

---

## CHAPITRE IV. STRUCTURE DE PROTEINES

### 1. Introduction

Les protéines sont un des composants principaux de la cellule. En effet, les protéines, au même titre que les acides nucléiques, les glucides et les lipides, constituent l'essentiel de l'activité moléculaire de la cellule. Plus particulièrement, les protéines exécutent la grande majorité des fonctions de celle-ci parmi lesquelles, nous pouvons citer la transmission de signaux inter et intracellulaires ou le contrôle du passage de molécules au travers de la membrane. Les protéines sont, en outre, les constituants essentiels de la membrane cellulaire et peuvent également agir en temps qu'anticorps, hormones ou fibres élastiques. Au vu de leurs fonctions avancées, il n'est pas surprenant de constater à quel point ces molécules sont complexes. Cette section a pour but de présenter rapidement quelques éléments clés concernant les protéines.

### 2. Définition de la protéine

Une protéine est une collection de chaînes polypeptidiques non ramifiées d'acides aminés liés entre eux par des liaisons peptidiques. Ces protéines jouent un rôle essentiel dans la cohésion des structures morphologiques et dans le fonctionnement cellulaire. On citera quelques grands groupes de protéines :

- Les enzymes (catalyseurs biologiques, responsables de la plupart des réactions chimiques de la cellule).
- Les anticorps (responsables de la défense des organismes supérieurs, ils forment, dans le sang, des complexes avec les corps étrangers).
- Les protéines de stockage.
- Les protéines de transport.
- Les hormones (certaines hormones sont de nature protéiques).
- Les histones (liées à l'ADN, elles participent au contrôle de l'expression génétiques).
- Les protéines de structure et de soutien.

### 3. La structure de la protéine

Elles se composent d'un enchaînement linéaire d'acides aminés liés par des liaisons peptidiques. Cet enchaînement possède une organisation tridimensionnelle (ou repliement) qui lui est propre. De la séquence au repliement, il existe quatre niveaux de structuration de la protéine.

### 3.1 Structure primaire

Dans cette structure, la protéine prend la forme d'une chaîne constituée d'un groupe d'acides aminés liés par des liaisons peptidiques aux enzymes ribosomales.

### 3.2 Structure secondaire

La structure secondaire décrit le repliement local de la chaîne principale d'une protéine et il existe deux principales catégories de structures secondaires selon l'échafaudage de liaisons hydrogène, et donc selon le repliement des liaisons peptidiques : les hélices  $\alpha$ , les feuilletts  $\beta$ . Cette structure maintient sa cohésion au moyen de liaisons hydrogène qui se forment entre les deux groupes CO et NH.

### 3.3 Structure tertiaire

La série de peptides contenant des structures secondaires implique des régions interstitielles appelées zones d'inflexion, Constitue une sous-unité. Cette structure maintient sa cohérence par :

- Liaisons hydrogène entre les fonctions chimiques des racines R.
- Les associations entre les groupes négatifs et positifs dans les racines R.
- Attirer des racines avides d'eau.
- Les ponts de soufre produits entre deux racines de deux acides aminés Cys.

### 3.4 Structure quaternaire

Deux ou plusieurs séries ont une structure triangulaire et chaque série est appelée une unité. Sous les unités, lier ensemble avec des liaisons faibles telles que des liaisons hydrogène. ou plus rarement des liaisons covalentes (ponts disulfures). Et sa c'est dans le cas de protéines formées de plusieurs sous unités.

Les protéines assurent une variété étonnante de fonctions étroitement liées à leur structure moléculaire. La description des protéines se fait traditionnellement selon quatre niveaux d'organisation:

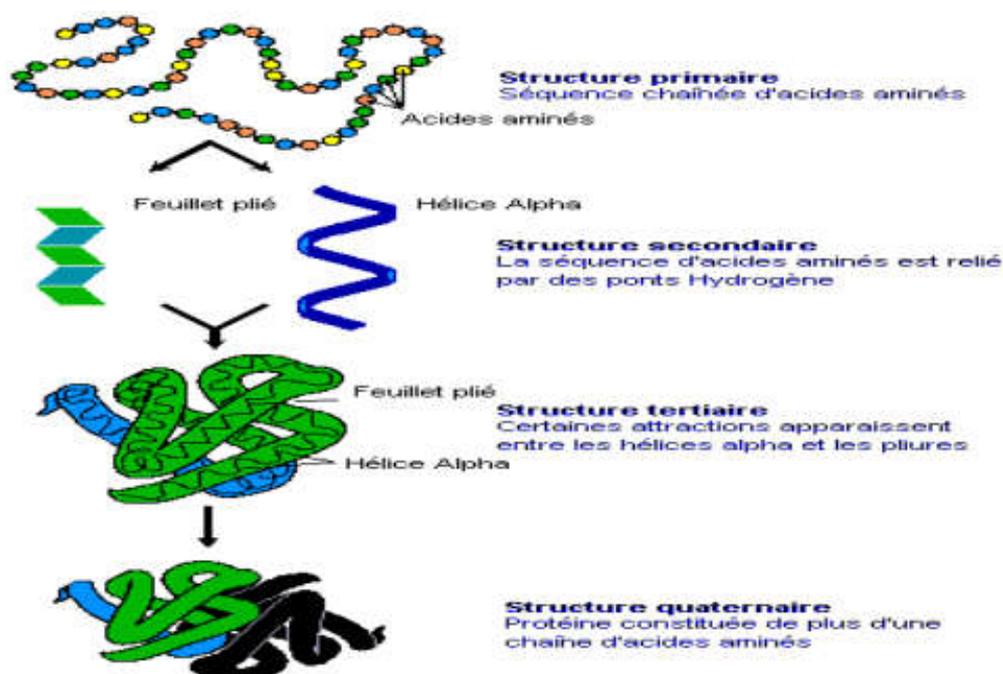


Figure 1 : Les quatre structures des protéines

#### 4. Interaction protéine

La plupart des protéines assurent leurs fonctions biologiques en interagissant avec une ou plusieurs autres protéines. Elles peuvent former de larges complexes protéiques tels que le protéasome, qui est un assemblage d'environ 50 sous-unités protéiques agissant ensemble pour dégrader d'autres protéines, et jouant un rôle primordial dans l'homéostasie. Ces interactions sont très diverses, selon leur composition, leurs affinités ou leur nature permanente ou transitoire.

##### 4.1. Type interaction

###### 4.4.1 Homo-oligomère ou hétéro-oligomère

Les interactions peuvent exister entre protéines identiques ou différentes, l'interaction peut avoir lieu sur une même surface pour les deux monomères (isologue), ou sur deux surfaces différentes (hétérologue), auquel cas la formation d'agrégats est possible.

###### 4.4.2 Obligatoire ou non-obligatoire

Les complexes formés peuvent être obligatoires ou non. Une interaction est obligatoire si les monomères impliqués n'ont pas de structure stable *in vivo* en l'absence de cette interaction. Dans ce cas, il est fréquent que la fonction des protéines impliquées soit dépendante de cette interaction. La plupart des complexes hétéro-oligomériques impliquent des interactions non-obligatoires, ce qui signifie que les protéines sont stables en l'absence d'interactions.

### 4.4.3 Permanente ou transitoire

On peut aussi distinguer les interactions selon leur dynamique. Les interactions permanentes sont très stables, et les protéines impliquées ne sont présentes que sous leur forme complexée. Les interactions transitoires sont beaucoup plus dynamiques, les partenaires s'associent et se dissocient rapidement *in vivo*. Les interactions transitoires peuvent être faibles, c'est à dire dépendantes de la concentration de chacun des partenaires dans le milieu.

L'interaction est contrôlée par un équilibre oligomérique dynamique en solution, dans laquelle les interactions se font et se défont continuellement. Les interactions transitoires peuvent aussi être fortes lorsqu'elles sont contrôlées par un mécanisme moléculaire. L'équilibre oligomérique dépend alors d'une phosphorylation ou d'une déphosphorylation, ou de la présence d'un autre partenaire tel que la guanosine triphosphate. Bien souvent, les interactions obligatoires sont aussi permanentes.

## 4.2. Mécanismes de régulation des protéines

Les interactions sont régulées par différents mécanismes. On identifie 3 types de contrôles :

### 4.2.1 La localisation

L'association de deux protéines dépend d'une rencontre des surfaces d'interaction, et requiert une co-localisation dans l'espace et le temps, c'est à dire une co-expression ou co-localisation dans un compartiment. Dans le cas où les protéines sont dans des compartiments différents, des transports dirigés entre les différentes localisations sont nécessaires.

### 4.2.2 La concentration locale

Ce paramètre est contrôlé par divers mécanismes, tels que : l'expression des gènes, les niveaux de sécrétion, la dégradation des protéines, le stockage temporaire, l'environnement moléculaire, et la diffusion ou viscosité du milieu.

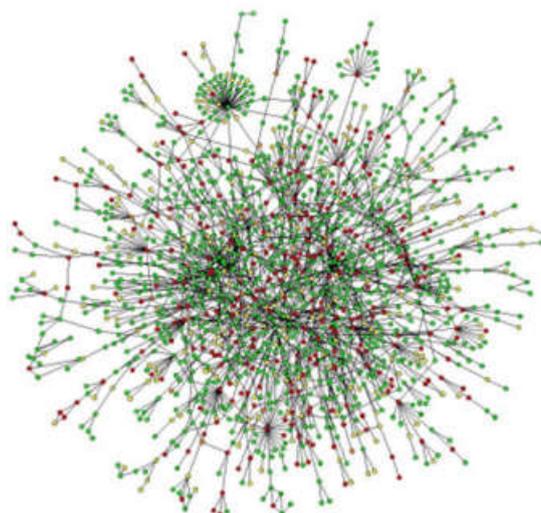
### 4.2.3 L'environnement physico-chimique local

Les affinités mutuelles des composants d'un complexe peuvent être modifiées par la présence d'une molécule effectrice (e.g. ATP, Ca<sup>2+</sup>) ou par un changement des conditions physiologiques.

## 5. Réseau d'interaction protéine-protéine

L'ensemble des interactions protéine-protéine ayant lieu dans un organisme, un organe ou un type cellulaire donné est appelé interactome. Un interactome est souvent représenté par un réseau, dans lequel les nœuds correspondent à des protéines, et où un arc entre deux nœuds signifie l'existence d'une interaction entre les deux protéines correspondantes. Ce réseau, appelé réseau d'interaction protéine-protéine, est non dirigé. Cependant, cette représentation

ne tient pas compte de la dynamique et de la localisation des interactions protéine-protéine dans la cellule.



**Figure 2: Visualisation d'un réseau d'interaction protéine-protéine chez la levure du boulanger.**

#### 6. Les bases de données d'interactions protéine-protéine

Les données d'interactions entre protéines sont publiquement disponibles dans différentes bases de données. Ces bases des données diffèrent par le type d'organisme couvert et les politiques d'acquisition des données. D'une part, les données d'interactions peuvent être succinctes ou détaillées, et d'autre part elles peuvent être récupérées de façon automatique ou entrées manuellement par une personne qui extrait ces informations de la littérature et des différents cribles effectués. Les principales d'entre elles sont:

**TABLE 1 : Description de quelques bases de données d'interactions**

Nom	Description	couverture organismes	Type de molécules
IntAct	Détaillée	Large	Tout
MINT	Détaillée	Large	Protéines
DIP	Détaillée	Large	Protéines
MatrixDB	Détaillée	Limité	protéines de la matrice extracellulaire
InnateDB	Détaillée	Limité	Protéines
MIPS	Détaillée	Mammifères	Protéines
MPACT	Détaillée	Levure	Protéines
BioGRID	Succincte	Limité	Protéines
HPRD	Succincte	Humain	Protéines
MPIDB	Succincte	Microbes	Protéines

## 7. Prédiction d'interactions

La prédiction des interactions permet d'augmenter le nombre d'interactions disponibles, notamment concernant les protéines peu ou pas étudiées. Les approches prédictives développées permettent de réduire l'espace des interactions à tester expérimentalement en proposant des interactions probables. De plus, ces méthodes prédictives permettent de proposer des interactions entre les protéines d'organismes peu étudiés, ou pour des systèmes inter-espèce dans lesquels très peu de données expérimentales d'interactions sont disponibles. La prédiction d'interactions a aussi l'avantage d'être applicable

La prédiction d'interactions a aussi l'avantage d'être applicable à grande échelle (génomique et protéomique) à peu de frais. Les méthodes de prédictions sont basées sur les séquences des protéines, et les caractéristiques structurales et génomiques liées aux interactions et aux relations fonctionnelles.

## 8. Méthodes de prédiction d'interaction

### 8.1 Méthodes de conservation du contexte génomique

L'analyse comparative des génomes, et en particulier de la conservation des contextes génomiques à travers les espèces, a permis de mettre en évidence des liens fonctionnels entre des gènes ou entre les protéines que ces derniers codent. Ces interactions fonctionnelles ne sont pas nécessairement des interactions physiques.

Différentes méthodes ont été comparées par Huynen et al. [Huynen et al., 2000]. et parmi ces méthodes : Transfert par interologues Cette méthode est basée sur le fait que deux protéines liées fonctionnellement ont tendance à co-évoluer. Si l'on considère deux protéines A et B en interaction dans un organisme donné, leurs orthologues A' et B' dans un autre organisme ont une forte probabilité d'interagir.

Cette méthode a permis de prédire des interactions chez l'homme, *C. elegans* et *D. melanogaster* à partir de données d'interaction de levure.

### 8.2 Méthodes de co-évolution

Les protéines qui interagissent physiquement évoluent en général de manière coordonnée, conservant ainsi les contacts entre elles [Pazos et al., 1997]. Ainsi, les méthodes basées sur ce principe sont susceptibles de prédire des interactions pas seulement fonctionnelles mais vraiment physiques. et parmi ces méthodes : Méthode des profils phylogénétiques Cette méthode n'utilise pas les interactions protéine-protéine déjà connues. Ce sont les relations d'homologie entre les protéines dans différents organismes qui sont utilisées.

Chaque protéine est représentée par un vecteur booléen dans lequel l'absence ou la présence d'un orthologue dans différents organismes est indiqué. Les protéines ayant des vecteurs similaires, donc des profils phylogénétiques proches sont identifiées et ont une forte

probabilité de participer à un même complexe ou une même voie de signalisation, et donc d'interagir physiquement.

### **8.3 Méthodes basées sur la structure**

D'autres méthodes utilisent des informations plus proches de la structure tridimensionnelle de la protéine, et cherchent d'abord à identifier les sites d'interaction. Ceci peut se faire par la reconnaissance de motifs de résidus, ou en utilisant les propriétés de la topologie de l'interface, de la surface accessible au solvant ou de l'hydrophobicité.

### **8.4 Méthodes basées sur les domaines**

Les domaines sont considérés comme les entités élémentaires de construction des protéines. Ce sont des unités structurales et/ou fonctionnelles qui sont conservées au cours de l'évolution. Chaque domaine contribue à la structure globale de la protéine et à ses différentes fonctions. L'hypothèse que les protéines interagissent entre elles par l'intermédiaire de leurs domaines est largement acceptée. L'idée est alors d'inférer des informations sur les interactions domaine-domaine en se basant sur les interactions protéine-protéine, puis de prédire des interactions protéine-protéine à partir des ces interactions domaine-domaine inférées. Mais ces méthodes sont évidemment très dépendantes de l'état de nos connaissances, car une grande partie de ces règles demeurent inconnues. C'est pourquoi on considère que les interactomes issues de ces approches sont encore peu fiables.

### **8.5 Méthodes d'apprentissage**

Différentes méthodes de classification avec apprentissage ont été utilisées pour prédire des interactions entre des protéines, ou entre des domaines. L'idée est d'apprendre les caractéristiques des paires de protéines en interaction, afin de prédire pour deux protéines quelconques si elles interagissent ou non, ou quelle est la probabilité qu'elles interagissent.