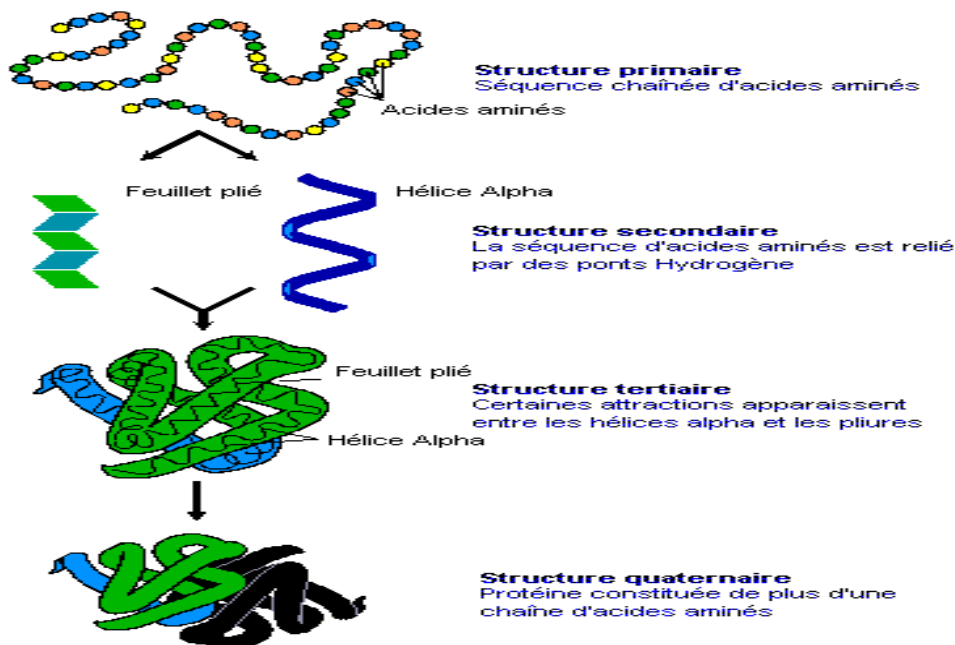


Méthodes d'études de la structure des protéines :

Généralités :

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:



I / Détermination de la structure primaire des protéines :

a) Séquençage de la protéine elle-même :

- méthode d'Edman
- spectrométrie de masse

b) Séquençage de l'ADN (ADNc) codant la protéine

Intérêts de la détermination de la séquence d'une protéine :

1. La séquence d'une protéine est son identité:

-indispensable pour comprendre son mécanisme d'action au niveau moléculaire et essentielle pour la détermination de la structure tridimensionnelle

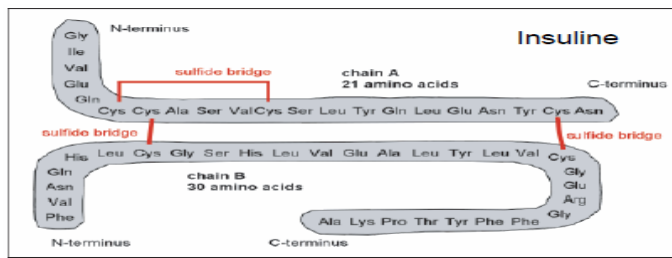
2. Permet d'identifier le gène

3. Comparaisons de séquences:

-applications cliniques car beaucoup de maladies héréditaires sont dues à des mutations qui modifient la nature d'un acide aminé dans une protéine

Séquençage des protéines :

Stratégie de base développée par Frederick Sanger en 1953 (Prix Nobel 1958).



Les liens disulfures doivent être coupés avant le séquençage pour séparer et déplier les sous-unités.

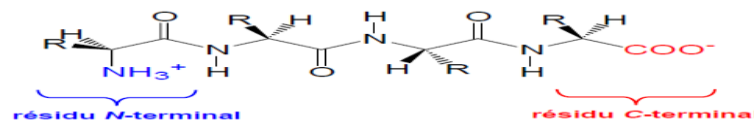
275

La première détermination de la séquence complète en acides aminés d'une protéine - l'insuline bovine par Fred Sanger en 1953. L'élucidation de la structure primaire a nécessité plus que 10 ans de travail et environ 100 g de protéine! Remarquez les ponts disulfure intra- et inter- caténaux

Étapes impliquées dans le séquençage d'une protéine:

- Détermination du *nombre de chaînes polypeptidiques* (sous-unités) dans la protéine

Puisque chaque chaîne polypeptidique possède un résidu *N*-terminal et *C*-terminal, le nombre de sous-unités distinctes dans une protéine peut être déterminé en identifiant le nombre de chacun des résidus terminaux.



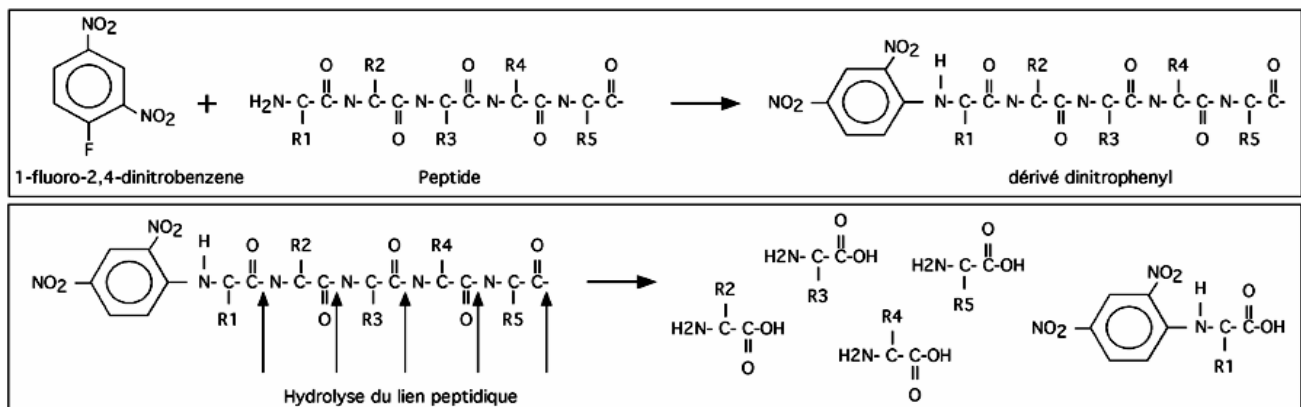
- Coupure des *liens disulfures* (inter- ou intramoléculaires)
- Le séquençage d'un peptide est un procédé classiquement décrit en six étapes :
 - 1- Détermination de la composition en acides aminés.
 - 2- Identification des bouts.
 - 3- Fragmentation de la chaîne polypeptidique.
 - 4- Séquençage des fragments.
 - 5- Seconde coupure de la chaîne polypeptidique avec un nouvel agent.
 - 6- Agencement des fragments obtenus et séquencés.
- Détermination de la structure de la protéine entière, incluant les liens disulfures entre les sous-unités.

1- Détermination de la composition en acides aminés : Cette étape a pour but d'évaluer la quantité de chaque acide aminé présent dans un peptide ou une protéine. On y parvient en brisant chaque lien peptidique de la protéine et en analysant ensuite le mélange obtenu par séparation chromatographique sur HPLC. Bien sûr, une fois qu'on a une idée de la nature et de la quantité des acides aminés contenus dans la protéine, il reste à en déterminer l'ordre.

Le nombre de chaque acide aminé dans une chaîne polypeptidique est un paramètre caractéristique de chaque protéine. Une protéine inconnue peut souvent être identifiée en

mesurant le pourcentage relatif des différents acides aminés et en comparant avec des bases de données.

2- Identification des bouts : Il existe des réactifs comme le 1-fluoro-2,4-dinitrobenzène ou le chlorure de dansyl qui réagissent avec la partie N-terminale des protéines et leur greffent un groupement supplémentaire. Alors que les liens peptidiques sont coupés par un traitement à l'acide, ces groupements supplémentaires restent greffés à au résidu situé le plus en 5' de la protéine; on peut alors identifier quel est l'acide aminé qui porte un appendice excédentaire.

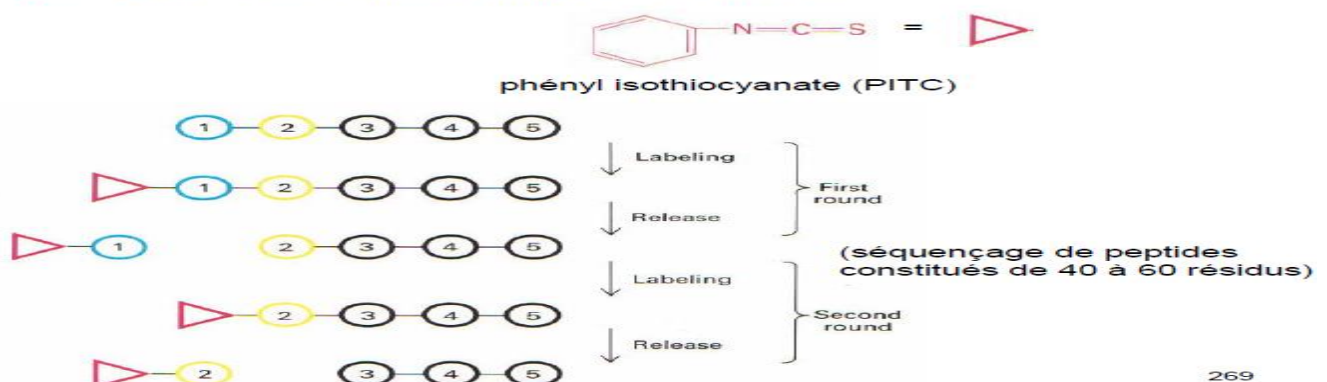


Pour l'extrémité C-terminale, on a recours à une protéolyse contrôlée avec la carboxypeptidase Y, C ou P (elles coupent le résidu situé le plus en 3'). Il faut alors faire attention à ne pas digérer plus d'un acide aminé à l'extrémité C-terminale.

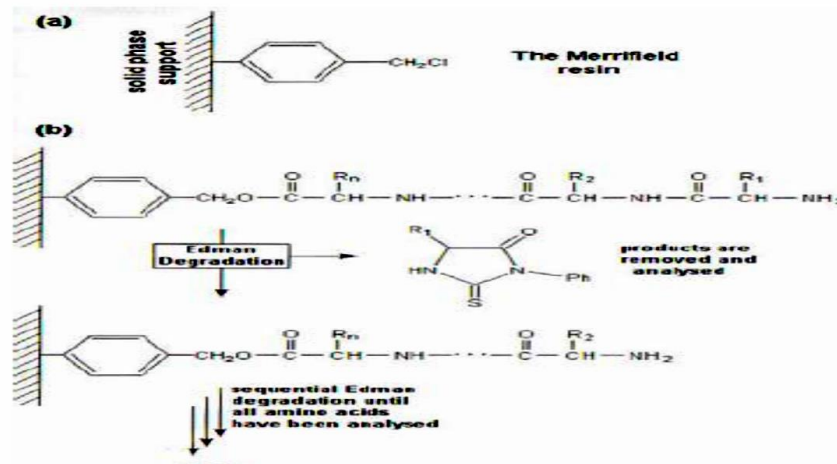
3- Fragmentation de la chaîne polypeptidique : Un traitement avec des agents chimiques ou des enzymes capables de couper la chaîne polypeptidique à des sites précis (la trypsine, par exemple, coupe après une lysine ou une arginine; la chymotrypsine après un résidu aromatique) permettra de réduire une grosse protéine en plus petits peptides. Ceux-ci pourront être séquencés séparément après séparation chromatographique.

4- Séquençage des fragments : La dégradation d'Edman permet d'enlever un par un les résidus en commençant par l'extrémité N-terminale. On arrive ainsi à déterminer la séquence de fragments d'environ 20 résidus.

❖ Dégradation d'Edman (analyse séquentielle):



Séquençage automatisé par dégradation d'Edman

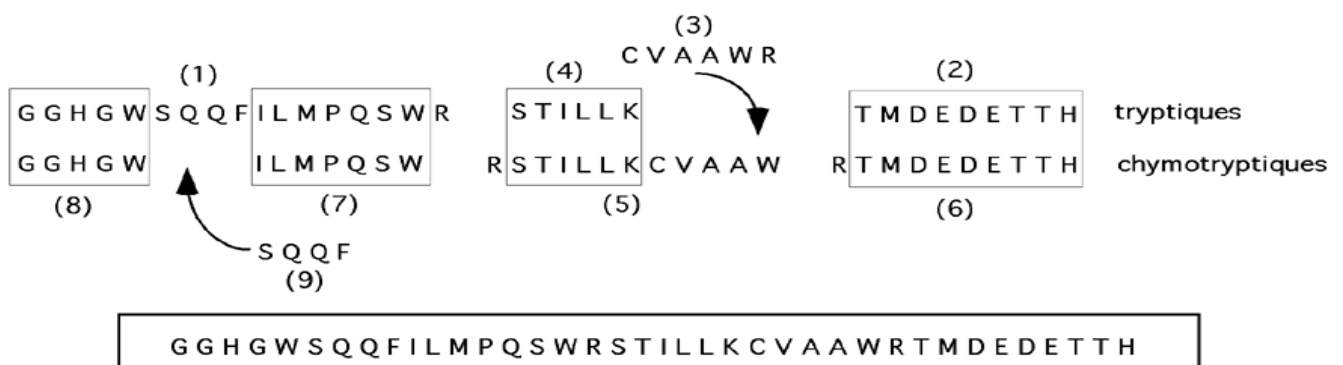


5- Seconde coupure de la chaîne polypeptidique avec un nouvel agent : Nous avons, en (3), coupé la chaîne avec un réactif ou un enzyme afin de pouvoir séquencer les fragments obtenus. Pour pouvoir enligner ces fragments, il nous faut maintenant recommencer (3) et (4) avec un nouvel échantillon de la protéine intacte et un agent différent. (Par exemple, si on a utilisé la trypsine en (3), on pourra maintenant utiliser la chymotrypsine).

6- Agencement des fragments obtenus et séquencés : Chaque fragment peptidique que nous avons séquencé se situe quelque part dans la séquence globale de la protéine. En comparant les fragments de la première digestion (en 3) à celle de la seconde digestion (en 5) et en ayant identifié les résidus des extrémités N- et C-terminale, on peut enligner toutes nos séquences par leur bout chevauchant.

- Fragments tryptiques**
- 1) GGHGWSQQFILMPQSWR
 - 2) TMDEDEETTH
 - 3) CVAAWR
 - 4) STILLK

- Fragments chymotryptiques**
- 5) RSTILLKCVAAW
 - 6) RTMDEDEETTH
 - 7) ILMPQSW
 - 8) GGHGW
 - 9) SQQF



Pour découvrir où se situent les ponts disulfures, on peut faire comme Sanger sur l'insuline: il a utilisé de l'acide per-formique, qui non seulement coupait les ponts disulfures mais en plus convertissait les chaînes latérales résultantes en -CH₂-SO₃H plutôt qu'en -CH₂-SH. Lors de la séparation pendant la séquence, on peut noter que ces cystéines modifiées ont une migration anormale.

Spectrométrie de masse :

En un mot, un spectromètre de masse c'est une balance. Il ne fonctionne pas du tout comme une balance, mais c'est tout de même ce qu'il est. Vous lui donnez un échantillon, et il va vous donner un poids, mais d'une manière vraiment précise, Pour une protéine de 40 000 Da, le spectromètre de masse peut vous indiquer un poids qui n'aura une marge d'erreur que de 4 Da!!! C'est amplement suffisant pour détecter le changement d'un acide aminé en un autre, ou encore la présence d'un groupement hydroxyle supplémentaire.

Mieux encore, si vous avez accès à une information génomique complète (comme c'est le cas pour de plus en plus d'organismes dont les génomes ont été séquencés) vous pouvez utiliser le spectromètre de masse pour identifier des protéines. Ainsi, si vous digérez la protéine XYZ (encore inconnue) à la trypsine et que vous en analysez les fragments par spectrométrie de masse, vous recevrez une liste de poids moléculaires. Un ordinateur peut facilement comparer les poids de cette liste avec les poids de toutes les séquences polypeptidiques codées par les cadres de lecture du génome qui nous intéressent, et identifier celles qui coïncident. Quand vous entendez un biochimiste dire qu'il a découpé une bande d'un gel SDS et l'a envoyé en *mass spec*, c'est tout juste ce qu'il a fait.

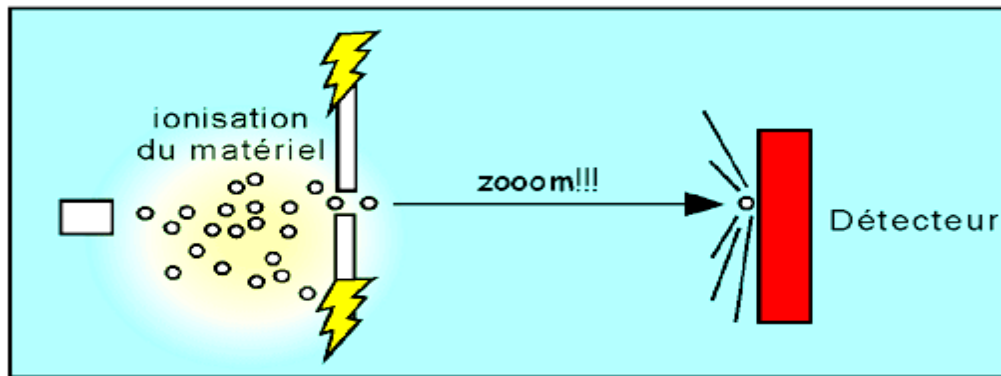
Le fractionnement du fragment en fragment plus petits permet d'identifier les résidus dans chacun d'entre eux, et l'alignement des résultats permet de déterminer l'ordre dans lequel on les retrouve dans le fragment initial.

Il est important de déterminer l'identité des acides aminés se trouvant aux extrémités d'un polypeptide, pendant sa séquence par MS/MS, pour permettre d'orienter l'assemblage obtenu en "pesant" chaque fragment: en effet, la séquence ABC donne les fragments suivants: A, B, C, AB, BC et ABC qui ont exactement le même poids que les fragments C, B, A, CB, BA et CBA qu'on obtiendrait avec le peptide CBA.

Il existe plusieurs types de spectromètres de masse. Ils portent des noms comprenant des sigles ésotériques comme TOF (time of flight), MALDI (matrix assisted laser desorption), ESI (electrospray ionisation), APCI (atmospheric pressure chemical ionisation), EI (electron impact), CI (chemical ionisation), FAB (fast atom bombardment), FD/FI (field desorption/ field ionisation), TSP (thermospray ionisation), Q (quadrupole), FT-ICR (Fourier transform-ion cyclotron resonance) etc.; ces sigles sont souvent combinés entre eux pour refléter la nature de

l'appareil ou ce à quoi ils sont connectés (LC-MS = liquid chromatography -mass spectrometer, etc).

Le principe général de la spectrométrie de masse repose sur une ionisation du matériel et au voyage du matériel chargé vers un détecteur. On distingue plusieurs systèmes de spectrométrie de masse selon la technique d'ionisation et la technique de détection.



Principe général de la spectrométrie de masse.

Avantages et inconvénients des différentes méthodes :

	<i>Avantages</i>	<i>Inconvénients</i>
❖ Edman	<ul style="list-style-type: none"> - Excellent pour 20 acides aminés standards - OK pour longs peptides 	<ul style="list-style-type: none"> - Lenteur/lourdeur/coût - Beaucoup modifications post-traductionnelles non détectées
❖ Spectrométrie de masse	<ul style="list-style-type: none"> - Rapidité relative - OK pour certaines modifications post-traductionnelles 	<ul style="list-style-type: none"> - Ambiguïté Ile/leu; Gln/Lys - Difficile avec longs peptides
❖ Séquençage de l'ADN	Rapidité	<ul style="list-style-type: none"> - Ignorance des modifications post-traductionnelles - Erreurs de phase de lecture

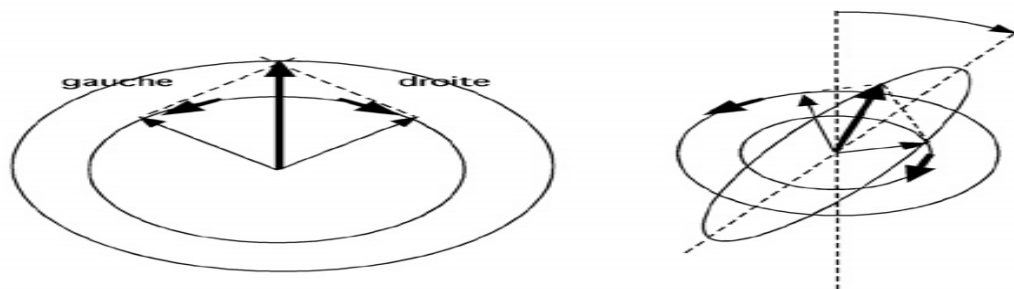
II/ Détermination de la structure secondaire :

Dichroïsme circulaire :

La spectroscopie par dichroïsme circulaire nous permet d'étudier la structure secondaire (hélices, feuillets, etc.) de polypeptides et de protéines en solution. Cette technique utilise peu de matériel et ne le détruit pas. Les changements dus aux conditions ambiantes de pH, d'agents dénaturants et de température peuvent y être facilement monitorés.

La technique repose sur la capacité qu'ont les structures optiquement actives d'absorber de façon inégale la lumière polarisée circulairement à droite de la lumière polarisée circulairement à gauche.

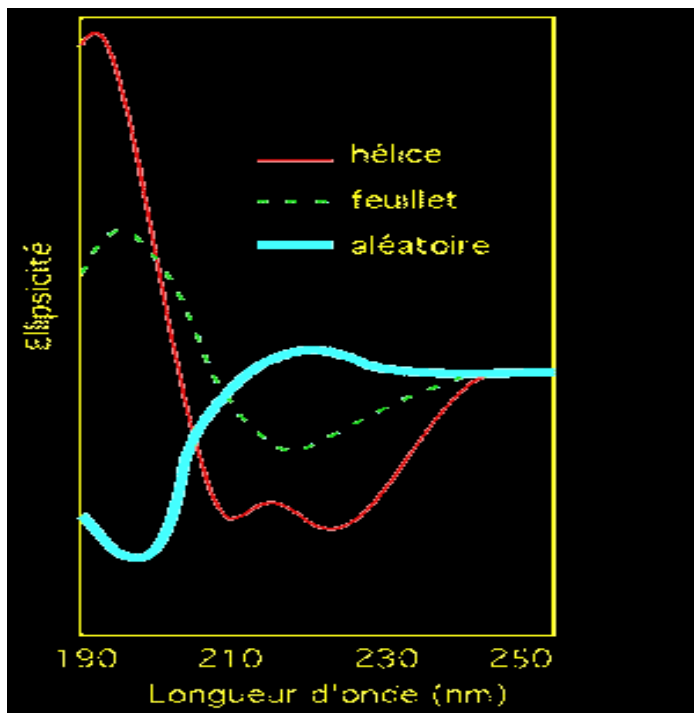
Les structures secondaires n'absorbent pas de façon égale la lumière polarisée circulairement vers la droite et la lumière polarisée circulairement vers la gauche. On raisonnera que l'absorption préférentielle de l'une de ces deux polarisations résultera en une déviation de la résultante (au lieu d'un cercle, le tracé de la résultante entre la spirale tournant à droite et la spirale tournant à gauche donnera une ellipse).



Cette déviation est appelée dichroïsme circulaire. Elle permet d'établir un spectre (appelé CD spectra en anglais) de forme caractéristique pour trois structures secondaires retrouvées dans les protéines: l'hélice, le feuillet et la forme aléatoire (ou random coil), qui chacune présentent des asymétries structurales.

Le spectre CD est particulièrement approprié pour juger rapidement l'état de repliement d'une protéine (ses hélices sont-elles bien hélicales comme prévu?); pour comparer la structure de protéines obtenues de sources différentes; pour analyser l'impact de mutations ponctuelles sur la structure d'une protéine; pour juger de la stabilité de la structure face à des changements environnementaux (pH, salinité); pour déterminer l'impact sur la structure de la présence d'un cofacteur comme un le zinc ou la magnésium.

On peut aussi suivre au fur et à mesure qu'il se déroule l'alignement d'une protéine au fur et à mesure qu'on la soumet à un traitement dénaturant.



La forme du spectre obtenu pour une protéine peut être décomposée en ses trois composantes :

- (a) hélicale,
- (b) en feuillet,
- (c) aléatoire.

Chaque composante donne une courbe spectrale caractéristique.

La dénaturation d'une protéine risque d'être irréversible. Le CD nous permet de la suivre "en direct", et de juger si elle est encore réversible. Trouver les conditions dans laquelle une protéine peut plus facilement retrouver son état premier après une dénaturation partielle peut nous être d'un grand secours pour la conserver longtemps.

III/ Détermination de la structure tridimensionnelle :

Pour être actives, elles adoptent un repliement au sein duquel les atomes s'organisent de façon spécifique dans l'espace. On parle de *structure tridimensionnelle* des protéines. Pour déterminer ces structures, il existe plusieurs méthodes qui diffèrent de par leur **résolution**. Deux techniques permettent d'atteindre une résolution atomique : la spectroscopie par résonance magnétique nucléaire **RMN** et la radiocristallographie.

1. Spectroscopie par Résonance Magnétique Nucléaire (RMN) :

La spectroscopie par résonance magnétique nucléaire (RMN) repose sur l'analyse des noyaux atomiques de la molécule.

L'environnement des noyaux influencera la façon dont ils se comportent quand on les irradie à la longueur d'onde appropriée ; la RMN permet donc de déterminer la nature des liens entourant un atome.

On peut faire transiter un noyau de l'état d'énergie faible à l'état d'énergie élevé en l'irradiant avec une fréquence donnée. Le retour du noyau à son état d'énergie antérieur est enregistré par un détecteur. Ce phénomène de transition d'un niveau d'énergie à un autre est la résonance qui donne le R de RMN.

L'apport de la RMN est essentiel non seulement pour déterminer la structure tridimensionnelle des protéines avec une résolution atomique, mais aussi pour élucider les mécanismes qui régissent l'assemblage dynamique de ces molécules.

On utilise la RMN pour étudier la structure de petites protéines (jusqu'à 30 KDa environ) en solution sans avoir besoin de les cristalliser. La RMN utilise le spin des particules subatomique pour nous renseigner sur la structure des protéines.

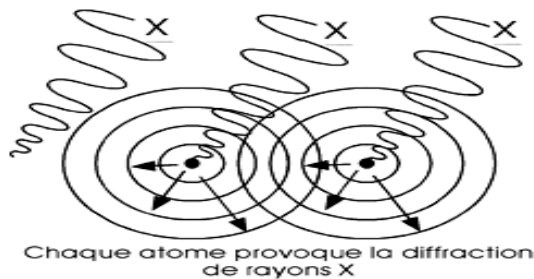
Le principe de la résonance magnétique nucléaire consiste à plonger une protéine en solution dans un **champ magnétique** intense (entre 10 et 20 **teslas**). Naturellement, seuls les atomes présentant un spin nucléaire non nul sont observables. Les protéines étant presque exclusivement composées d'atomes de carbone, d'oxygène, d'azote et d'**hydrogène**, les **protons** ont longtemps été les observables privilégiés de la RMN des protéines.

L'environnement électronique des protons module leurs fréquences de résonance qui varient de quelques parties par million (ppm) par rapport à une fréquence de référence. La valeur de la fréquence de résonance rapportée à une fréquence de référence et exprimée en ppm est appelée *déplacement chimique*. Celui-ci est fortement sensible au type d'atome lié au proton (carbone, oxygène, azote) (1), à son implication dans des liaisons non **covalentes** comme les liaisons hydrogène (2), à la proximité de noyaux aromatiques (3), ou encore d'un centre **paramagnétique**, **ion** métallique ou radical (4) contenant un **électron** non apparié. Ainsi, chaque proton résonne à un déplacement chimique qui le caractérise.

2. Diffraction aux rayons X :

Le principe de la détermination d'une protéine par cristallographie au rayon X est de déterminer la position de ses atomes afin d'en déduire un modèle structural. La position précise de chaque atome d'une molécule peut être déterminée seulement si la molécule est cristallisée (l'organisation des atomes forme des cristaux). Lorsque les rayons X frappent une molécule cristallisée, les électrons entourant chaque atome courbent ou diffractent le faisceau rayon X; ce phénomène permet de déduire un modèle de diffraction par rayon X par analyse de la densité des électrons. Un ordinateur interprète alors mathématiquement ce modèle et reconstruit la position des atomes pour finalement donner un modèle quasi exact de la structure de la molécule.

Le nombre d'électrons dans les atomes de la protéine détermine le degré de diffraction des rayons, et c'est pourquoi les atomes lourds seront plus efficaces pour faire dévier les rayons X.



Chaque atome peut induire la diffraction de rayons X incidents. Les rayons déviés se répartissent dans l'espace. Si plus d'un atome est présent, les rayons déviés par l'un et l'autre se feront de l'interférence; cette interférence peut être constructive quand deux pics, ou encore deux creux, se superposent. Elle peut aussi être destructive quand ce sont un pic et un creux qui se superposent.

La succession d'interférence constructive et destructive donne le graphique du milieu; vue en 2D elle donne la figure du bas, ci-contre.

En combinant le patron de diffraction d'un cristal obtenu à partir de différents angles, on obtient assez d'information pour nous aider à déterminer l'agencement des atomes de la protéine dans l'espace.

La cristallisation demande une grande quantité de protéine pure.

Comparaison entre RMN et cristallographie :

La RMN appliquée aux macromolécules biologiques se distingue de la cristallographie par plusieurs points forts :

- les mesures de structures ne sont pas dommageables pour les échantillons protéiques et elles sont réalisées en solution, alors que la radiocristallographie des protéines nécessite au préalable de disposer d'un cristal de protéine de très bonne qualité.
- La RMN permet également de mettre en évidence les fluctuations conformationnelles des protéines, qui jouent souvent un rôle majeur dans la reconnaissance entre molécules ou dans le fonctionnement intime et dynamique des enzymes.

La banque de structures protéiques PDB

Protein Data Bank est la banque de référence des structures protéiques obtenues expérimentalement par cristallographie rayon X ou spectroscopie RMN. Jusqu'en mai 2003, 21000 structures protéiques ont été entrées dans la banque ; de 50 à 100 nouvelles structures sont déposées chaque semaine.

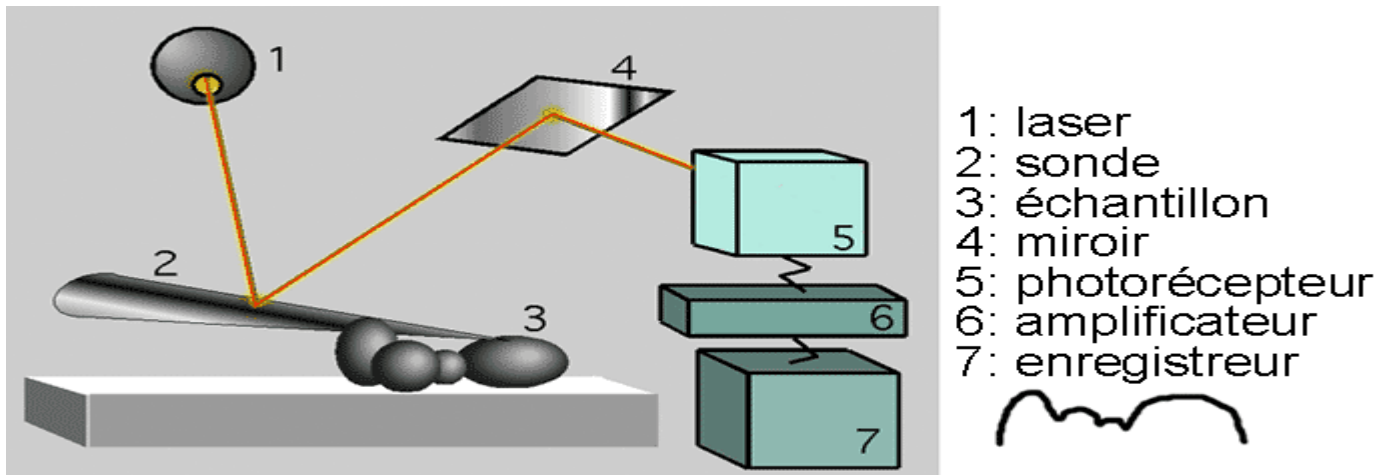
Les coordonnées des atomes formant la structure d'une protéine, le détail de la séquence, les conditions de cristallisation sont les principales informations disponibles pour chaque structure de la banque PDB.

C'est à partir de cette banque que sont détectés les homologues structuraux. La majorité des séquences ont une homologie structurale inférieure à 20% ; on évalué à environ 1000 le nombre total de structures protéiques originales qui suffirait à modéliser la quasi-totalité des protéines connues.

3. Microscopie de force atomique (AFM) :

Cette technique non-destructrice permet de relever des détails topographiques à très petite échelle. Elle détecte les formes des molécules un peu comme vous le faites en tâtant une surface

avec un doigt. Dans ce cas-ci, le doigt est remplacé par un levier extrêmement ténu (la sonde). La position de cette sonde est suivie par la déviation du tracé d'un rayon laser qui rebondit en permanence sur sa surface réfléchissante et se rend jusqu'à un photodétecteur. Tout mouvement de la sonde quand elle grimpe par dessus un obstacle ou descend dans une cavité se traduit par une déviation du rayon, déviation qui est notée par le photodétecteur.



La microscopie de force atomique a permis de visualiser la structure de la chromatine, la position de polymérase sur l'ADN, et de nombreuses autres structures ne relevant pas toujours de la biologie.

Ce [lien](#) mène à de très belles images obtenues par AFM. On peut y suivre la condensation croissante d'un fragment de chromatine en fonction de la concentration en sel dans le milieu ambiant.

Les avantages de l'AFM sont nombreux. Elle ne requiert pas la destruction de l'échantillon, ne demande pas de coloration et donne une excellente idée de la forme de l'échantillon.