pondération des éléments qui composent le motif (Fig.1 exemple de motif protéique).

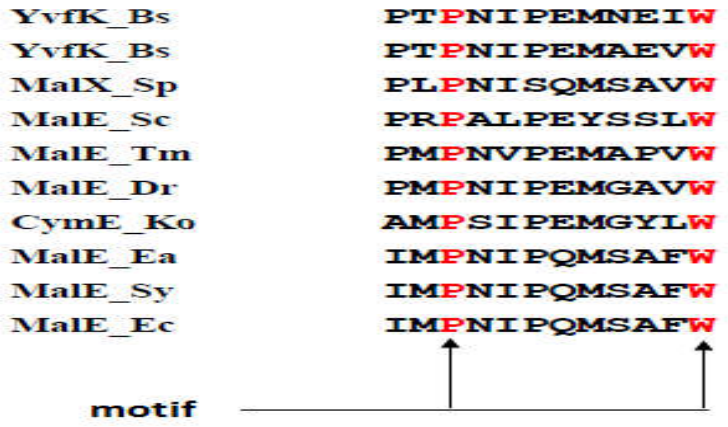

Figure 1 : représente un motif dans les séquences de Maltose Binding Protein

L'importance des motifs : la recherche des motifs est impliquée dans des systèmes de régulation ou définissent des fonctions biologiques.

- l'identification de régions codantes dans une séquence nucléique (par exemple en repérant les codons d'initiation et de terminaison...).

- la recherche d'un site de fixation de facteur de transcription dans une séquence.

- la recherche des sites (sites palindromique) de coupure d'une enzyme de restriction dans une séquence.

1.1 Types de motifs

1.1.1 Le motif nucléique :

Un motif nucléique est définis a partir de l'analyse d'un alignement multiple des séquences nucléotidiques connues .l'analyse de ces séquences permet de déterminer la variabilité en nucléotides dans chaque position de l'alignement.

L'alignement multiple des séquences permet de produire une séquence dite séquence consensus.

La séquence consensus : est une séquence de longueur n contenant, à chaque position, le symbole le plus fréquent à la même position dans l'alignement multiple. La séquence consensus est construite à partir de l'alphabet IUPAC ou en retenant une seule base, la plus fréquente, pour chacune des positions de la séquence, l'alphabet d'IUPAC représentée dans le tableau 1.

Tableau 1 ; l'alphabet IUPAC

A	**A**	**Adénine**
T	**T**	**Thymine**
C	**C**	**Cytosine**
G	**G**	**Guanine**
R	**A , G**	**Purines**
Y	**T , C**	**Pyrimidines**
W	**A T**	**Weak hydrogen bonding (2 liaisons)**
S	**C G**	**Strong hydrogen bonding (3 liaisons)**
M	**A C**	**Groupe amino (amine)**
K	**T G**	**Groupe keto (cétone)**
H	**A, C ou T**	**Non G**
B	**C, G ou T**	**Non A**
V	**A, C ou G**	**Non T**
D	**A, G ou T**	**Non C**
N	**A, G, C,T**	**N' importe nucléotide**