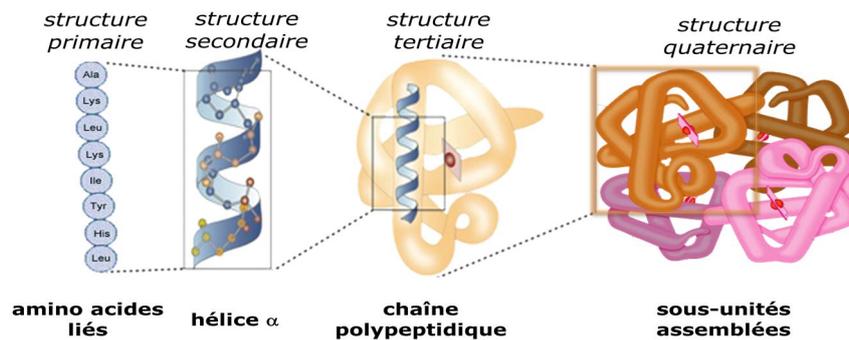
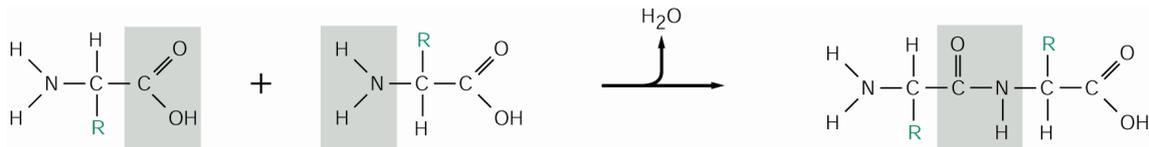


Protéines

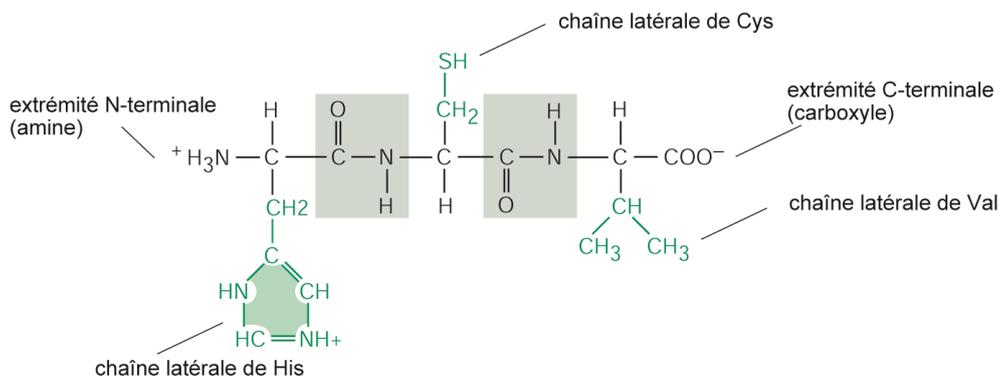


I. La liaison peptidique

La présence des fonctions amine et carboxyle sur le carbone α d'un acide aminé favorise les **réactions intermoléculaires** qui conduisent à la formation de structures polypeptidiques (peptides et protéines), donnant une certaine rigidité au squelette ainsi formé. Les petits peptides restent assez flexibles. Les grosses structures protéiques sont plus rigides. D'un point de vue chimique, la condensation d'un acide carboxylique et d'une amine engendre une fonction amide, CO-NH :

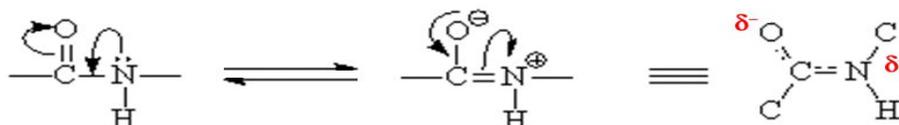


La liaison covalente entre le C et le N de la fonction amide reliant deux acides aminés est appelée **liaison peptidique**. Peptides et protéines sont donc constitués de l'enchaînement linéaire d'acides aminés liés entre eux par des liaisons peptidiques. Ainsi, dans l'exemple ci-dessous, la liaison des acides aminés His, Cys et Val conduit à la formation du tripeptide His-Cys-Val.

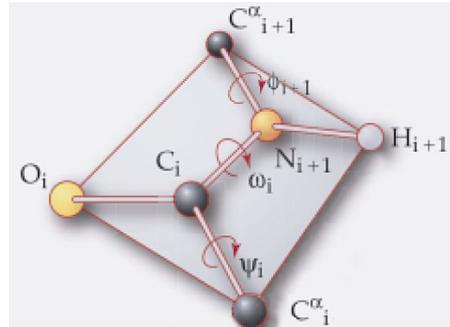


Les extrémités d'une chaîne polypeptidique sont dissymétriques. L'une correspond à une fonction α -amine libre, l'autre à une fonction α -acide carboxylique libre. Elles sont respectivement appelées extrémités N- et C-terminales de la chaîne. Par convention, on positionne à gauche l'extrémité N-terminale et à droite l'extrémité C-terminale (tant dans la structure chimique développée que dans la désignation de la succession des acides aminés de la séquence, ici His-Cys-Val). Aux pH physiologiques, les extrémités N- et C-terminales sont toutes deux ionisées.

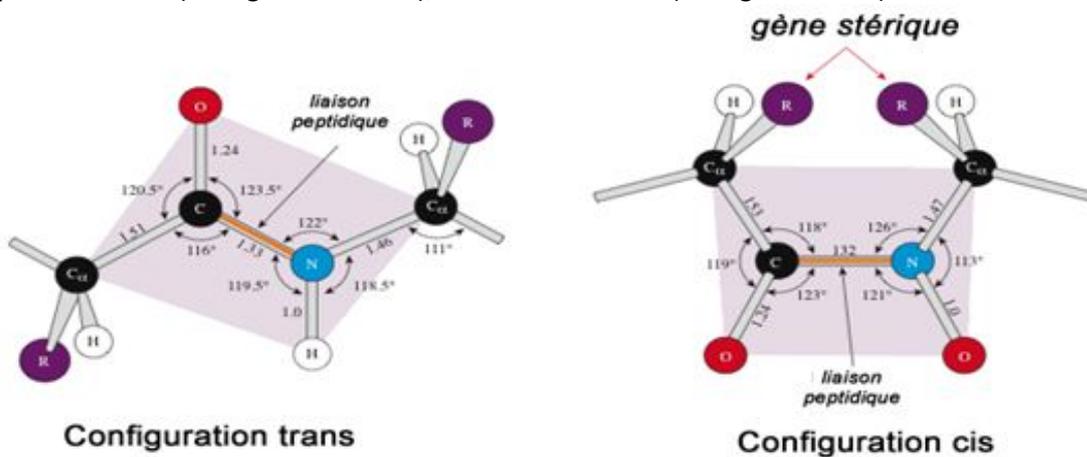
La fonction amide est stabilisée par mésomérie (résonance), ce qui confère à la liaison peptidique un caractère partiel de double liaison :



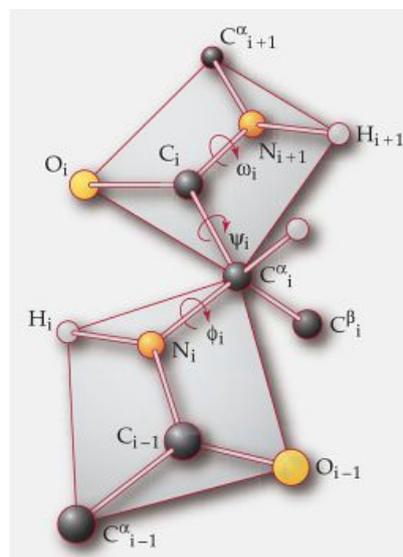
Le phénomène de résonance exige la planéité de la liaison peptidique (établie dans la figure ci-dessous entre les acides aminés i et $i+1$) : les 4 atomes O_i , C_i , N_{i+1} et H_{i+1} forment en conséquence un plan rigide. De ce fait, il n'y a pas de rotation libre autour de la liaison peptidique : l'angle dièdre associé ω_i ne peut prendre que 2 valeurs, 0 ou 180°.



Les carbones α des acides aminés i et $i+1$, soit C_i^α et C_{i+1}^α , se trouvent également dans ce plan, mais la rotation est cette fois possible autour des liaisons $C_{i-1}^\alpha - C_i$ (angle Ψ_i) et $C_{i+1}^\alpha - N_{i+1}$ (angle Φ_{i+1}). Selon la valeur (0 ou 180°) de l'angle ω_i , les chaînes latérales portées par les deux carbones α sont situées soit de part et d'autre (configuration **trans**), soit du même côté (configuration **cis**) de la liaison C-N :



La configuration cis est très peu fréquente car stériquement défavorisée (sauf dans le cas de la proline, où elle est facilitée par la formation d'une liaison covalente entre la chaîne latérale de l'imino-acide et un atome d'azote du squelette polypeptidique, ce qui engendre une contrainte très forte dans la géométrie de ce dernier). La manière dont les plans peptidiques consécutifs vont s'orienter les uns par rapport aux autres autour des angles Ψ et Φ va déterminer la mise en place de motifs structuraux d'un ordre supérieur, les structures secondaires (voir § III).



II. Structure primaire des protéines

La structure primaire d'un peptide ou d'une protéine correspond à l'ordre dans lequel s'enchaînent les acides aminés dans la molécule. Comme indiqué plus haut, on écrit conventionnellement l'extrémité N-terminale du côté gauche. Les acides aminés sont notés soit selon leur code à trois lettres, soit sous celui à une seule lettre. Ainsi, le glucagon, petite protéine constituée de 29 acides aminés et sécrétée par le pancréas, a-t-elle pour structure primaire (code à 3 lettres) :

His Ser Gln Gly Thr Phe Thr Ser Asp Tyr Ser Lys Tyr Leu Asp Ser Arg Arg Ala Gln Asp Phe Val
Gln Trp Leu Met Asn Thr

Bien qu'un pont disulfure (liaison covalente formée entre 2 cystéines) soit un élément de structure tertiaire (car il ne se forme généralement qu'une fois la protéine repliée dans l'espace), il est courant d'en indiquer la localisation au niveau de la structure primaire de la protéine.

1. Détermination de la structure primaire d'une protéine : méthode chimique (Edman)

La méthode est basée sur la réaction d'Edman où un réactif (le phénylthiocyanate) réagit avec la fonction amine N-terminale pour donner un complexe qui, après action d'un acide, libère une phénylthiohydantoïne (à partir duquel l'acide aminé N-terminal sera identifié) et le peptide amputé de son dernier résidu. La chaîne peptidique résultante peut être soumise à un nouveau cycle de traitement, ce qui va permettre d'identifier le résidu en position $n-1$ dans la séquence originale. Et ainsi de suite, en progressant vers l'extrémité C-terminale.

Cependant, le réactif d'Edman n'agit que sur des peptides de petite taille, non sur des protéines entières. Aussi est-il nécessaire de mettre en œuvre une stratégie élaborée, avant de procéder à l'analyse chimique. Les différentes étapes permettant de déterminer la séquence primaire d'une protéine purifiée sont les suivantes :

- 1) Dénaturation et séparation des chaînes polypeptidiques (dans le cas où la protéine en comporte plusieurs). La solution est ensuite divisée en 2 lots qui, à une exception près (voir étape 2), vont subir le même traitement.
- 2) Hydrolyse partielle de la chaîne en peptides, par des moyens chimiques ou, plus souvent, enzymatiques au moyen de protéases qui sont des enzymes hydrolysant, avec une spécificité propre à chacune, les liaisons peptidiques. Deux ensembles de coupures différentes (mettant en œuvre deux protéases de spécificités distinctes) doivent être effectués, chaque lot ne subissant qu'un type de coupure. On obtiendra ainsi des zones de recouvrement entre les peptides issus des deux lots (voir dernière étape). Parmi les principales protéases disponibles, la trypsine coupe les liaisons peptidiques dont le carbonyle est fourni par *Lys* ou *Arg*, alors que la chymotrypsine permet la coupure du côté C-terminal des acides aminés aromatiques *Phe*, *Trp* et *Tyr*.
- 3) Au sein de chacun des deux lots, les peptides sont séparés en combinant des méthodes d'électrophorèse et de chromatographie. Une technique fréquemment employée est l'électrophorèse 2D où les peptides sont séparés dans une dimension selon leur taille et dans l'autre selon la valeur de leur point isoélectrique (en réalisant un gradient de pH).
- 4) La séquence de chacun des peptides ainsi séparés est déterminée chimiquement (et automatiquement) par réaction d'Edman, comme indiqué précédemment.
- 5) La séquence complète de la chaîne polypeptidique est reconstruite grâce à la présence de zones de recouvrement entre les séquences des peptides issus des deux lots.

A titre d'exemple, imaginons que l'on veuille déterminer la séquence d'une protéine dont on aurait, pour commencer, identifié les résidus N- et C-terminaux comme correspondant à un résidu A en position N-terminale et à un résidu R en position C-terminale (cette étape liminaire est facultative, bien que souvent réalisée expérimentalement).

Le tableau de gauche présente les peptides obtenus après digestion par la protéase A, qui coupe après les résidus S et G. Les peptides figurant dans le tableau de droite résultent de l'action de la protéase B qui hydrolyse les liaisons peptidiques qui suivent les résidus R, N et D.

N° peptide	séquence
1a	ADTEVVCG
2a	AKLS
3a	APRKFFVG
4a	APS
5a	ELIHTLNG
6a	G
7a	IYLDFAR
8a	LG
9a	NWKMNG
10a	DKKS

N° peptide	Séquence
1b	APR
2b	FAR
3b	GAKLSAD
4b	GD
5b	KFFVGGN
6b	KKSLGELIHTLN
7b	TEVVCGAPSIYLD
8b	WKMN

Dans le tableau de la série (a), un seul peptide - le numéro 7a - possède une extrémité C-terminale qui n'est ni un résidu S, ni un résidu G : c'est donc que cette extrémité ne résulte pas de l'action protéolytique mais qu'elle constitue l'extrémité C-terminale de la protéine.

```

                                IYLDFAR      (7a)
                        TEVVCGAPSIYLD      (7b)
                                APS          (4a)
                        ADTEVVCG           (1a)
GAKLSAD                       (3b)
AKLS                            (2a)

```

L'extrémité N-terminale est un résidu A, c'est-à-dire l'un des peptides numérotés 1a, 2a, 3a et 4a (pour la série a); d'après les alignements ci-dessus, il ne peut s'agir des peptides 1, 2 ou 4. Il s'agit donc nécessairement de 3a.

```

(3a)      APRKFFVG
(1b)      APR
(5b)      KFFVGGN
(6a)      G
(9a)      NWKMNG
(8b)      WKMN

```

Quel peptide (commençant nécessairement par G, d'après 9a) suit 8b dans la série b ? Il pourrait *a priori* s'agir de 3b ou de 4b. S'il s'agissait de 3b, cela signifierait alors que l'on aurait rejoint l'enchaînement déterminé à partir de l'extrémité N-terminale, sans avoir réussi à placer le peptide 4b. C'est donc nécessairement le peptide 4b qui fait suite à 8b.

```

(8b)      WKMN
(4b)      GD
(10a)     DKKS
(6b)      KKSLGELIHTLN
(8a)      LG
(5a)      ELIHTLNG

```

La séquence de la protéine et l'ordre d'enchaînement des peptides (par exemple pour la série a) sont donc les suivants :

APRKFFVG G NWKMNG DKKS LG ELIHTLNG AKLS ADTEVVCG APSIYLD DFAR

3a
6a
9a
10a
8a
5a
2a
1a
4a
7a

2 Détermination de la structure primaire d'une protéine à partir de la séquence de son gène

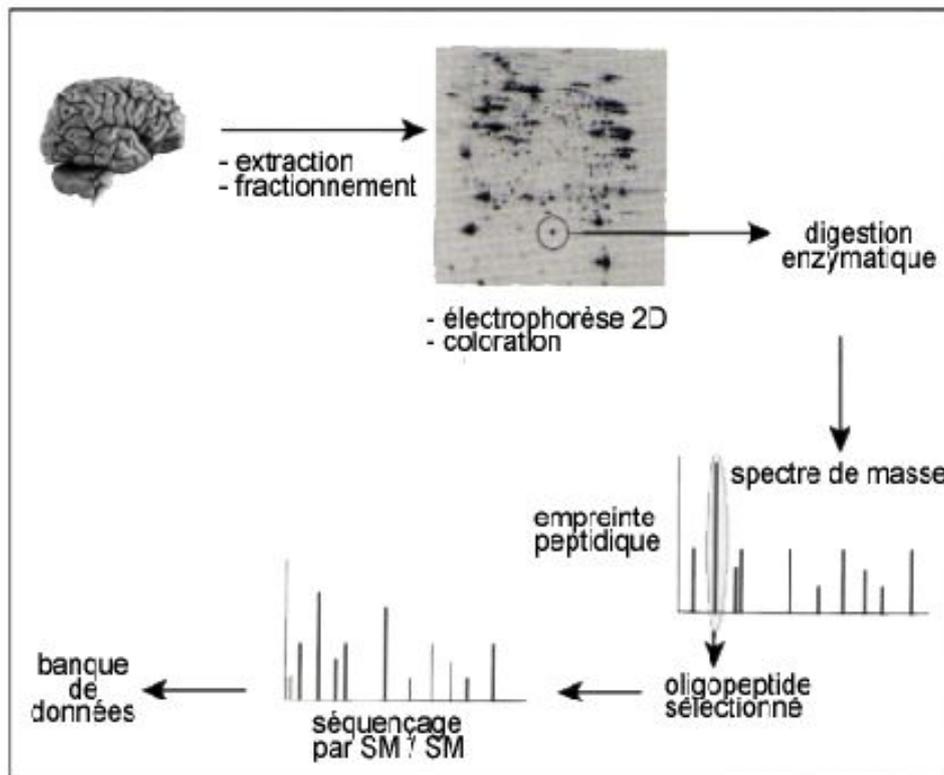
Les méthodes issues de la biologie moléculaire permettent de se contenter de n'établir que la séquence (dite *micro-séquence*) d'une partie d'un seul peptide. Encore faut-il avoir la chance de bien choisir celui-ci, afin que sa séquence soit spécifique de la protéine analysée (pour ce faire, on la compare aux séquences recensées dans les bases de données). On réalise ensuite un *primer* constituée d'une **séquence d'ADN** correspondant à la microséquence protéique déterminée, primer que l'on va tenter d'hybrider avec l'ADN de l'espèce concernée afin d'identifier puis de séquencer le gène correspondant (ce qui est beaucoup plus rapide que de séquencer en totalité la protéine).

La séquence du gène ne permet cependant pas d'identifier la présence éventuelle de modifications post-traductionnelles, modifications qui auraient pu être décelées par l'analyse protéique.

3 Structure primaire des protéines et techniques de la protéomique¹

L'analyse protéomique met en œuvre deux techniques principales, l'électrophorèse bidimensionnelle et la spectrométrie de masse² (SM).

La fraction cellulaire contenant les protéines est soumise à une électrophorèse bidimensionnelle, séparant les protéines selon leur charge dans une dimension et selon leurs poids moléculaire dans l'autre. Les images des cartes ainsi obtenues peuvent déjà être comparées à celles existantes dans les banques de données, afin d'identifier certaines protéines.



La protéine d'intérêt (elles le sont potentiellement toutes) séparée sur le gel est ensuite éluée puis soumise à une digestion enzymatique qui va fournir une collection de peptides. Ceux-ci sont analysés par spectrométrie de masse, ce qui permet d'obtenir une " empreinte peptidique " spécifique de la protéine considérée. Chaque pic correspond à un peptide ayant un rapport masse/charge donné. La technique MS / MS (ou spectrométrie de masse en tandem) consiste à récupérer chaque peptide isolé de la première spectrométrie et à le soumettre

¹ La protéomique a pour objet d'identifier l'ensemble des protéines exprimées (ou protéome) dans un type cellulaire ou un organisme donné, dans un jeu de conditions environnementales ou à un stade de développement particulier.

² La spectrométrie de masse est une technique où la molécule ou macromolécule est soumise à un bombardement électronique qui rompt certaines liaisons chimiques. Les différents fragments ionisés sont ensuite séparés suivant leur rapport masse/charge par application d'un champ électrique ou magnétique. Le spectre de masse d'un polypeptide permet d'en déduire sa structure primaire.

à une deuxième analyse qui va le fragmenter en coupant (principalement) les liaisons peptidiques (à partir des deux extrémités du peptide). Il est alors possible de reconstituer la séquence peptidique. La recherche dans les banques de données (qui contiennent l'ensemble des protéines correspondant aux génomes séquencés) permet d'identifier la protéine à partir d'un séquençage partiel de ses peptides (à condition, naturellement, que le génome correspondant soit entièrement connu).

Plusieurs variantes existent autour de ce schéma général et correspondent à différentes applications ainsi qu'à la mise en œuvre de différents modes opératoires (couplage chromatographie / MS en particulier).

III. Structure secondaire des protéines

Les structures secondaires sont des éléments de structure locale mis en place au niveau de segments du squelette peptidique, indépendamment de la présence des chaînes latérales des acides aminés ou de leur conformation. Néanmoins, les chaînes latérales peuvent venir perturber cette organisation du squelette peptidique. Dans le cas des feuillettes β qui va être évoqué, la notion de contingence locale est en particulier assez relative.

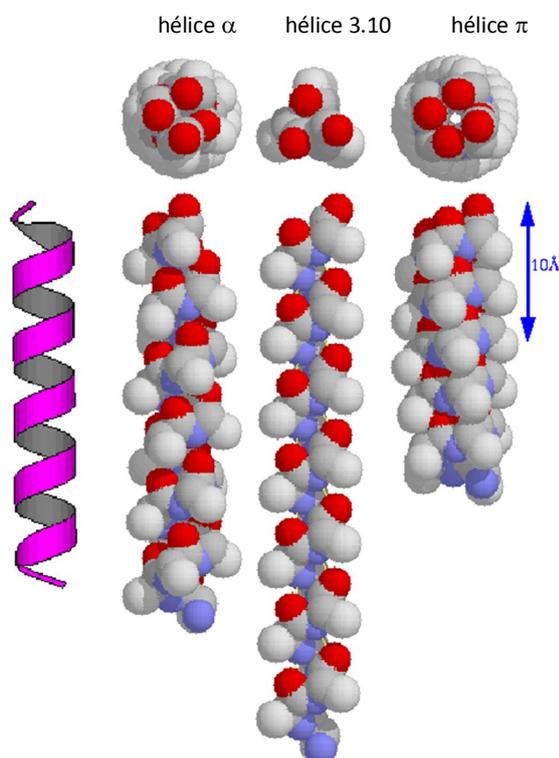
1 Hélices

32 à 38% des résidus de protéines globulaires forment des hélices. Ce sont les motifs de structures secondaires protéiques les plus fréquemment rencontrés. Dans une hélice, les paires d'angles ϕ et Ψ des carbones α de tous les résidus ont mêmes valeurs (données dans le tableau ci-dessous). Inversement, si ces angles sont les mêmes pour tous les résidus, la structure adoptée par la chaîne est une structure en hélice. Il existe plusieurs types d'hélice, selon leur pas (définissant en particulier le nombre de résidus par tour) et leur sens d'enroulement (vers la droite ou vers la gauche).

Type d'hélice	fréquence	ϕ	Ψ	n
α	élevée	-57.8	-47.0	+3.6
3.10	faible	-74.0	-4.0	+3.0
π	très faible	-57.1	-69.7	+4.4

Quelques caractéristiques d'hélices rencontrées dans les protéines

n représente le nombre de résidus par tour. Une valeur positive dénote une hélice enroulée vers la droite.



Exemple d'hélices régulières

Les trois hélices sont reproduites à la même échelle. Il s'agit d'un isopeptide (polyalanine) dont l'extrémité N-terminale est située vers le bas. Les atomes les plus sombres (rouges) représentent l'oxygène. Le carbone des chaînes latérales de l'alanine est très légèrement plus clair que celui des atomes du squelette polypeptidique.

L'abondance relative des hélices de type α est due à plusieurs facteurs :

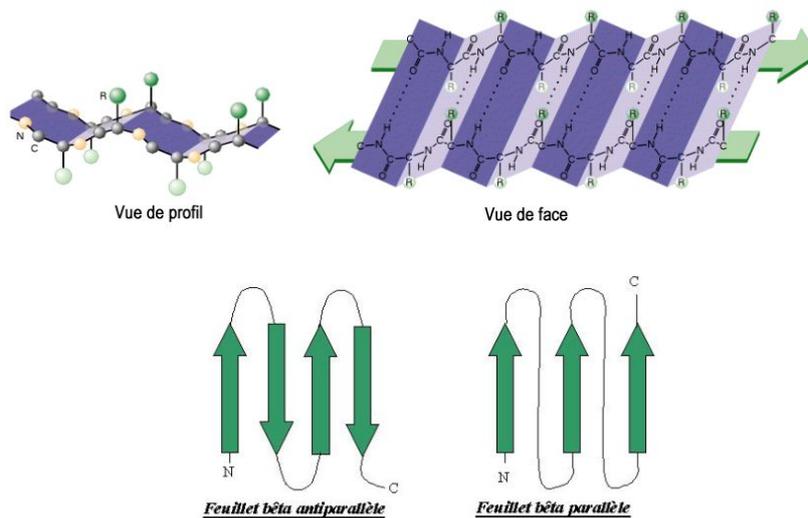
- Les angles ϕ et Ψ d'une telle hélice sont associés à une énergie interne minimum (donc à une stabilité maximum).
- Les atomes du squelette peptidique participant aux liaisons hydrogènes (entre le CO du résidu i et le NH du résidu $i+3$) au sein de l'hélice α sont dans un alignement parfait.
- Le rayon de l'hélice α est optimum pour la mise en place des interactions de van der Waals.
- Les chaînes latérales (rejetées vers l'extérieur) interfèrent très peu avec la structure en hélice.

2 Feuilletts β

Les feuilletts β forment une autre classe de structure secondaire fréquemment rencontrée dans les protéines : 20 à 28% des résidus de protéines globulaires sont engagés dans de telles structures. L'unité de base d'un feuillet β est le brin β (qui peut être vu comme une hélice ayant 2 résidus par tour) dont les angles ϕ et Ψ de son squelette peptidique valent -120° et $+120^\circ$, respectivement. Le brin β est donc formé de la répétition d'un motif. Cependant, il n'existe, **au sein d'un brin**, aucune stabilisation possible par formation de liaisons hydrogènes ou d'interactions de van der Waals entre résidus voisins, du fait de la structure étirée du brin. C'est la raison pour laquelle celui-ci n'acquiert sa stabilité qu'une fois inséré dans un feuillet, structure dans laquelle les interactions stabilisantes entre brins peuvent se mettre en place.

On distingue les feuilletts β parallèles et antiparallèles, selon l'orientation relative des deux brins en interaction. Les chaînes latérales de résidus adjacents d'un brin se trouvent de part et d'autre du feuillet, ce qui empêche toute interaction entre eux. C'est pourquoi les feuilletts β , tout comme certaines hélices α , ont un caractère amphiphile (une de leur face est polaire, l'autre est apolaire). En revanche, à la différence des hélices, les structures β mettent en jeu des résidus (appartenant à des brins différents) qui peuvent être très distants dans la séquence polypeptidique.

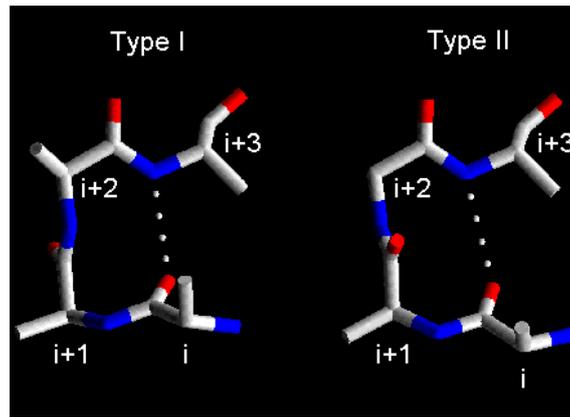
Les feuilletts β antiparallèles sont intrinsèquement les plus stables de par l'orientation optimale des résidus engagés dans la formation de liaisons hydrogènes entre brins.



Feuilletts β antiparallèles et parallèles. Les liaisons hydrogènes s'établissent entre l'azote et l'oxygène. Les flèches symbolisant les brins sont dirigées vers l'extrémité C-terminale.

3 Tours

Un tour (parfois aussi appelé "tonneau") est une structure de retournement de la chaîne polypeptidique. La boucle ainsi formée est stabilisée par une liaison hydrogène établie entre un atome d'oxygène en position i dans la chaîne et l'atome d'azote de la liaison peptidique située en position $i+3$.

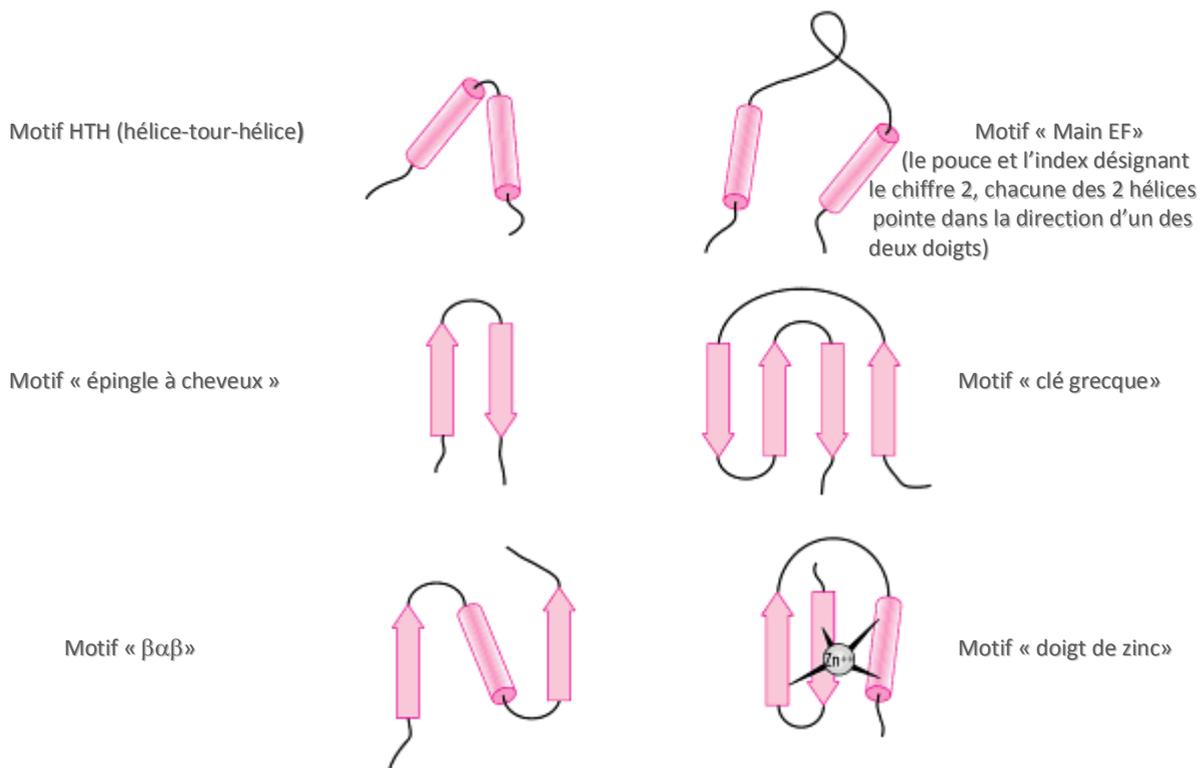


Les tours sont nombreux dans les protéines globulaires. Ils se trouvent généralement à la surface des molécules. Les sites de reconnaissance de l'antigène dans les anticorps, les sites de phosphorylation, de glycosylation ou d'hydroxylation sont fréquemment situés à proximité d'un tour.

On distingue deux types de tours, selon la valeur des angles ϕ et Ψ des résidus en position $i+1$ et $i+2$ (figure ci-dessus).

4 Structures super-secondaires ou motifs

Au sein de la protéine, les éléments de structure secondaire créent des arrangements géométriques tridimensionnels appelés structures super-secondaires ou encore motifs. Ainsi les feuilletés β peuvent-ils former une classe particulière de tours appelés "épingles à cheveux". Certains de ces motifs peuvent présenter par eux-mêmes une activité biologique (fixation de cations, liaison à l'ADN, etc.). D'autres sont inclus dans des structures macromoléculaires fonctionnelles plus complexes.



Les cylindres figurent des hélices α , les flèches (dirigées vers l'extrémité C-terminale) des brins β

IV. Structure tertiaire des protéines. Notion de domaines

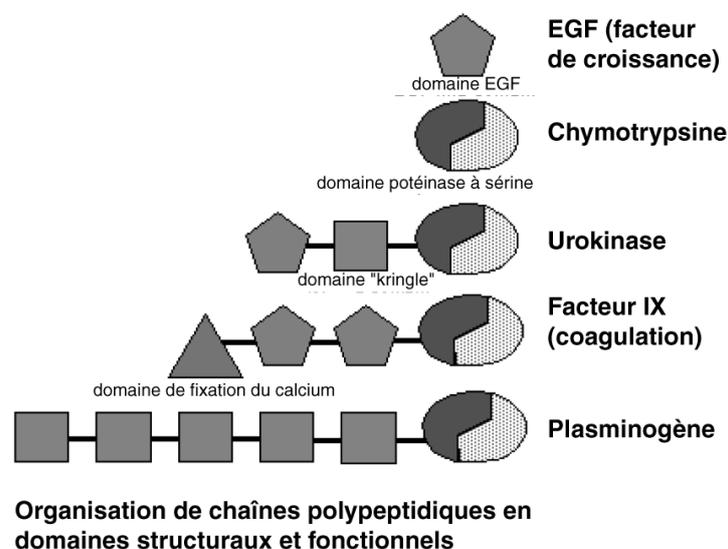
La structure tertiaire d'une protéine correspond à l'arrangement tridimensionnel de ses groupements constitutifs (incluant les éventuels ponts disulfures). Cet arrangement permet le rapprochement de domaines éloignés dans la structure primaire. La structure tertiaire est souvent la structure active de la protéine, sauf dans le cas où celle-ci comporte plusieurs chaînes polypeptidiques associées (leur arrangement définira un niveau d'organisation supérieur qualifié de *quaternaire*). Il arrive néanmoins qu'en plus de son repliement dans l'espace, la protéine ait à subir une ou plusieurs modifications chimiques post-traductionnelles avant d'être fonctionnelle.

La notion de **domaine** protéique est aussi importante qu'ambiguë. Les structuralistes définissent le domaine comme une unité structurale capable de se replier indépendamment du reste de la protéine. En biochimie, les domaines sont fréquemment décrits comme des régions protéiques dont la fonction a été expérimentalement caractérisée (indépendamment de la structure). En génomique comparative, les domaines sont des séquences homologues³ que l'on peut rencontrer dans des contextes moléculaires différents. Ces trois définitions sont souvent compatibles (il existe une relation structure / fonction) et définissent la même entité, même s'il existe des exceptions.

Les fonctions distinctes associées à différents domaines peuvent par exemple correspondre à la fixation d'un ligand, à la reconnaissance d'un partenaire, à un ancrage membranaire...

Une protéine constituée de différents domaines structuraux est à même de réaliser plusieurs fonctions distinctes. C'est typiquement le cas des récepteurs transmembranaires qui sont des protéines *intégrales* des membranes (voir chapitre sur les lipides). Ces protéines comportent généralement trois domaines, un domaine extracellulaire en interaction avec l'environnement cellulaire et qui joue un rôle de récepteur, un domaine intracellulaire au contact du compartiment cytoplasmique et un domaine intramembranaire, la plupart du temps structuré en hélice α . Le domaine intracellulaire est souvent responsable de l'activation d'une cascade de signalisation, après avoir été lui-même activé (par un changement subtil de forme ou de conformation) suite à la fixation du récepteur à l'autre extrémité de la protéine.

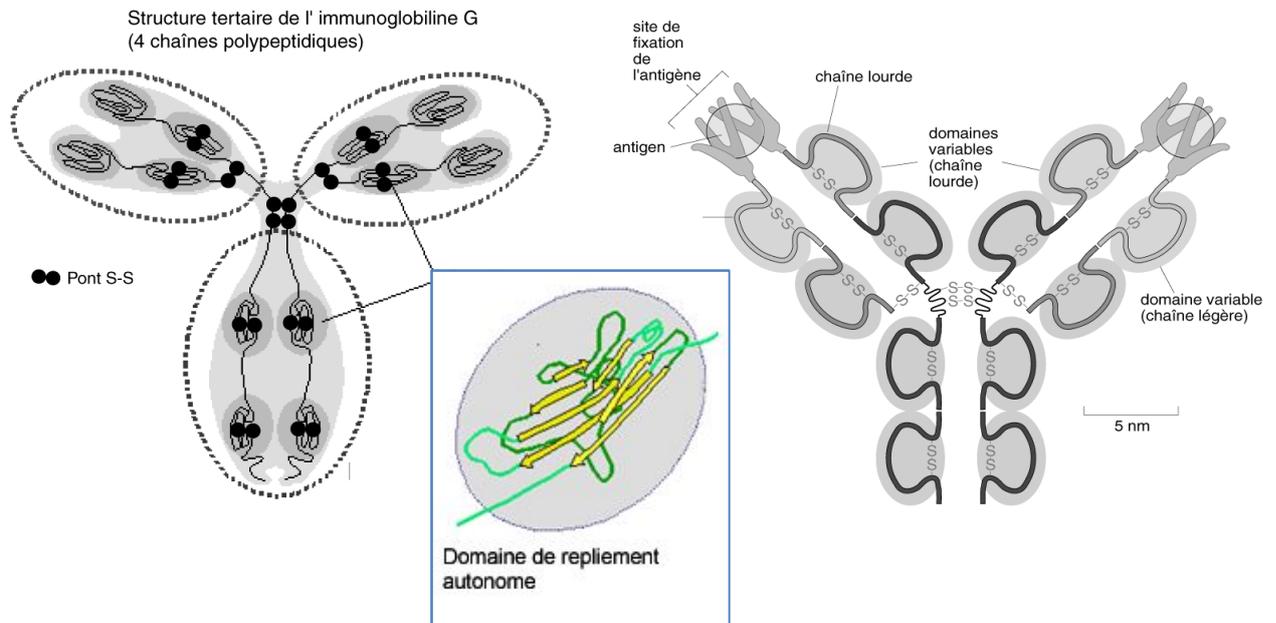
Certains **motifs** de structures super-secondaires peuvent former des domaines et se retrouver dans des protéines sans lien phylogénétique. De tels motifs sont généralement déterminés par quelques acides aminés et associés à des interactions bien précises. Par exemple, le motif « doigt de zinc » fixe l'ion Zn^{2+} , qui est impliqué dans des interactions spécifiques avec l'ADN. Le motif « main EF » est retrouvé dans beaucoup de protéines (mais pas toutes) qui fixent l'ion calcium Ca^{2+} .



³ L'homologie désigne une relation d'évolution entre des gènes. Des gènes sont dits « homologues » s'ils sont issus d'un gène ancestral commun.

La Figure ci-dessus illustre la vision moderne selon laquelle il n'existe, dans la diversité des protéines existantes, qu'un répertoire limité de domaines qui se dupliquent et se combinent de différentes manières pour former l'ensemble des protéines d'un génome.

La notion de domaine peut être illustrée à partir de la structure de l'immunoglobuline G (IgG). Cette protéine est formée de l'association de 4 chaînes polypeptidiques liées entre-elles par des ponts disulfures, c'est-à-dire par des liaisons covalentes. C'est la raison pour laquelle on parle de **structure tertiaire** de l'IgG. Si ces chaînes avaient été en interactions non-covalentes, on aurait parlé de la structure quaternaire de la protéine (ce que font d'ailleurs certains).



Deux méthodes (que nous ne ferons ici que mentionner) permettent d'étudier expérimentalement la structure tridimensionnelle d'une protéine. Ce sont :

- La diffraction des rayons X. Les protéines doivent auparavant avoir été cristallisées.
- La résonance magnétique nucléaire (RMN), cependant limitée, aujourd'hui, à des protéines de petite taille (ne dépassant pas 25 000 daltons de poids moléculaire). Cette technique présente l'avantage d'opérer sur des protéines en solution (*i.e.* non cristallisées).