

Génomique fonctionnelle chez les végétaux

I- Généralités

1- L'ère des -omics

- **Génomique** : analyse le fonctionnement d'un organisme, d'un organe, d'un cancer à l'échelle du génome et non d'un seul gène.
- **Transcriptomique** : analyse l'ensemble des ARN messagers (quantification) produits lors du processus de transcription d'un génome.
- **Protéomique** : analyse l'ensemble des protéines d'une cellule d'un tissu, d'un organe ou d'un organisme à un moment et sous des conditions données.
- **Metabolomique** : analyse l'ensemble des métabolites (sucres, acides aminés, acides gras, etc.) présents dans une cellule, un organe, un organisme. utilise la spectrométrie de masse et la résonance magnétique nucléaire.
- **Interactomique** : étude des interactions entre les différentes molécules biochimiques, la compréhension des réseaux d'interactions protéiques.
- **Lipidomique** : analyses qualitatives et quantitatives de lipides sur microéchantillons biologiques de toutes natures : fluides, tissus ou encore cellules.

2- Génome

- l'ensemble du matériel génétique d'un individu ou d'une espèce codé dans son ADN (Homo sapiens) ou dans son ARN (rétrovirus).

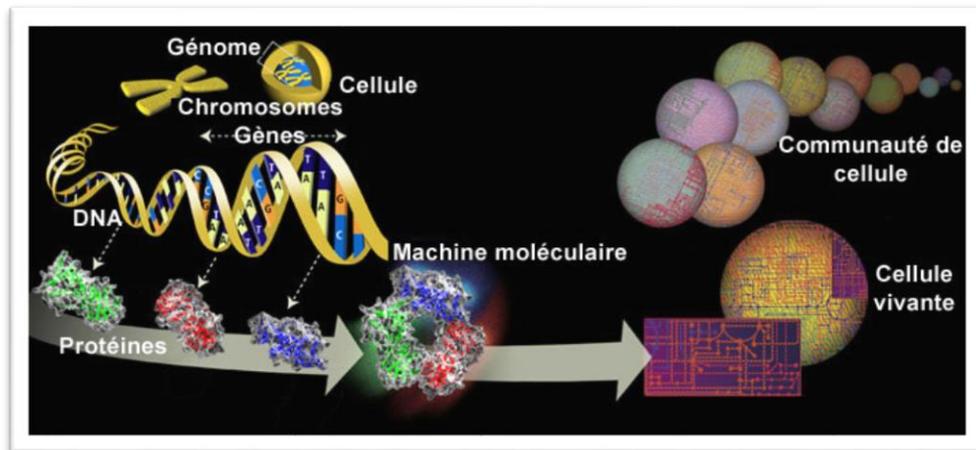


Figure 01. Du génome à la cellule

3- Génotype

- Le **génotype** est l'ensemble ou une partie donnée de la composition génétique (information génétique) d'un individu. Le génotype d'un individu est donc la composition allélique de tous les gènes d'un individu.
- Un gène est une partie du génome dont la séquence sert à coder une molécule qui intervient dans la vie de la cellule (enzyme, ARNt, ARNr..), il est caractérisé par la fonction qu'il assure.

- Le terme allèle désigne chacune des différentes formes ou versions possibles d'un même gène, relatives au même caractère.
- On peut avoir dans les populations naturelles plusieurs séquences différentes pour un même locus. On parle de différents allèles.
- Ce polymorphisme génétique est dû à l'apparition de mutations qui sont des variations de la séquence nucléotidique.

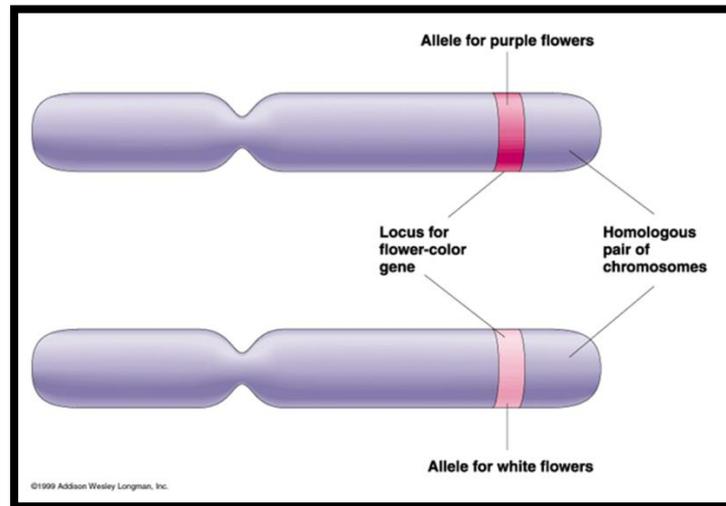


Figure 02. Les Allèles dans les chromosomes homologues

4- Phénotype

- L'état d'un caractère observable (caractère anatomique, morphologique, moléculaire, physiologique, ou éthologique) chez un organisme vivant.



Figure 03. Polymorphisme du phénotype

5- Théorie fondamentale(centrale) de la biologie moléculaire

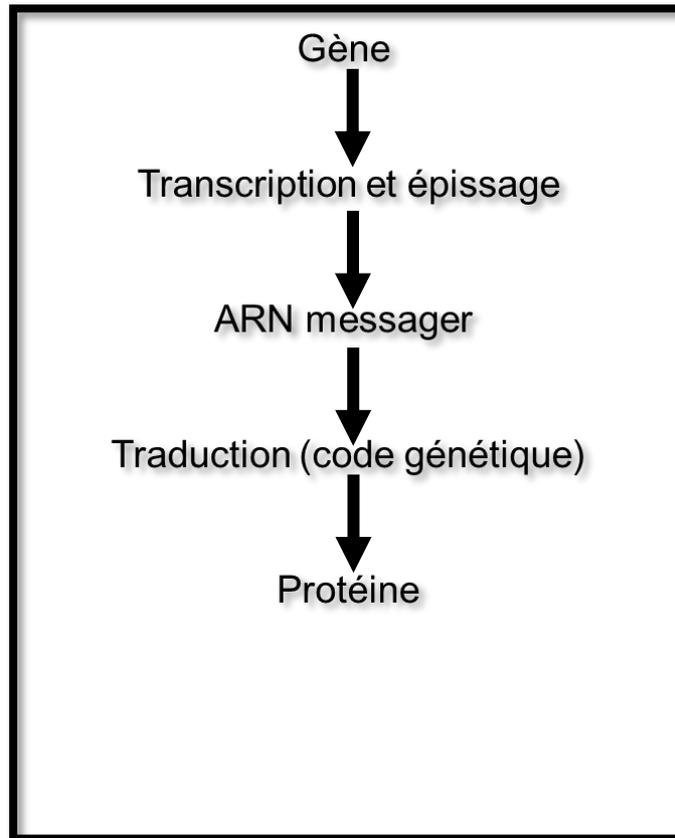


Figure 04. Le dogme central de la biologie moléculaire

- Par **théorie fondamentale de la biologie moléculaire**, les biologistes représentent le modèle schématique de la conservation et de l'utilisation de l'information génétique. L'ADN est le support stable et transmissible de l'information génétique qui définit les fonctions biologiques d'un organisme (reproduction, nutrition, excrétion, action sur l'environnement, communication, etc.). Il est transcrit en ARN (un langage similaire) qui n'a qu'une vie temporaire, cet ARN assure soit :
 - ✓ Une fonction structurale (squelette de complexes nucléo-protéiques).
 - ✓ Une fonction enzymatique (synthèse des protéines, exportation des protéines, épissage des ARNm eucaryotes, etc.).
 - ✓ Une fonction de transport de l'information génétique.
- L'ARN de type messager (ARNm) est traduit en protéines par le ribosome(complexe nucléo-protéique). On parle de traduction des ARNm en protéines car le langage change.
- La théorie fondamentale se résume ainsi : **L'ADN dirige sa propre répllication en ADN identique, ainsi que sa transcription en ARN, pouvant ou non être traduit en protéines.**
- Mais ces hypothèses semblent ne pas résister à l'expérience, ni aux faits. Ainsi, la théorie initiale n'envisageait pas la possibilité d'un « retour » à une forme ADN à partir de l'ARN. Or il fut mis en évidence chez les rétrovirus l'existence d'une enzyme (catalyseur protéique), la **transcriptase inverse**, capable de rétrotranscrire l'ARN viral en ADN. Cette découverte ne remet pas complètement en cause la théorie centrale de la biologie moléculaire, mais permet

de préciser les liens existants entre les différents supports moléculaires intervenant dans l'expression de l'information génétique.

- Lorsque Francis Crick formula cette théorie en 1958, il utilisa l'expression *central dogma of molecular biology*. Le mot dogme prête ici à confusion, car il s'agit plutôt d'une hypothèse scientifique et non pas d'une doctrine établie comme une vérité incontestable. En anglais, le terme *dogma* renvoie également à « une idée qui n'est pas étayée par des preuves rationnelles », ce qui à l'époque était exact. Sa formulation fut malheureusement entérinée par l'usage et on parle couramment de « dogme central de la biologie moléculaire » pour désigner les relations entre ADN, ARN et protéines. Le dogme renvoie à l'idée que cette relation soit unidirectionnelle (de l'ADN vers les protéines en passant par l'ARN). Cette thèse a été confrontée pendant la deuxième moitié du vingtième siècle, notamment par les découvertes en épigénétique et par les hypothèses de constructions de niches.

II - La Génomique

- La génomique est une discipline de la biologie moderne. Elle étudie le fonctionnement d'un organisme, d'un organe, d'un cancer, etc. à l'échelle du génome, au lieu de se limiter à l'échelle d'un seul gène.

La génomique se divise en deux branches :

- ✓ La *génomique structurale*, qui se charge du séquençage du génome entier ;
 - ✓ La *génomique fonctionnelle*, qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le transcriptome et le protéome.
- La génomique est l'équivalent de la métabolomique pour les métabolites.

1- Histoire de la génomique

- L'essor de cette discipline a été facilité par le développement des techniques de séquençage des génomes et la bio-informatique.
- ✓ en 1972, le premier véritable séquençage d'un génome est publié, avec la lecture de la séquence ARN du gène du virus bactériophage MS2¹.
- ✓ Elle a été très médiatisée à la fin du xx^e siècle avec la compétition entre différentes équipes scientifiques pour la publication de la première carte du génome humain, annoncée conjointement le 26 juin 2000 par Bill Clinton et Tony Blair.
- ✓ Depuis, un nombre croissant de génomes complets sont séquencés chez des espèces vivantes très différentes : le ver *Caenorhabditis elegans* en 1998, la mouche drosophile et la plante *Arabidopsis thaliana* en 2000 ou encore, le chien en 2005. En septembre 2007, une équipe menée par le biologiste et entrepreneur Craig Venter a publié le premier génome complet d'un individu qui se trouve être Craig Venter lui-même. Le génome du codécouvreur de la structure de l'ADN et ancien directeur du Projet génome humain, James Watson, a aussi été séquencé dans son intégralité à la même période.

2- Utilité de la génomique

- Connaitre la séquence nucléotidique permet une multitude d'études :
 - ✓ Exploration des fonctions associées aux gènes, aux variations alléliques et au polymorphisme génétique ;
 - ✓ Reconstruction d'arbres phylogénétique d'espèces vivantes ;
 - ✓ Analyse de l'histoire évolutive des êtres vivants en lien avec leur écosystème
 - ✓ Compréhension de maladies liées aux gènes.

Les branches de la génomique

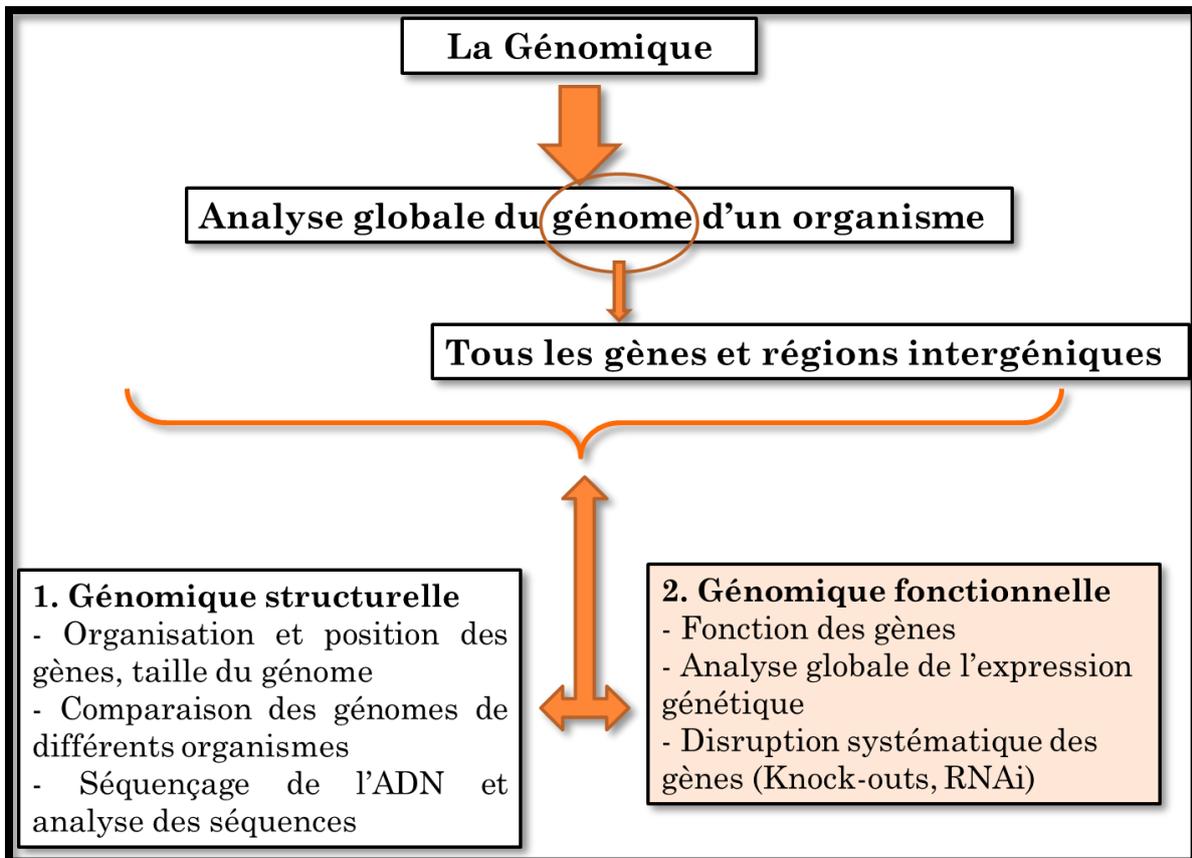


Figure 05. Différentes branches de la Génomique

a-Génomique structurale

- Cette branche de la génomique regroupe toutes les analyses de la structure des génomes (Ici « *structure* » est entendu au sens « *organisation des génomes* ») ; Les méthodes concernées sont donc le séquençage des génomes, l'identification des gènes, des séquences régulatrices, des séquences répétées, etc.

Séquençage des génomes

- Les projets de séquençage de génome sont des projets scientifiques qui ont pour but d'obtenir les séquences complètes des génomes de différents organismes: bactéries (*E.coli*), plantes(*Arabidopsis Thaliana*), champignons (Champignon de Paris, *Agaricus bisporus*), animaux (Le chien, *Canis lupus familiaris*; abouti en 2005), et humain.
- Ce travail nécessite la séquence de l'ADN de chacun des chromosomes de l'espèce. Pour une bactérie, il n'y a qu'un chromosome à séquencer. Pour l'espèce humaine, qui possède 22 paires de chromosomes et 2 chromosomes sexuels (X et Y), il y a 24 chromosomes à séquencer.
- Le projet génome humain est abouti depuis 2003 (1999-2003).

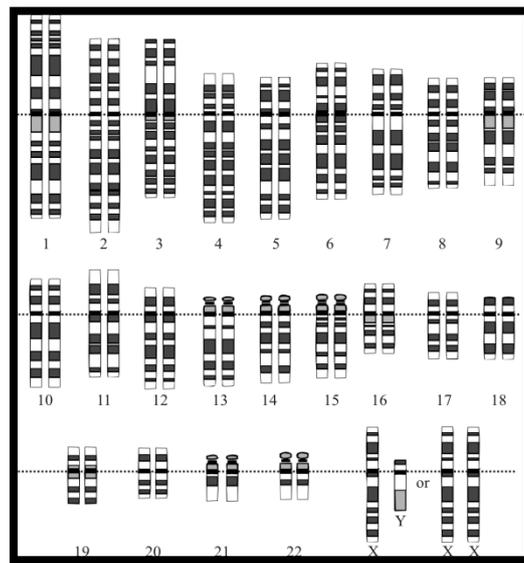


Figure 06. Le génome humain (constitué de l'ensemble de l'information portée par nos 23 paires de chromosomes)

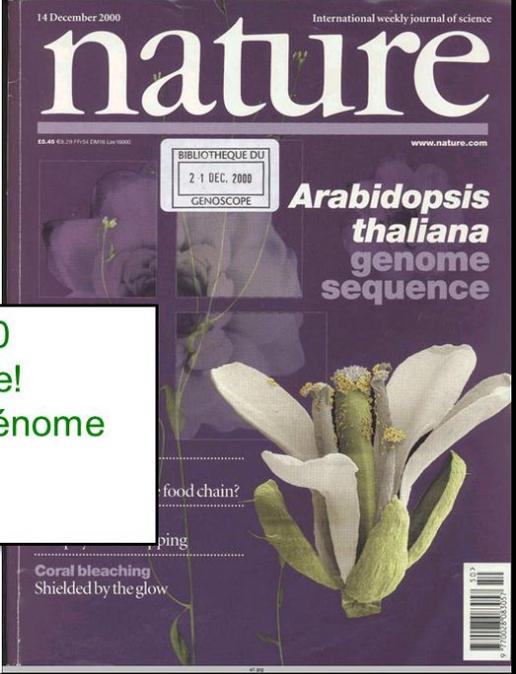
Assemblage génomique

- L'assemblage génomique consiste à prendre un grand nombre de séquences d'ADN, qui ont été obtenues par la méthode *Whole Genome Shotgun*, pour les remettre ensemble et recréer le chromosome original. Dans la méthode *Whole Genome Shotgun*, l'ADN d'un seul organisme est fragmenté en millions de morceaux. Ceux-ci sont lus par des séquenceurs d'ADN automatiques (qui peuvent déterminer 900 bases par minute). Les quatre bases sont adénine, guanine, cytosine, et thymine, abrégées AGCT. Un logiciel d'algorithme d'assemblage génomique analyse chaque fragment et les aligne tous les uns derrière les autres, en détectant les zones de chevauchement entre les fragments. Si deux séquences se correspondent elles sont considérées comme étant identiques ce qui relie les fragments.
- L'assemblage génomique peut être problématique pour les informaticiens, car les génomes contiennent des séquences qui se répètent, et ces répétitions peuvent être longues de plusieurs milliers de nucléotides (ou bases) et sur plusieurs endroits du génome. C'est surtout le cas pour les grands génomes des plantes et des animaux.

Pourquoi *Arabidopsis* a-t-elle été choisie comme plante modèle ?

Arabidopsis thaliana (L.) Heynh
Arabette des Dames, crucifère modèle

Petit génome 125 Mb



14 Décembre 2000
une date historique!
La séquence du génome
d'*Arabidopsis*
a été publiée

Arabidopsis thaliana présente de nombreux avantages pratiques :

- Une taille réduite ;
- Un cycle de végétation rapide ;
- Allofécondation possible après castration ;
- Les croisements sont prolifiques ;
- Génome est simple, : 5 chromosomes. 5500 gènes codants pour des protéines appartenant à 11000 familles. ;
- La taille du génome est relativement petite : 125 Mb ;
- Données de séquençage : moins de la moitié de l'ADN de cette plante est unique ; le reste est copié au moins une fois quelque part dans son génome.

b-Génomique fonctionnelle

- Etude de la fonction des gènes ;
- Partie de la génomique qui étudie la fonction des gènes, leur régulation et les interactions de leurs produits d'expression, ARN et protéines ;
- Analyse la fonction des gènes et autres parties du génome. Elle inclue l'analyse du transcriptome (ARN messagers) ou transcriptomique. Elle contribue aussi très largement à l'annotation des génomes et à l'identification des séquences informatives ;
- L'étude nécessite l'analyse simultanée du transcriptome et du protéome, dans diverses conditions physiologiques et sur divers génotypes sauvages et mutants ainsi que l'intégration des données obtenues ;
- Un ensemble de méthodes permettant d'étudier de nombreux gènes simultanément, pour élucider leurs fonctions. En recherche, cette discipline permet de développer des modèles quantitatifs de cellules et d'organismes ; dans un cadre plus strictement médical, elle facilitera considérablement l'identification de nouvelles cibles pour les médicaments ;
- Une fois les génomes annotés, l'étape suivante sera la recherche de la fonction des séquences informatives identifiées. La génomique fonctionnelle peut être considérée comme de la génétique à « haut débit ». Les techniques utilisées seront comparables mais généralement appliquées à un grand nombre de gènes en parallèle, cela peut par exemple être la création de mutants et l'analyse de leurs phénotypes pour toute une famille de gènes, ou l'analyse de l'expression de tous les gènes d'un organisme entier.

Les objectifs de la génomique fonctionnelle

La génomique fonctionnelle a pour principaux buts de déterminer :

- Le moment dans le cycle cellulaire où un gène est transcrit (expression d'un gène) ;
- Les conditions environnementales liées à la transcription ou la non-transcription d'un gène ;
- L'intensité (nombre de copies du ou des transcrits) avec laquelle ce gène est transcrit ;
- Le compartiment cellulaire où un gène est transcrit (ADN nucléaire, mitochondrial, chloroplastique) ;
- Le compartiment cellulaire où est adressé le (ou les) produit(s) de la transcription d'un gène ;
- Les interactions que le produit d'un gène peut établir avec d'autres produits de gènes et/ou d'autres types de molécules (interactomique). Ce type d'analyse débouche sur la construction de réseaux d'interactions ("*interaction networks*") ;
- Le rôle que peut avoir un **ARN** (interférence ARN / RNAi) dans la régulation post-transcriptionnelle, Si le produit est un **ARN** ;

- Le rôle qu'une **protéine** peut avoir dans une voie métabolique et/ou dans la régulation de cette voie métabolique, Si le produit intermédiaire est un **ARN messenger** et le produit final est donc une **protéine**.

***Attention** : on confond souvent l'expression d'un gène, c'est-à-dire sa transcription (dans le noyau chez les Eucaryotes) et l'expression du produit de ce gène, c'est-à-dire l'activité biologique de ce produit (qui n'est pas destiné systématiquement à rester dans le noyau). Le produit d'un gène est un (ou plusieurs) ARN. Si cet ARN est un ARN messenger, le produit final est une (ou des) protéines.

Plus précisément, la génomique fonctionnelle permet :

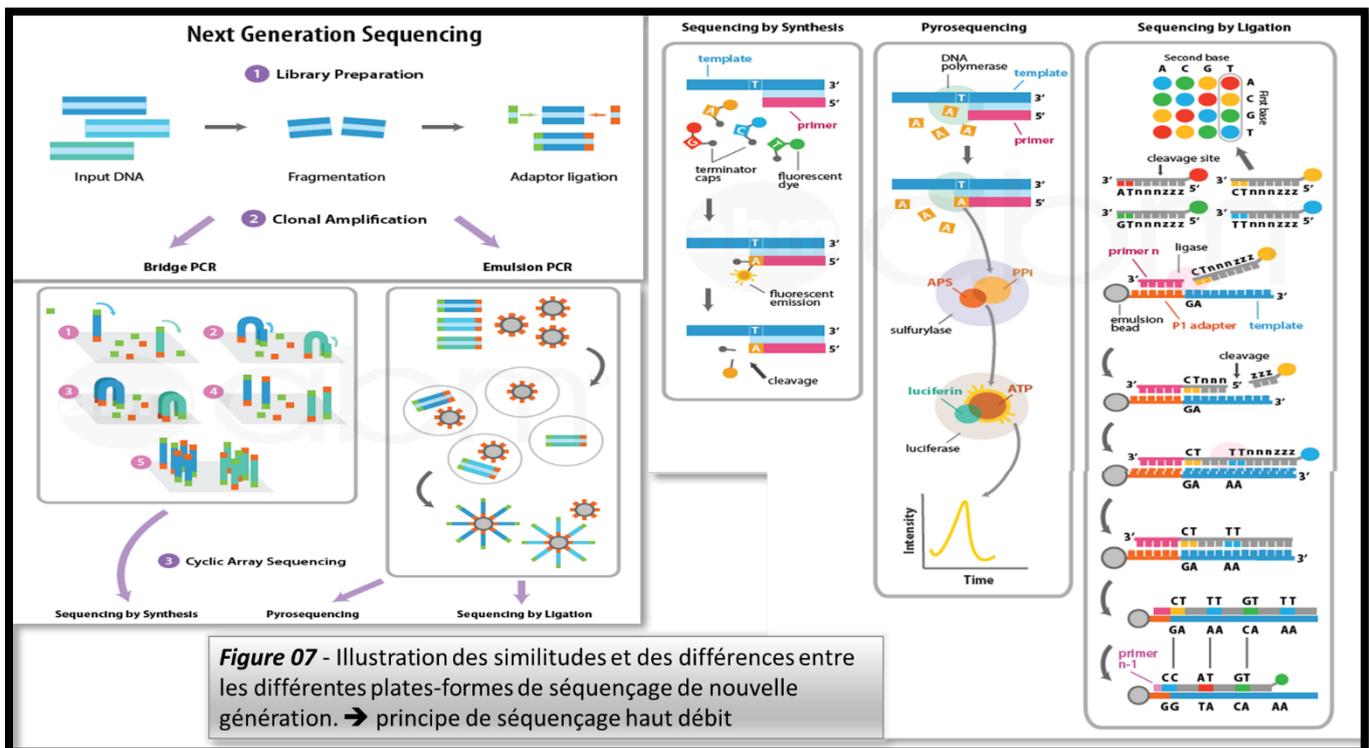
- D'identifier les éléments constitutifs d'un gène (introns, exons, séquences de régulation de la transcription, ...);
- D'identifier les régions des génomes dont on ignore encore le rôle et élucider ce rôle ;
- D'étudier les différences de transcription des gènes dans le temps et pour chaque type de tissus et de cellules ;
- D'étudier les différences d'activité biologique des produits des gènes dans le temps et pour chaque type de tissus, de cellules, de compartiments sub-cellulaires ;
- D'apporter des éléments qui contribuent à déterminer la fonction des ARN et des protéines pour lesquelles les gènes codent ;
- D'intégrer toutes ces informations dans un ensemble plus vaste, celui du métabolisme (métabolome) en décrivant les interactions entre tous ces types de macromolécules biologiques (interactome) ;
- D'obtenir ces données pour le plus grand nombre d'organismes possibles.

Les nouvelles techniques de séquençage en masse (à très haut débit)

Ces techniques ont encore élargi le champ d'investigation de la génomique fonctionnelle. On peut citer :

- Le séquençage de novo ou le reséquençage d'un génome connu ;
- L'étude de la variabilité génétique et du polymorphisme de nucléotide simple (SNP) ;
- Le séquençage d'haplotypes particuliers lors du clonage positionnel d'un gène d'intérêt ;
- L'étude de plus en plus fine du transcriptome :
 - ✓ Identification de transcrits rares, étude des phénomènes d'épissage alternatif, identification des séquences frontières intron/exon, analyse quantitative du niveau de transcription des gènes ;
 - ✓ Étude du profil en petits ARN non codants ("*small ncRNAs*"), découverte de gènes codant ces types d'ARN.

- L'étude des interactions ADN / protéines (régulation de la transcription, facteurs de transcription, ...);
- La génomique médicale;
- La génomique comparative qui compare la structure et les fonctions des génomes de différentes espèces (organisation et évolution des génomes);
- La métagénomique : étude du génome d'un organisme prélevé directement dans un environnement complexe (intestin, océan, sols, ...), à l'inverse d'un organisme de laboratoire. Le but est d'obtenir des informations sur l'incidence de cet environnement. Le préfixe "méta" signifie "après, au-delà de, avec, ...".
- L'épigénétique et l'épigénomique : étude de l'influence de l'environnement et de l'histoire individuelle sur les modifications de l'expression des gènes d'une génération à l'autre. Le préfixe "épi" signifie "sur, au-dessus, ...";
- L'étude du profil de méthylation (processus épigénétique).

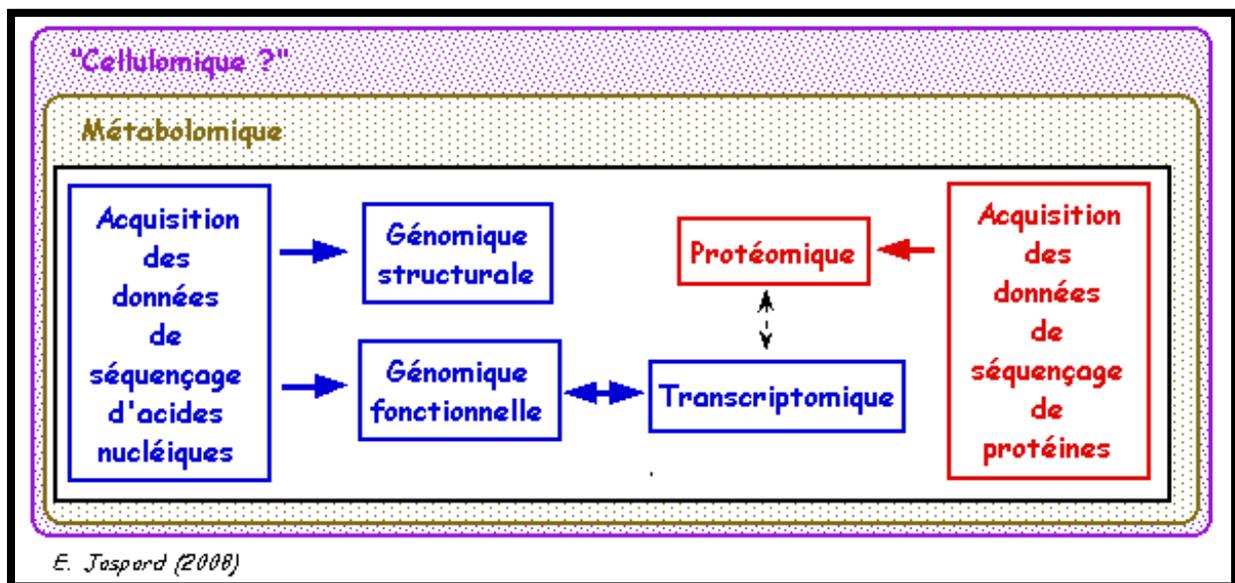




	Pyroséquençage	Termineurs réversibles	Légation
Fournisseur	Roche	Illumina	Applied Biosystems
L'année	2005	2006	2007
Appareils (Nom commercial)	454 GS FLX 454 GS Junior	Génome Analyzer HiSeq MiSeq	SOLiD 5500XL
Longueur des lectures (pb)	400	2x 150	2x 75
Nombre de lectures	1 million	3 milliards	3 milliards
Données produites	600 Mb	600 Gb	600 Gb
Durée du séquençage	10 heures	11 jours	8 jours

Tableau 01 : Présentation des caractéristiques des différentes méthodes de séquençage accessibles en routine aujourd'hui en France.

En résumé : La génomique structurale, la génomique fonctionnelle, la transcriptomique et la protéomique sont des **approches complémentaires**.



3- Annotation des génomes

- L'annotation des génomes est une analyse informatique des séquences obtenues lors du séquençage permettant d'identifier les séquences informatives des génomes. Ces séquences sont principalement les gènes (on parle alors de prédiction de gènes).
- La plupart de ceux-ci sont identifiés soit par leur similitude avec des gènes déjà connus, soit par une prédiction en fonction de la séquence (présence d'un cadre de lecture ouvert caractérisé par un codon d'initiation de la traduction, puis au moins 100 codons et enfin un codon stop).

Mais il existe aussi des « gènes morcelés » (contenant des introns) ou codant des ARN fonctionnels, ceux-ci doivent être prédit par des algorithmes différents.

- Les gènes ne sont pas les seules cibles de l'annotation des génomes, il existe de nombreux autres types de séquences importantes dans les génomes ; les séquences régulatrices, les éléments transposables, etc.

Exemple d'Annotation des génomes de plantes

Une fois la séquence d'un génome obtenue, il faut déchiffrer cette séquence par des approches bio-informatiques

- -Présence de répétitions de séquences.
- -Présence de gènes codant des protéines.
- -Identification de séquences proches de celles de protéines déjà connues.

ATACOU~~LN~~INSERTEURDINSERTE~~SP~~PETA~~INS~~ERTLESGETAR

ATACOU~~LN~~INSERTEURDINSERTE~~SP~~PETA~~INS~~ERTLESGETAR

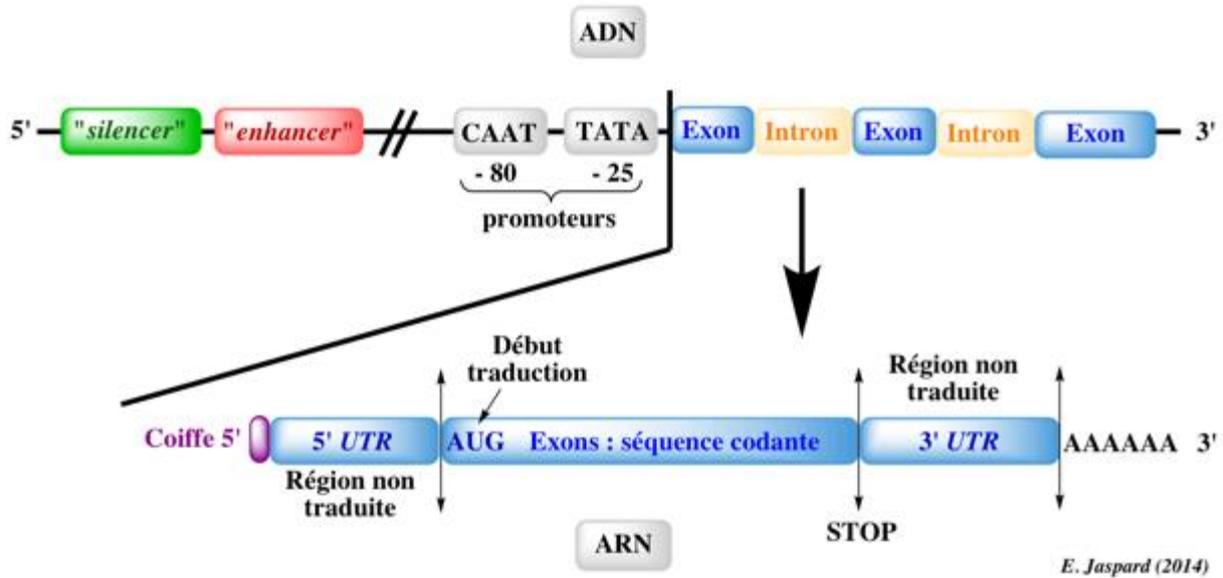
ATACOU~~LN~~INSERTEURDINSERTE~~SP~~PETA~~INS~~ERTLESGETAR

COULEURDESPETALES
| | | | | | | | | | | | | | | |
COULEURDESSEPALES

Etapas de l'annotation

a- Identification des gènes

- Après avoir séquencé un génome, il faut identifier les gènes qu'il contient. Plus précisément, cela consiste à :
 - ✓ Déterminer la structure globale du gène ;
 - ✓ Rechercher les cadres de lecture ouvert ("Open Reading Frame" - ORF) ;
 - ✓ Localiser les motifs de régulation de la transcription des gènes (sites d'initiation ou de terminaison de la transcription, ...) ;
 - ✓ Localiser les sites d'épissage aux bornes exon / intron ;
 - ✓ Localiser les régions codantes ;
 - ✓ Déterminer les pseudogènes (gènes non fonctionnels, des copies inactives de gènes fonctionnels), les éléments transposables (séquences d'ADN capable de se déplacer et de se multiplier de manière autonome dans un génome) ...



b- Annotation des gènes

- Une fois un gène identifié, il faut l'annoter, c'est-à-dire le relier aux maximum de données biologiques (par exemple : données de génétique concernant sa fonction, son expression et les variations phénotypiques des mutants pour la protéine codée, ...).
- En d'autres termes, on tente d'assigner aux molécules pour lesquelles les gènes codent :
 - ✓ Une fonction biologique / biochimique ;
 - ✓ Une localisation sub-cellulaire ;
 - ✓ Leur implication dans des processus de régulation ;
 - ✓ Leur interactions avec d'autres molécules biologiques ;
 - ✓ Un profil de transcription dite "spatio - temporelle" des gènes.
- Des logiciels bioinformatiques sont nécessaires pour l'étude de la structure des gènes, Par exemple :
 - ✓ **BLAST** qui permet d'aligner la séquence du génome avec les séquences d'ADNc ou rechercher des similarités entre ce génome et d'autres génomes déjà connus et annotés ;
 - ✓ **"ORF Finder"** (NCBI) : Entrer le N° d'accession : AK094782.1 - ARNm de la glutamate déshydrogénase ;
 - ✓ **Suite logicielle** pour l'annotation de *Arabidopsis thaliana*.
- Des bases de données regroupent l'ensemble des données biologiques informatives quant à la structure et la fonction des gènes et elles permettent leur annotation :
 - ✓ **Gene Ontology Annotation (GOA) Database** ;
 - ✓ **Rice Genome Annotation Project** ;
 - ✓ **Vertebrate Genome Annotation (VEGA) database** ;

✓ [Ensembl Genome browser](#).

➤ L'ensemble de ces données sont intégrées dans les grandes bases de données biologiques mondiales que sont :

✓ [NCBI / Entrez](#) (National Center for Biotechnology Information) ;

✓ [EMBL - EBI](#) ;

✓ [UniProt / Swissprot](#).

III- La protéomique

➤ C'est l'analyse du protéome (protéines). Elle contribue aussi à l'annotation des génomes et à l'identification des séquences informatives.

➤ La protéomique a pour but d'identifier et quantifier l'ensemble des protéines synthétisées ou protéome, à un moment donné et dans des conditions données au sein d'un tissu, d'une cellule ou d'un compartiment cellulaire.

➤ Le protéome est extrêmement complexe à plusieurs titres :

✓ compte-tenu de l'épissage alternatif des transcrits primaires (plusieurs ARNm pour un gène) et compte-tenu des modifications post-traductionnelles des protéines, on peut estimer à plusieurs dizaines de milliers les formes des protéines synthétisées dans les différents tissus humains par exemple.

✓ Pour chaque condition environnementale (condition physiologique normale vs. conditions de stress) une cellule est caractérisée par un protéome adapté à cette condition alors qu'elle a toujours le même génome. Le cas des plantes est un exemple flagrant compte-tenu de leur nécessité de s'adapter tant aux variations de la lumière qu'aux effets de stress biotiques ou abiotiques.

✓ outre les modifications post-traductionnelles, les protéines subissent des transformations une fois synthétisées : clivage du peptide signal (séquence d'adressage), activation de la forme native à partir d'un précurseur (zymogène), assemblage en complexes oligomériques, association à des cofacteurs.

✓ il existe une grande dynamique de la synthèse des protéines : le rapport entre les protéines les moins abondantes et les plus abondantes dans une cellule dépasse 10^6 pour atteindre 10^{12} dans le sérum.

✓ Les protéines ont des demi-vies très variables : ornithine décarboxylase 11 min - tryptophane oxygénase 2 h - myosine 30 j.

➤ La transcriptomique analyse l'ensemble des transcrits (produits d'expression des gènes). La protéomique et la transcriptomique sont donc des approches complémentaires très puissantes qui peuvent être utilisées pour des études fondamentales ou appliquées en biologie, en médecine, en agriculture.

➤ En effet, dans les deux cas, les informations recueillies permettent d'aborder l'ensemble des réponses cellulaires dans leur globalité et non plus de manière partielle.

➤ La protéomique apporte des réponses auxquelles la transcriptomique ne peut répondre :

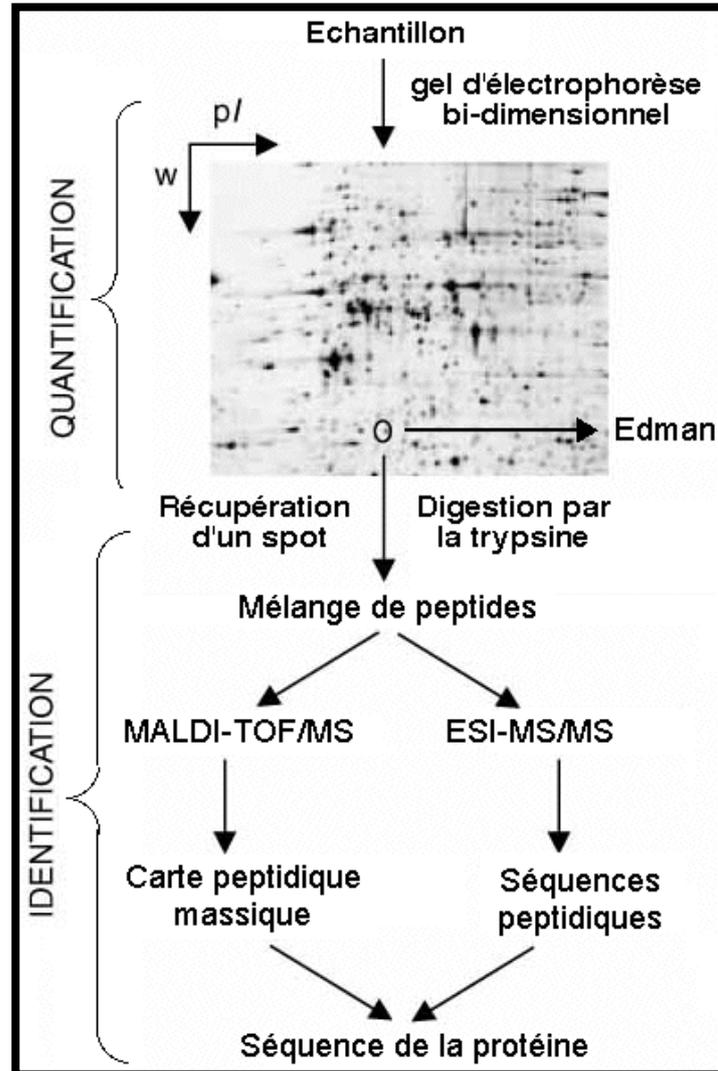
✓ Compléments d'informations sur les modalités d'expression des gènes pour les organismes dont le génome n'a pas encore été séquencé ou pour lesquels les programmes de prédiction

de séquences codantes sont moins fiables. Un exemple est l'aide au repérage des bordures d'exons ce qui permet en retour une meilleure annotation des génomes.

- ✓ Estimation quantitative des concentrations des protéines synthétisées (méthode de marqueurs d'affinité contenant un isotope d'identification : ICAT).
- ✓ Obtention de données sur la fonction des protéines et les interactions entre protéines ou entre protéines et autres molécules biologiques (approche double-hybride ou approche "tandem affinity purification by tag" - TAP/TAG).

Les étapes de l'analyse protéomique

1. L'extraction des protéines ;
2. La séparation des protéines par électrophorèse sur gel bi-dimensionnel ;
3. La révélation des protéines dans les gels puis l'analyse d'image des gels ;
4. La récupération des spots de protéines et la digestion par des protéases ;
5. *le cas échéant, la détermination de la séquence N-terminale par la dégradation d'Edman dans le but de rechercher des candidats dans les banques de données ;*
6. L'obtention de cartes peptidiques massiques par des techniques de spectrométrie de masse ;
7. La détermination de la séquence complète des protéines par d'autres techniques de spectrométrie de masse ;
8. L'analyse bioinformatique (identification des protéines, annotation des protéines et des gènes, recherche de motifs structuraux).



La protéomique et la spectrométrie de masse : différentes stratégies

- Le développement récent de nouveaux spectromètres de masse comme Orbitrap et de nouvelles méthodes de dissociation telles que le transfert d'électrons dissociation (ETD) ont rendu accessible à de nouveaux domaines d'application en protéomique. Bien que la stratégie protéomique **bottom-up (analyse des mélanges de peptides protéolytiques)** soit la plus utilisée par son efficacité et sa simplicité par rapport à la stratégie **top-down (analyse des peptides plus longs ou des protéines intactes)**, mais cette dernière permet une caractérisation plus complète des isoformes de protéines et des modifications post-

traductionnelles. Enfin, les stratégies d'étiquetage des isotopes stables ont transformé la spectrométrie de masse d'un simple outil d'identification à un outil complète pour quantifier et mesurer les changements dynamiques dans l'expression des protéines, l'interaction et la modification.

1- Stratégie Bottom-up en MS/MS

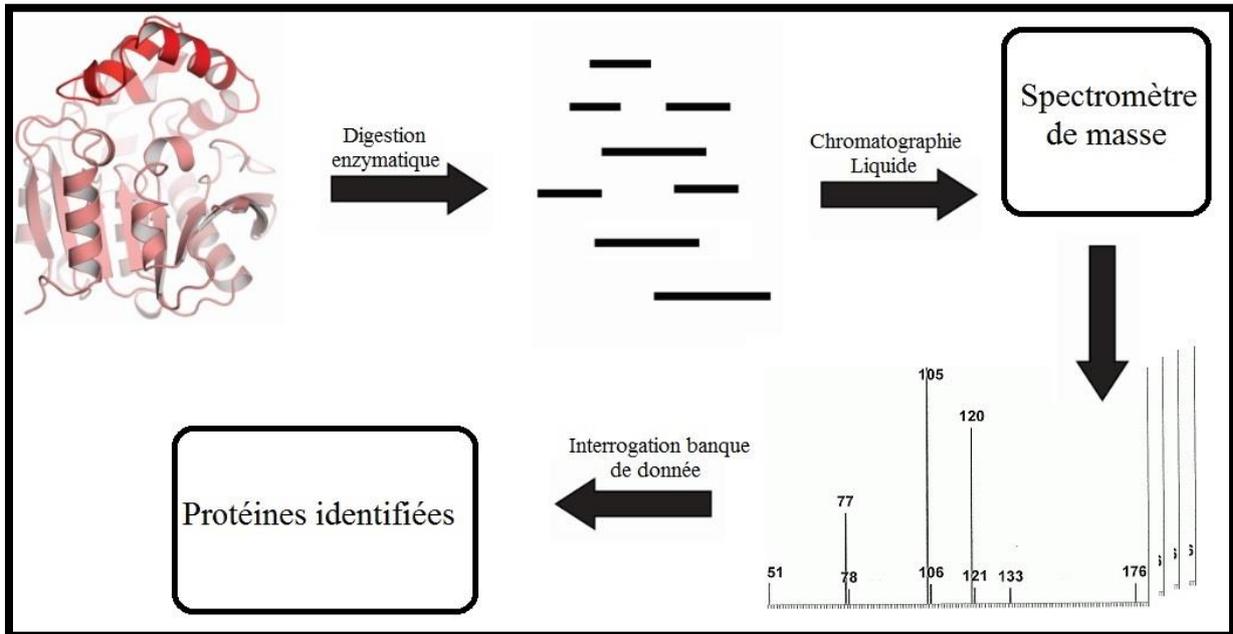


Figure 08. La stratégie bottom-up

- La stratégie bottom-up (Figure 7) est la plus utilisée en protéomique pour identifier des protéines dans les laboratoires de spectrométrie de masse. Les protéines sont d'abord digérées par une enzyme (souvent c'est la trypsine) et on obtient un mélange de peptides.
- Ces peptides sont séparés par la chromatographie liquide avant d'être analysés par la spectrométrie de masse, cela permet de ne pas surcharger le spectromètre de masse.
- Les peptides sont d'abord identifiés par un spectre MS, puis ils sont fragmentés un par un, pour mieux les identifier.
- Généralement le spectromètre de masse scanne un spectre MS pour voir tous les peptides présents à un moment t dans le flux de la chromatographie, ensuite il sélectionne N ions peptidiques et les fragmente un par un.
- Cette méthode est appelée TopN, dont N représente le nombre de spectre MS/MS effectué après un scan MS. Il faut noter que le nombre N est dépendant à la vitesse de scan du spectromètre de masse, si N est élevé (par exemple Top20) et la vitesse de scan n'est pas suffisante, on risque de ne pas attraper les derniers peptides, car pendant que le spectromètre de masse effectue les spectres MS/MS pour les premiers peptides, le flux de la chromatographie continue et les derniers peptides auraient déjà tous partis.
- Les peptides sont identifiés par un logiciel (Protéome discoverer, Maxquant...), à partir de ces peptides identifiés, on peut identifier la ou les protéines, il faut noter qu'il est possible d'identifier une protéine par un seul peptide si c'est un peptide unique.

2- Stratégie top-down en MS/MS

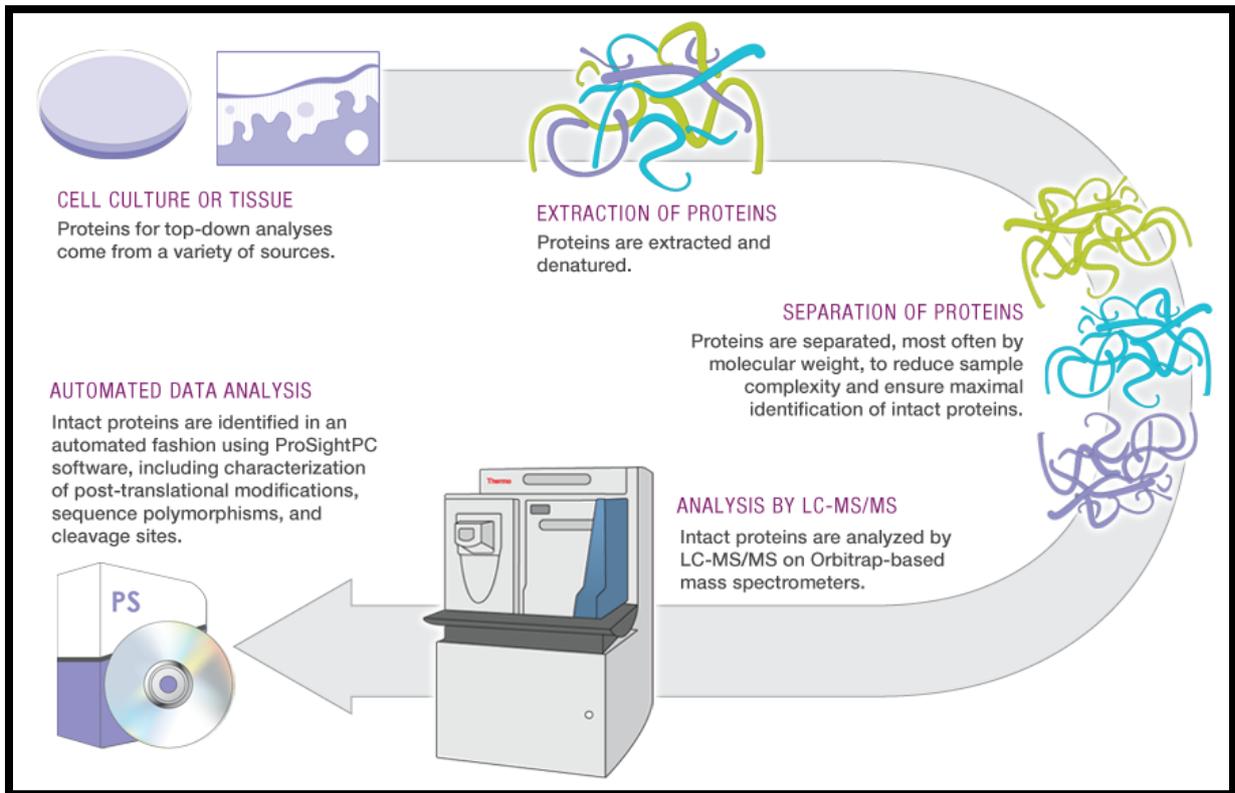


Figure 09. Stratégie top-down en protéomique (Image à partir de planetorbitrap)

- L'identification protéomique en mode top-down consiste à analyser un mélange de protéines intacts sans avoir à digérer (Figure 8), le mélange de protéines est séparé par la chromatographie liquide, puis les protéines sont ionisées et ensuite la protéine à étudier sera isolée puis fragmentée dans l'analyseur.
- La masse sur charge (m/z) de l'ion de protéine et celles de ses ions fils servent à identifier la protéine.
- Les avantages de la stratégie top-down sont la capacité à détecter les produits de dégradation, des variantes de séquence, et des combinaisons de modifications.