

# AI for Health

Chroniques AI for Health - Série des « Omiques »

## La protéomique : le royaume de la fonction des gènes

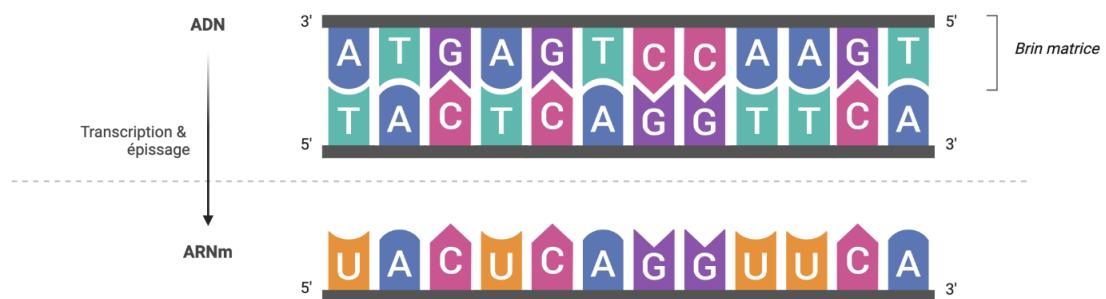
Temps de lecture : moins de 15mn

Bienvenue dans ce troisième épisode de notre série des omiques sur la **protéomique** !

### Partie 1 : Résumons

Dans les **deux premiers volets**, génomique et transcriptomique, nous avons vu que certaines séquences des gènes permettaient de coder pour des ARN messagers.

Dans le noyau, le processus de **transcription** permet d'ouvrir l'ADN et transcrire une séquence matrice en un brin d'ARN complémentaire. A se lie à T (U pour l'ARN) et C se lie à G. Par exemple, ci-dessous, ATGAG sera transcrit en UACUC.



L'**épissage** permettra par la suite de retirer des blocs de séquences, les **introns**, et de ne laisser que des **exons** dans l'ARN messager (ARNm) mature. L'ARNm rejoindra dès lors le cytoplasme, à l'extérieur du noyau, où se déroulera le processus de **traduction** que nous détaillerons dans cet épisode.

### Partie 2 : les protéines

On entend beaucoup le mot “protéines” lorsque l'on parle nutrition, sport, santé,... Mais qu'est-ce qu'une protéine ? Zoomons !

#### Types de protéines

Anticorps, enzymes, kératines, collagène, ... Tous ces termes assez connus sont des familles ou des noms de protéines !

Les protéines généralement sont divisées en deux groupes : **fonctionnelles ou structurales**, c'est-à-dire ayant des rôles notables ou de “charpente” du vivant.

# AI for Health

Quelques exemples avec les familles de protéines citées précédemment :

Protéines de fonction		Protéines de structure	
Enzymes	Catalysent (accélèrent/permettent) les réactions chimiques du vivant	Collagène	Protéine constituant principalement le tissu conjonctif, "remplissant" les intervalles entre les organes
exemple : les lipases "découpent" les graisses pour permettre leur digestion		fun fact : le collagène représente près d'1/4 des protéines totales des mammifères	
Anticorps (immunoglobulines)	Détection/neutralisation d'éléments étrangers de l'organisme	Kératines	Constituantes principaux des cheveux, poils, ongles et également présentes dans la peau
exemple : les immunoglobulines A (IgA), contenues notamment dans le lait maternel, participent activement à la protection des muqueuses		fun fact : la kératine étant insoluble, un excès de cette protéine peut rendre les cheveux gras à la place de l'effet brillant et lisse. C'est pourquoi certains shampoings sont sans kératine !	

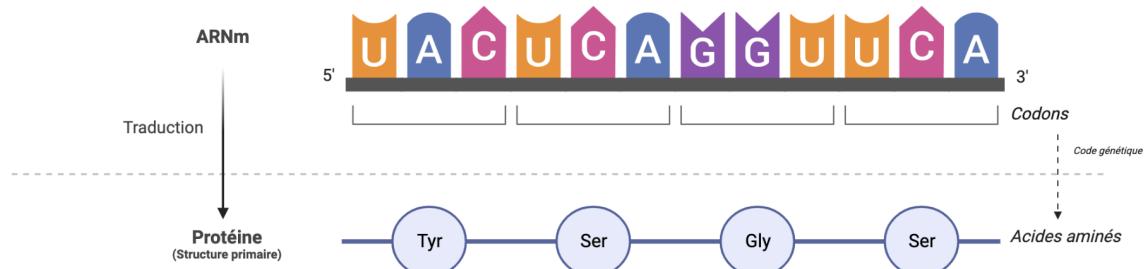
Les protéines permettent, entre autres, la compréhension de la fonction du gène : ce pour quoi il "code". L'occasion de rappeler que la transcription de l'ADN peut également engendrer des **ARN non codants**, parfois fonctionnels, comme c'est le cas des ARN de transfert dont le rôle est clé dans la formation des protéines.

## Partie 3 : la traduction

### Composition d'une protéine

Une protéine est une macromolécule composée d'un enchaînement d'**acides aminés**, au nombre de 20 différents chez l'Homme. C'est la séquence de l'ARNm qui sert de base de lecture pour permettre de lier les différents acides aminés entre eux et former la protéine.

Comment ? Grâce au fameux **code génétique**. Ce code agit comme un dictionnaire qui permet de **traduire** la séquence de 3 bases azotées consécutives - un codon - en un acide aminé. Par exemple, chez l'Homme, un codon UAC donnera une tyrosine, UCA une sérine ou GGU une glycine. Ce principe permet de donner la **structure I** ( primaire) de la protéine, à l'image d'un collier où des acides aminés sont des perles reliées entre elles sans structure 3D apparente.



Toute une machinerie cellulaire permet d'assembler les différents acides aminés à la chaîne sur l'ARNm matrice. Vous l'aurez compris, **la complémentarité** est un concept très important en biologie.

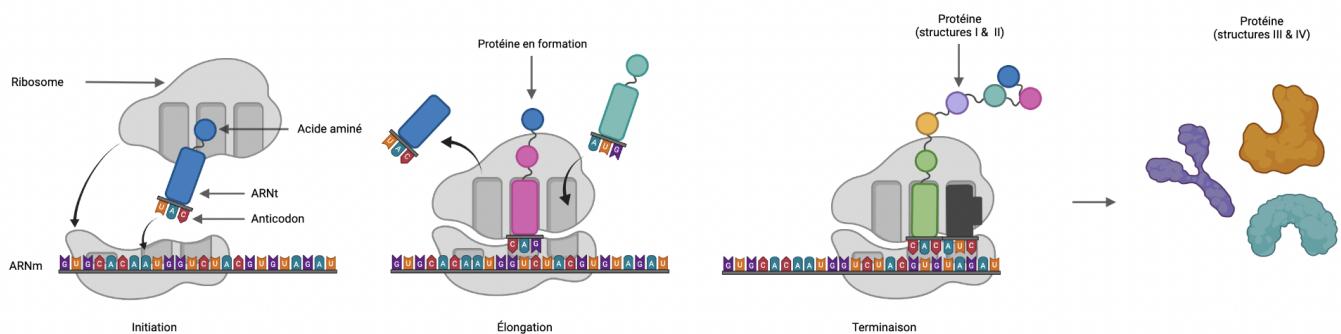
# AI for Health

Reparlons des **ARN de transfert** (ARNt) : ce sont eux qui vont se lier à l'ARNm grâce à un anticodon, complémentaire d'un codon de l'ARNm. Cette liaison de l'ARNt s'effectue grâce aux ribosomes, pour permettre de greffer les différents acides aminés et ainsi former la protéine dans son entièreté 😊

**Simplifions** : Imaginez vouloir surligner jusqu'à 20 champs lexicaux de 20 couleurs différentes dans un livre. Prenez 20 surligneurs (ARN de transfert) de couleurs différentes (acides aminés), un support pour permettre de surligner de manière précise (ribosomes) et reliez tous les mots surlignés dans le même ordre : vous avez désormais une protéine lexicale !

Ci-dessous, voici schématiquement comment se forme une "vraie" protéine :

## Schéma succinct du processus de traduction



**Fun fact** : la métaphore du langage est encore une fois omniprésente ! On parle par exemple de **cadre de lecture** pour désigner la succession des codons lus dans un sens. Il ne suffit donc parfois que d'une base azotée décalée pour générer des mutations, la précision est de mise !

Comme toute molécule, les acides aminés peuvent avoir des interactions entre eux et ainsi se rejeter, s'attirer ou autre. Ces différentes interactions vont donc remodeler ce fameux collier de perles dans un premier temps **en 2D**, comme si vous le posiez sur une table et vous amusiez à le mettre dans différentes positions à plat (**structure II**).

Dans un second temps, l'ensemble des interactions et d'éventuelles modifications post-traductionnelles vont permettre de donner une **structure 3D** à la protéine, comme si vous preniez le collier et l'emméliez dans votre main (**structures III et IV**). C'est à partir de la structure III que la protéine acquiert sa potentielle fonction.

## **Vous suivez toujours ?**

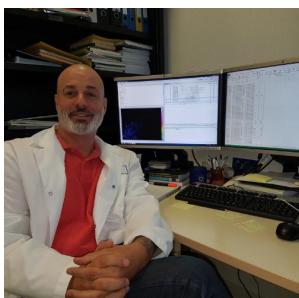
Super ! Prenez une petite pause et installez-vous confortablement.

Saviez-vous que **la chenille et le papillon** en devenir avaient le même génome ? En effet, l'ensemble des gènes de la chenille ne va pas varier lors de sa transformation en papillon. Cependant, de nombreuses différences vont survenir au niveau de l'ensemble de ses **protéines**.

# AI for Health

Maintenant que la formation des protéines n'a plus de secrets pour vous, explorons ensemble la science permettant de les étudier, aux côtés d'un expert.

## Partie 4 : la Protéomique



Partie rédigée grâce aux réponses du **Pr Christophe Hirtz**, spécialiste R&D et essais cliniques dans les champs de la protéomique clinique et des maladies neurodégénératives. Le Pr Hirtz est également directeur de la [plateforme de protéomique clinique de Montpellier](#).



### C'est quoi, la protéomique ?

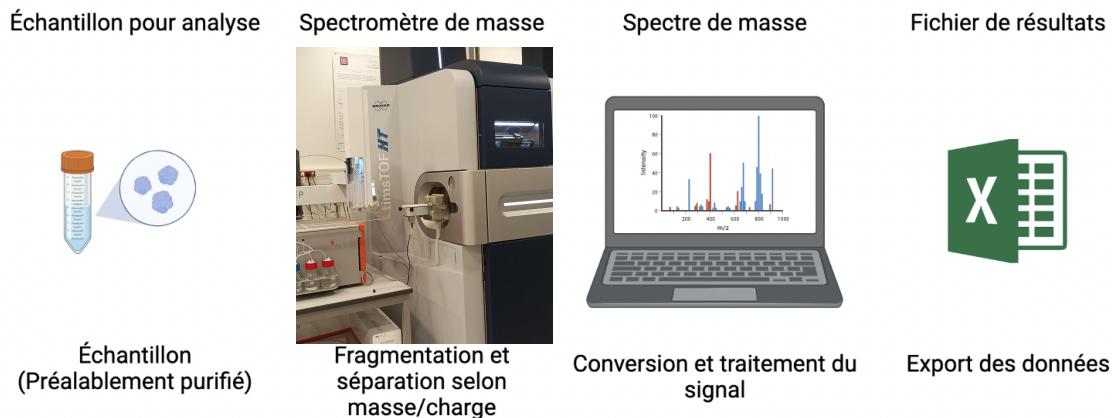
À l'instar de la génomique et de la transcriptomique, la protéomique est une photographie à un moment donné d'un ensemble de **protéines** cette fois-ci, dans un fluide biologique, un tissu ou une cellule donnés.

On retrouve des dimensions d'études communes aux Omiques, à savoir une analyse à large échelle - des protéines - permettant leur **identification**, leur **quantification** voire même la modélisation de leur **structure**. Ces différentes informations permettent un large panel d'applications, de la recherche fondamentale au diagnostic clinique en passant par le suivi de biomarqueurs par exemple.

### Quels sont les outils associés ?

Afin de mettre en place une analyse en protéomique, les chercheurs s'équipent de **spectromètres de masse**, machines capables de fragmenter les protéines et de séparer les fragments en fonction d'un rapport de leur masse sur leur charge électronique.

Le signal sera ensuite traité et traduit en signal électrique afin de produire un graphique appelé **spectre de masse**, permettant l'analyse des résultats.



# AI for Health

Ces outils sont de plus en plus performants ! Là où des analyses pouvaient prendre des heures voire des jours, les appareils de pointe d'aujourd'hui sont capables d'**analyser jusqu'à 6500 protéines en 20mn** tout en prenant en compte la complexité de l'échantillon de départ.

Rappelez-vous, on dénombre environ 25000 gènes codant pour des protéines (*épisode 1 : La génomique, voyage au centre de nos gènes*). Pour illustrer, on est donc aujourd'hui capable de **traiter un quart du protéome en moins d'une demi-heure** !

Pour aller plus loin, la sensibilité et la précision d'analyse de nos jours est telle qu'elle permet même de différencier des protéines très semblables, appelées protéoformes, issues notamment des différents variants d'ARNm après épissage alternatif (*épisode 2 : La transcriptomique, l'ARN passé au crible*).

Il est maintenant également possible d'analyser en "single-cell", c'est-à-dire **sur une cellule isolée**. Cela revient à avoir des informations précises sur le contenu en protéines d'un seul échantillon 1000 fois plus petit qu'un grain de sable !

## Quelles sont les promesses en clinique ?

Outre son intérêt dans la recherche et le développement, la protéomique s'inscrit entre autres parfaitement dans l'essor de la médecine personnalisée et des biothérapies.

### La biothérapie

Les **thérapies cellulaires** permettent notamment de corriger ou restaurer une fonction perdue en remplaçant et/ou ajoutant des cellules saines, palliant ainsi le dysfonctionnement des cellules malades.

La protéomique permettrait de **phénotyper** efficacement des cellules médicamenteuses, c'est-à-dire en donner la caractérisation la plus complète possible : sa nature, ses fonctions, sa composition, ses interactions,... De cette manière, la R&D pourra développer des cellules médicamenteuses toujours plus spécifiques et adaptées à une pathologie et un patient donné.

### La médecine personnalisée

Jetez un œil à vos derniers résultats de prises de sang. Si vous les comparez entre eux, vous noterez une sorte de "redondance" des protéines que l'on va rechercher : albumine, TSH, hGH, HDL/LDL, Créatine,...

Sur les milliers de protéines analysables, **seules quelques dizaines sont analysées en clinique** lors de diagnostics de routine. La protéomique est remplie de promesses ! Imaginez pouvoir obtenir un profil protéomique complet permettant de poser un diagnostic et une indication thérapeutique encore plus précisément.

# AI for Health

On pourrait se prendre à rêver d'un changement de paradigme en passant de 15 à pourquoi pas 1000 protéines analysées par échantillon, une avancée majeure dans le suivi complet de la santé du patient. Mais ce n'est pas si simple !

## Quels freins ?

La technologie est aujourd'hui globalement prête, mais quelques aspects freinent l'intégration optimale de la protéomique en clinique :

### Sur la technique en elle-même : les données

Qui dit haut débit ou augmentation de la précision des résultats et de la richesse des informations, dit augmentation du volume et de la typologie de données générées.

- **L'extraction des données** est laborieuse :
  - L'export des données de spectres de masse se réalise via des tableurs allant de 10 à 100 Mo, ce qui est conséquent lorsque l'on sait qu'un tableur dépasse rarement le Mo.
- **L'analyse des données** subséquente fait face à des pertes d'informations :
  - D'une part, à cause d'un fort manque de complétude des **bases de données** de protéomique.
    - Résultat, sur un spectre de qualité, environ 80% des pics correspondant à des protéines ne sont pas identifiés car elles sont non répertoriées.
    - le dosage par spectrométrie de masse ne donne pas les mêmes concentrations qu'en dosage clinique ELISA classique, il faudra donc établir les nouveaux seuils pathologiques pour ces nouveaux biomarqueurs.

### Sur son intégration en clinique : le financement et l'appropriation par les professionnels

Se procurer des machines de pointe a un **coût** conséquent, sans compter les licences additionnelles de logiciel d'analyse des données ou encore les ordinateurs permettant de les traiter. Sur le long terme, les questions du coût de la maintenance, des consommables ainsi que de la formation des opérateurs spécialisés sont également à prendre en compte.

**L'appropriation et l'intelligibilité** des résultats de l'analyse protéomique pour un usage en clinique sont primordiales pour permettre son utilisation. Un praticien doit pouvoir tirer une conclusion d'un rapport protéomique facilement et de manière fiable.

De grands enjeux de développement de **logiciels basés sur l'IA** sont alors posés, afin de permettre à terme de renvoyer des interprétations de résultats rapides, fiables et intelligibles par les praticiens.

# AI for Health

La protéomique est un domaine rempli de promesses et n'a rien à envier à ses cousines de la famille des Omiques, dont l'essor commun révolutionnera la santé de demain.

Vous souhaitez nous contacter pour nous rejoindre lors de nos ateliers ou conférences, nous proposer une mission ou simplement vous renseigner ?  
**Ecrivez-nous à [ai@startupinside.com](mailto:ai@startupinside.com).**

*Figures adaptées de différents éléments de l'outil biorender.com.*