

PROTEINE ET SON INTERACTION:

1.1 Introduction

L'étude des interactions protéine-protéine (IPP) peut être très importante pour comprendre les fonctions cellulaires biologiques. Cependant, la découverte des IPP dans les laboratoires est longue et coûteuse. Pour cette raison, il y a eu beaucoup d'efforts récents pour développer des techniques de prédiction informatique des interactions protéine-protéine, car cela peut compléter les procédures de laboratoire et fournir un moyen peu coûteux de prédire le groupe d'interactions le plus probable sur toute la gamme de protéines.

L'objectif de ce chapitre est de présenter le contexte biologique relatif à cette thèse. Nous introduisons les notions de protéine et d'interaction protéine-protéine avec les méthodes de prédiction d'interaction protéine-protéine.

1.2 Définition de la Bioinformatique

La bioinformatique est une discipline émergente qui s'appuie sur les forces de l'informatique, les mathématiques et la technologie de l'information qui déterminerait l'analyse de l'information génétique. La bioinformatique tire partie des synergies entre sciences computationnelles et biologiques. [1] donc on peut dire aussi que la bioinformatique est l'application d'informatique sur la biologie.[2]

1.3 Application de la Bioinformatique

La Bioinformatique n'est pas limitée dans ses applications et en général peut être appliqué à toute recherche informatique pour résoudre des problèmes biologiques. Les applications courantes de la bioinformatique sont énumérées ci-dessous : [1]

- La bioinformatique des séquences, qui traite de l'analyse de données issues de l'information génétique contenue dans la séquence de l'ADN ou dans celle des protéines qu'il code. Cette branche s'intéresse en particulier à l'identification des ressemblances entre les séquences, à l'identification des gènes ou de régions biologiquement pertinentes dans l'ADN ou dans les protéines, en se basant sur l'enchaînement ou séquence de leurs composants élémentaires (nucléotides, acides aminés).

CHAPITRE 1:PROTEINE ET SON INTERACTION

- La bioinformatique structurale, qui traite de la reconstruction, de la prédiction ou de l'analyse de la structure 3D ou du repliement des macromolécules biologiques (protéines, acides nucléiques), au moyen d'outils informatiques.
- La bioinformatique des réseaux, qui s'intéresse aux interactions entre gènes, protéines, cellules, organismes, en essayant d'analyser et de modéliser les comportements collectifs d'ensembles de briques élémentaires du Vivant. Cette partie de la bioinformatique se nourrit en particulier des données issues de technologies d'analyse à haut débit comme la protéomique ou la transcriptomique pour analyser des flux génétiques ou métaboliques.
- La bioinformatique statistique et la bioinformatique des populations. [1]

1.4 Terminologies biologiques

1.4.1 ADN

L'ADN est une molécule très longue, composée d'une succession de nucléotides accrochés les uns aux autres par des liaisons phosphodiester. Il existe quatre nucléotides différents : A(adénosine), C(cytidine), G(guanosine), et T(thymidine), dont l'ordre d'enchaînement est très précis et correspond à l'information génétique. [30]

Un nucléotide d'ADN (Figure 1) a 3 composants :

- Un sucre (désoxyribose).
- Un composant d'acide phosphorique (phosphate).
- Une base d'azote (un des quatre types : Adénine ou Adénosine (A), Guanine (G), Cytosine (C) et Thymine (T)).

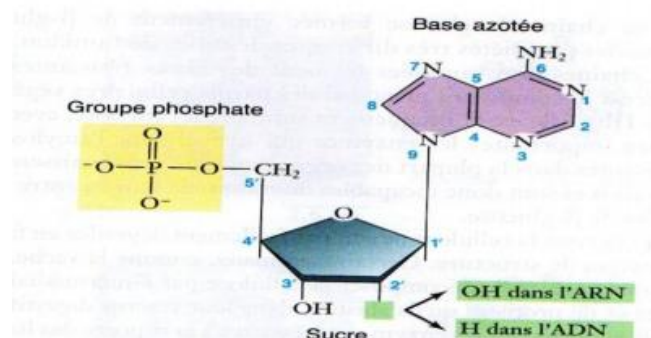


Figure 1.1 :- Structure du nucléotide (Raven et al., 2014).

CHAPITRE 1:PROTEINE ET SON INTERACTION

L'ADN est composé de deux brins se faisant face, et formant une double hélice. Ceci est rendu possible par la complémentarité des nucléotides qui peuvent interagir par des liaisons hydrogènes. Il y a deux liaisons hydrogènes entre **A** et **T** et trois entre **C** et **G**, ce qui conduit aux interactions possibles : A-T et T-A, d'une part et G-C et C-G, d'autre part. Les brins d'ADN sont orientés dans le sens 5' vers 3' (et ceci en raison de notations liées à la géométrie du désoxyribose).

Les deux brins d'une double hélice sont complémentaires et antiparallèles, c'est-à-dire assemblés tête bêche (l'extrémité 5' de l'un est en contact avec l'extrémité 3' de l'autre et inversement).

Nous pouvons également dire que l'ADN est la molécule portant l'information génétique dans le noyau. [4]

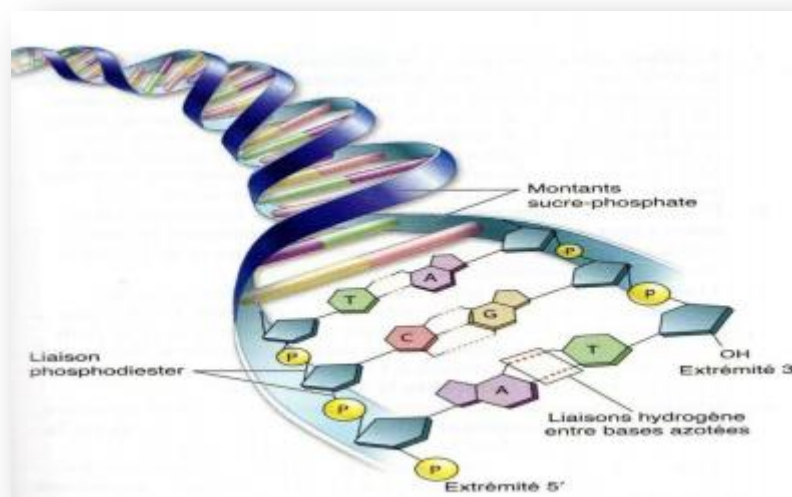


Figure 1.2 :- La structure de l'ADN (Raven et al., 2014) .

1.4.2 ARN et ARNm

ARN : L'acide ribonucléique (ARN) est l'acide nucléique obtenu à partir d'un des deux brins d'ADN. Produit au cours du processus de transcription (c'est-à-dire de l'ADN à l'ARN).

ARNM : est une molécule qui porte une séquence nucléotidique qui forme le code génétique sur lequel les protéines sont fabriquées. [29]

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.4.3 La reproduction

Est la première étape de la fabrication des protéines [4]. Sont réalisées à l'intérieur du noyau dans les cellules qui ont un noyau et ceux qui n'ont pas un noyau dans le cytoplasme et au cours de laquelle le traitement biomoléculaire de la molécule ARNm.

Éléments principaux du processus de reproduction :

- ADN.
- Enzyme ARN Polymères.
- Quatre types de nucléotides sont inclus dans la synthèse d'ARN.
- L'énergie d'ATP.

La reproduction passe par les étapes suivantes :

- **L'étape de démarrage** : L'enzyme ARN polymérase est associée au début du gène, élimine la circonférence et ouvre la série ADN après avoir cassé les liaisons hydrogène entre les paires de bases nucléotidiques. L'enzyme commence par lire la séquence des bases sur l'une des liaisons ADN de la chaîne clonée et reliant les noyaux correspondants pour former une chaîne d'ARN. Les nucléotides de l'acide nucléique sont comparés aux nucléotides en fonction de l'intégration des bases azotées.
- **L'étape d'allongement** : L'enzyme ARN polymérase se déplace le long du gène pour lire les informations sur la molécule d'ADN et lie les noyaux d'ARN en fonction de leur séquence dans la chaîne clonée de l'ADN, conduisant à l'allongement de la molécule d'ARN.
- **L'étape de fin** : L'enzyme atteint la fin du gène et l'élongation allongée de l'ARN est séparée et l'enzyme se décompose et la chaîne d'ADN se ferme.

CHAPITRE 1:PROTEINE ET SON INTERACTION

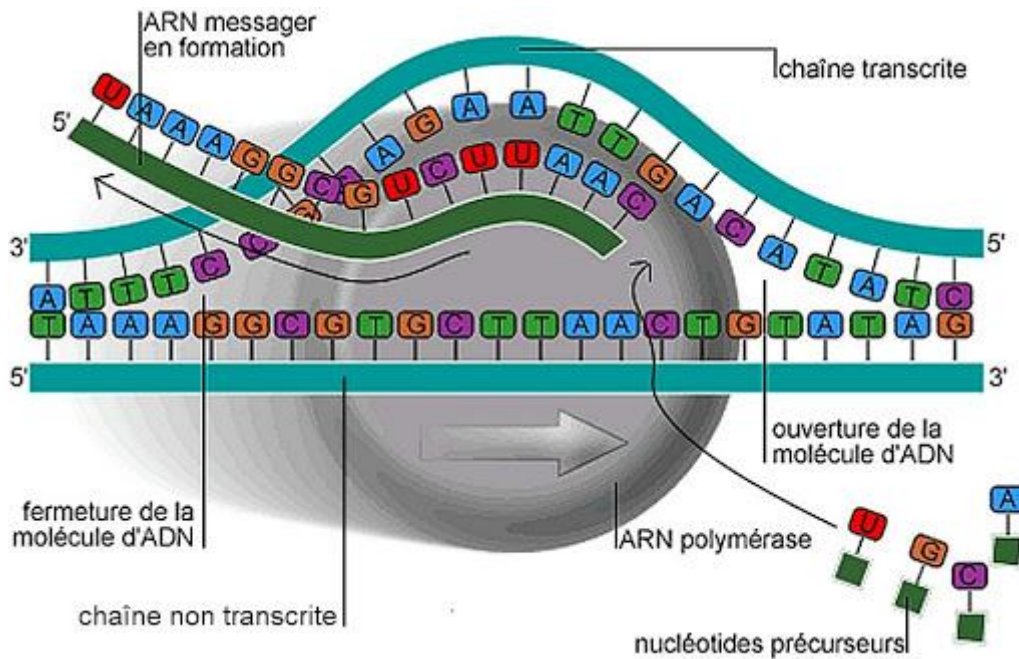


Figure 1.3:- la reproduction.

1.4.4 Les acides aminés

Il y'a 20 acide aminée, chaque acide aminée a La formule générale suivante :

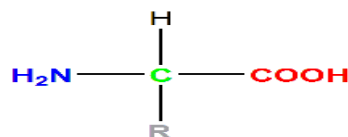


Figure 1.4 :- La structure d'acide amine.

Les acides aminés se distinguent donc par la nature de leur radical **R** plus communément appelé chaîne latérale. Ces dernières se distinguent par leur dimension, leur forme, leur charge, leur capacité de contracter des liaisons hydrogènes et leur réactivité chimique. [3]

Pour construire une protéine, il est nécessaire de se munir d'un mécanisme permettant de lier les acides aminés entre eux : la liaison peptidique. Son principe consiste à lier le groupe α -carboxyle d'un acide aminé à la fonction α -amide d'un autre acide aminé par une liaison amide. [3]

CHAPITRE 1:PROTEINE ET SON INTERACTION

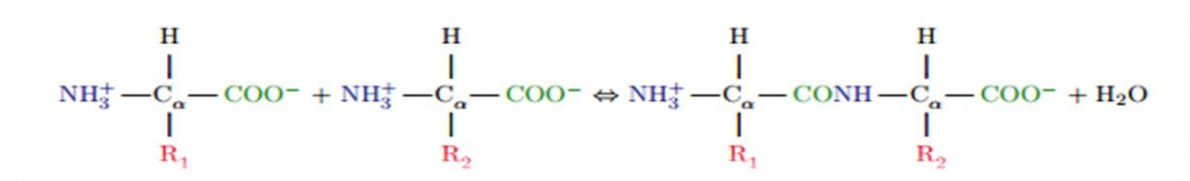


Figure 1.5 :- La structure des acides aminés liés.

1.4.5 Codon

Est une séquence de trois nucléotides sur un acide ribonucléiques messenger(ARNm) spécifiant l'un des 22 acides aminés protéinogènes dont la succession sur l'ARN messenger détermine la structure primaire de la protéine à synthétiser .Il existe quatre bases nucléiques A ,T,C,G de sorte qu'il existe $4^3=64$ codons différents, on distingue le codon de démarrage AUG et le codon d'arrêt UAG ,UGA,UAA.

1.4.6 La Traduction

La traduction est un processus permettant la synthèse d'une chaîne polypeptidique (protéine) à partir d'un brin d'ARN messenger (ARNm) et assure l'expression des gènes portés par l'ADN, et constitue la deuxième grande étape de ce processus après la transcription (qui consiste en la conversion de l'ADN en ARNm).

Acteurs de la traduction : La traduction nécessite un grand nombre d'acteurs dont les plus importants sont :

- L'ARNm.
- Le ribosome .
- L'ARN de transfert (ARNt).
- L'acide aminé.

La traduction se passe par les étapes suivantes :

- **L'étape de démarrage :** L'ARNm est placé sous la micro-unité et ARNt et l'acide aminé méthionine est placé dans la position du ribosome P. L'ARNt est connu comme le marqueur "AUG" aqueux dans l'ARNm à travers l'antioxydant. Ils sont connectés sous le grand corps pour former le complexe de départ. Le deuxième ARNt de l'acide

CHAPITRE 1:PROTEINE ET SON INTERACTION

aminé est situé sur le site A du ribosome selon le second brin de l'ARNm La liaison peptidique entre le premier acide.

- **L'étape d'allongement** : Le ribosome se déplace par un marqueur sur l'ARN m .Conduisant à la séparation de l'ARNt de l'acide aminé et du site P. L'emplacement du second ARNt portant le dipeptide change du site A au site P et le site A devient vide pour recevoir un nouvel ARNt portant un troisième acide aminé. Ainsi, les mêmes étapes sont répétées et la chaîne peptidique est absorbée par la quantité d'acide aminé.
- **L'étape de fin** : Le ribosome est l'un des arrêts sur l'ARNm .La chaîne peptidique résultante se décompose et le dernier ARNt est séparé et séparé sous les unités de ribosome. Le premier acide aminé de la chaîne peptidique est également implanté.

Les ribosomes peuvent répéter le cycle et former une autre série de peptides.

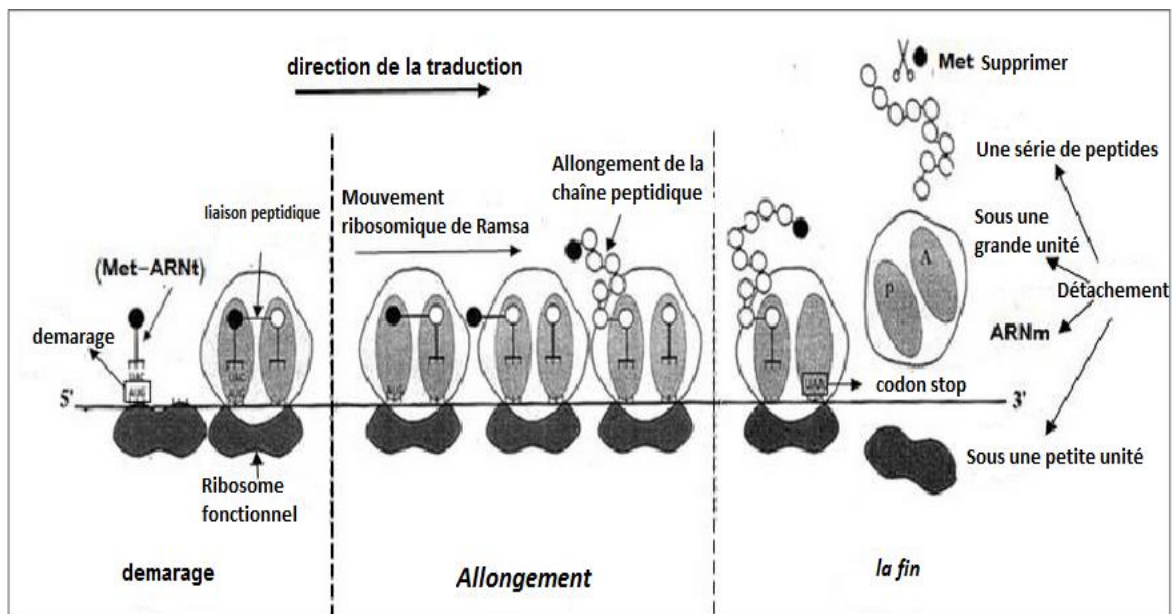


Figure 1.6 :- La traduction.

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.5 La protéine

1.5.1 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 pour mémoire, 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. [28]

1.5.2 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.

1.5.2.1 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.

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.5.2.2 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 α , les feuillets β .

Cette structure maintient sa cohésion au moyen de liaisons hydrogène qui se forment entre les deux groupes CO et NH.

1.5.2.3 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. [28]

1.5.2.4 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. [28]

CHAPITRE 1:PROTEINE ET SON INTERACTION

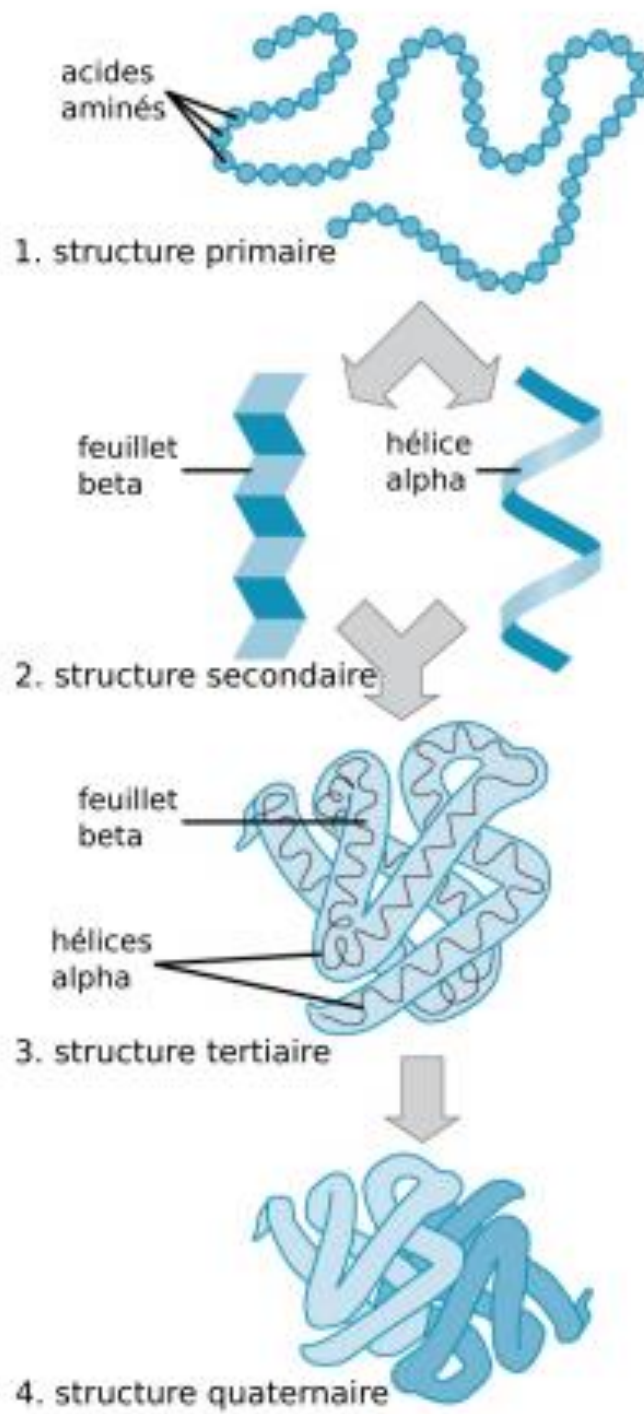


Figure 1.7 :- Les 4 structures des protéines. [5]

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.5.3 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 leurs composition, leurs affinités ou leur nature permanente ou transitoire. [24]

1.5.4 Type interaction

1.5.4.1 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. [5]

1.5.4.2 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. [5]

1.5.4.3 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 pas un équilibre oligomérique dynamique en solution, dans

CHAPITRE 1:PROTEINE ET SON INTERACTION

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. [5]

1.5.5 Mécanismes de régulation des protéines

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

1.5.5.1 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. [5]

1.5.5.2 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. [5]

1.5.5.3 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, Ca²⁺) ou par un changement des conditions physiologiques. [5]

1.5.6 Réseau d'interaction protéine-protéine

L'ensemble des interactions protéine-protéine ayant lieu dans un organisme, un organe

CHAPITRE 1:PROTEINE ET SON INTERACTION

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. [27]

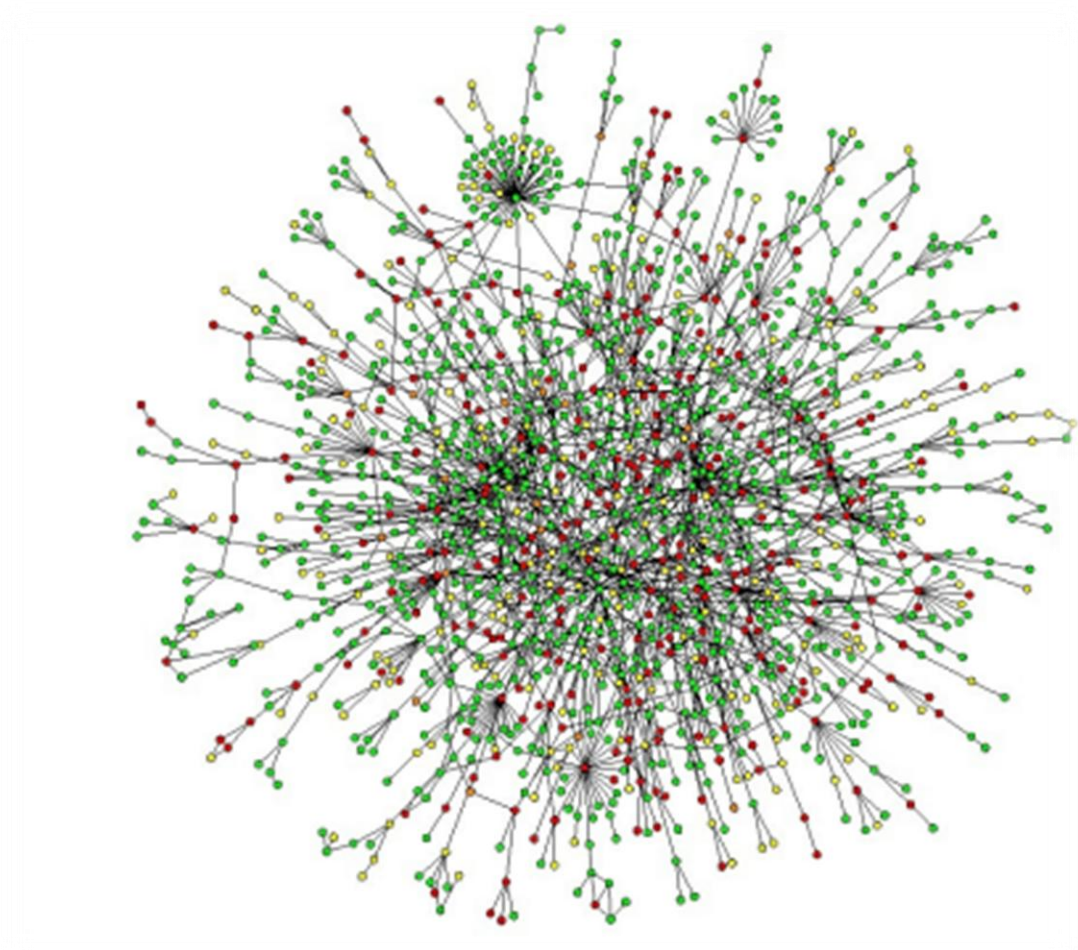


Figure 1.8 :- Visualisation d'un réseau d'interaction protéine-protéine chez la levure du boulanger. [Jeong et al., 2001].

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.5.7 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:

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

TABLE 1.1 :- Description de quelques bases de données d'interactions. [26]

1.5.8 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.

CHAPITRE 1:PROTEINE ET SON INTERACTION

La prédiction d'interactions a aussi l'avantage d'être applicable à grande échelle (génome et protéome) 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. [5]

1.5.9 Méthodes de prédiction d'interaction

1.5.9.1 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. [6] 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 [7]. 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]

1.5.9.2 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.

CHAPITRE 1:PROTEINE ET SON INTERACTION

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. [9]

1.5.9.3 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 [Kini et Evans, 1996], ou en utilisant les propriétés de la topologie de l'interface, de la surface accessible au solvant ou de l'hydrophobicité [Jones et Thornton, 1997]. [6]

1.5.9.4 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 de 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. [25]

1.5.9.5 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.

CHAPITRE 1:PROTEINE ET SON INTERACTION

1.6 Conclusion

Nous avons introduit dans ce chapitre les terminologies biologique et le contexte des interactions protéine-protéine. Et nous avons également parcouru ici un ensemble de méthodes de prédiction d'interaction .Dans le chapitre suivant, nous nous intéressons aux méthodes d'apprentissage qui peuvent être utilisées pour la prédiction d'interactions protéine-protéine.