



ROYAUME DU MAROC
UNIVERSITE MOHAMMED V DE RABAT
FACULTE DE MEDECINE
ET DE PHARMACIE
RABAT



Année: 2021

Thèse N°: 78

LA PROTEOMIQUE : GENERALITES METHODES ET APPLICATIONS

THESE

Présentée et soutenue publiquement le : / /2021

PAR

Monsieur Younesse BAGACHOUL
Né le 01 Juillet 1995 à Massa (Chtouka Ait Baha)

Pour l'Obtention du Diplôme de
Docteur en Pharmacie

Mots Clés : Protéine; Protéomique; Protéome; Spectrométrie de masse; Electrophorèse

Membres du Jury :

Monsieur M'hammed ANSAR

Professeur de Chimie Organique

Monsieur Azeddine IBRAHIMI

Professeur de Biotechnologie

Monsieur Jaouad EL HARTI

Professeur de Chimie Thérapeutique

Monsieur Youssef RAMLI

Professeur de Chimie

Président

Rapporteur

Juge

Juge

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

قالوا سبحانك لا علم لنا إلا ما علمتنا
إنك أنت العزيز الحكيم

سورة البقرة: الآية: 31

صَلَّى
الْعِظِيمِ



**UNIVERSITE MOHAMMED V
FACULTE DE MEDECINE ET DE PHARMACIE RABAT**

DOYENS HONORAIRES :

1962 - 1969: Professeur Abdelmalek FARAJ
1969 - 1974: Professeur Abdellatif BERBICH
1974 - 1981: Professeur Bachir LAZRAK
1981 - 1989: Professeur Taieb CHKILI
1989 - 1997: Professeur Mohamed Tahar ALAOUI 1997 - 2003: Professeur Abdelmajid BELMAHI
2003 - 2013: Professeur Najia HAJJAJ - HASSOUNI

ADMINISTRATION :

Doyen :

Professeur Mohamed ADNAOUI

Vice-Doyen chargé des Affaires Académiques et estudiantines

Professeur Brahim LEKEHAL

Vice-Doyen chargé de la Recherche et de la Coopération

Professeur Taoufiq DAKKA

Vice-Doyen chargé des Affaires Spécifiques à la Pharmacie

Professeur Younes RAHALI

Secrétaire Général

Mr. Mohamed KARRA

****Enseignant militaire***

1 - ENSEIGNANTS-CHERCHEURS MEDECINS ET PHARMACIENS

PROFESSEURS DE L'ENSEIGNEMENT SUPERIEUR:

Décembre 1984

Pr. MAAOUNI Abdelaziz
Pr. MAAZOUZI Ahmed Wajdi
Pr. SETTAF Abdellatif

Médecine Interne - [Clinique Royale](#)
Anesthésie -Réanimation
Pathologie Chirurgicale

Décembre 1989

Pr. ADNAOUI Mohamed
Pr. OUAZZANI Taïbi Mohamed Réda

Médecine Interne –[Doyen de la EMPR](#)
Neurologie

Janvier et Novembre 1990

Pr. KHARBACH Aïcha
Pr. TAZI Saoud Anas

Gynécologie -Obstétrique
Anesthésie Réanimation

Février Avril Juillet et Décembre 1991

Pr. AZZOUZI Abderrahim
Pr. BAYAHIA Rabéa
Pr. BELKOUCHI Abdelkader
Pr. BENSOUDA Yahia
Pr. BERRAHO Amina
Pr. BEZAD Rachid
Pr. CHERRAH Yahia
Pr. CHOKAIRI Omar
Pr. KHATTAB Mohamed
Pr. SOULAYMANI Rachida
Pr. TAOUFIK Jamal

Anesthésie Réanimation
Néphrologie
Chirurgie Générale
Pharmacie galénique
Ophtalmologie
Gynécologie Obstétrique [Méd. Chef Maternité des Orangers](#)
Pharmacologie
Histologie Embryologie
Pédiatrie
Pharmacologie- [Dir. du Centre National PV Rabat](#)
Chimie thérapeutique

Décembre 1992

Pr. AHALLAT Mohamed
Pr. BENSOUDA Adil
Pr. CHAHED OUAZZANI Laaziza
Pr. CHRAIBI Chafiq
Pr. EL OUAHABI Abdessamad
Pr. FELLAT Rokaya
Pr. JIDDANE Mohamed
Pr. ZOUHDI Mimoun

Chirurgie Générale [Doyen de EMPT](#)
Anesthésie Réanimation
Gastro-Entérologie
Gynécologie Obstétrique
Neurochirurgie
Cardiologie
Anatomie
Microbiologie

Mars 1994

Pr. BENJAAFAR Noureddine
Pr. BEN RAIS Nozha
Pr. CAOUI Malika
Pr. CHRAIBI Abdelmjid
Pr. EL AMRANI Sabah
Pr. ERROUGANI Abdelkader
Pr. ESSAKALI Malika

Radiothérapie
Biophysique
Biophysique
Endocrinologie et Maladies Métaboliques [Doyen de la EMPA](#)
Gynécologie Obstétrique
Chirurgie Générale - [Directeur du CHUIS](#)
Immunologie

**Enseignant militaire*

Pr. ETTAYEBI Fouad
Pr. IFRINE Lahssan
Pr. RHRAB Brahim
Pr. SENOUCI Karima

Chirurgie Pédiatrique
Chirurgie Générale
Gynécologie –Obstétrique
Dermatologie

Mars 1994

Pr. ABBAR Mohamed*
Pr. BENTAHILA Abdelali
Pr. BERRADA Mohamed Saleh
Pr. CHERKAOUI Lalla Ouafae
Pr. LAKHDAR Amina
Pr. MOUANE Nezha

Urologie [Inspecteur du SSM](#)
Pédiatrie
Traumatologie - Orthopédie
Ophtalmologie
Gynécologie Obstétrique
Pédiatrie

Mars 1995

Pr. ABOUQUAL Redouane
Pr. AMRAOUI Mohamed
Pr. BAIDADA Abdelaziz
Pr. BARGACH Samir
Pr. EL MESNAOUI Abbes
Pr. ESSAKALI HOUSSEINI Leila
Pr. IBEN ATTYA ANDALOUSSI Ahmed
Pr. OUAZZANI CHAHDI Bahia
Pr. SEFIANI Abdelaziz
Pr. ZEGGWAGH Amine Ali

Réanimation Médicale
Chirurgie Générale
Gynécologie Obstétrique
Gynécologie Obstétrique
Chirurgie Générale
Oto-Rhino-Laryngologie
Urologie
Ophtalmologie
Génétique
Réanimation Médicale

Décembre 1996

Pr. BELKACEM Rachid
Pr. BOULANOUEAR Abdelkrim
Pr. EL ALAMI EL FARICHA EL Hassan
Pr. GAOUZI Ahmed
Pr. OUZEDDOUN Naima

Chirurgie Pédiatrie
Ophtalmologie
Chirurgie Générale
Pédiatrie
Néphrologie

Pr. ZBIR EL Mehdi*

Cardiologie [Directeur HMI Mohammed V](#)

Novembre 1997

Pr. ALAMI Mohamed Hassan
Pr. BIROUK Nazha
Pr. FELLAT Nadia
Pr. KADDOURI Nouredine
Pr. KOUTANI Abdellatif
Pr. LAHLOU Mohamed Khalid
Pr. MAHRAOUI CHAFIQ
Pr. TOUFIQ Jallal
Pr. YOUSFI MALKI Mounia

Gynécologie-Obstétrique
Neurologie
Cardiologie
Chirurgie Pédiatrique
Urologie
Chirurgie Générale
Pédiatrie
Psychiatrie [Directeur Hôp.Ar-razi Salé](#)
Gynécologie Obstétrique

Novembre 1998

Pr. BENOMAR ALI
Pr. BOUGTAB Abdesslam
Pr. ER RIHANI Hassan

Neurologie [Doyen de la FM Abulcassis](#)
Chirurgie Générale
Oncologie Médicale

****Enseignant militaire***

Pr. BENKIRANE Majid*

Hématologie

Janvier 2000

Pr. ABID Ahmed*
Pr. AIT OUAMAR Hassan
Pr. BENJELLOUN Dakhama Badr Sououd
Pr. BOURKADI Jamal-Eddine
Pr. CHARIF CHEFCHAOUNI Al Montacer
Pr. ECHARRAB El Mahjoub
Pr. EL FTOUH Mustapha
Pr. EL MOSTARCHID Brahim*
Pr. TACHINANTE Rajae
Pr. TAZI MEZALEK Zoubida

Pneumo-phtisiologie
Pédiatrie
Pédiatrie
Pneumo-phtisiologie
Chirurgie Générale
Chirurgie Générale
Pneumo-phtisiologie
Neurochirurgie
Anesthésie-Réanimation
Médecine Interne

Novembre 2000

Pr. AIDI Saadia
Pr. AJANA Fatima Zohra
Pr. BENAMR Said
Pr. CHERTI Mohammed
Pr. ECH-CHERIF EL KETTANI Selma
Pr. EL HASSANI Amine
Pr. EL KHADER Khalid
Pr. GHARBI Mohamed El Hassan
Pr. MDAGHRI ALAOUI Asmae

Neurologie
Gastro-Entérologie
Chirurgie Générale
Cardiologie
Anesthésie-Réanimation
Pédiatrie - [Directeur Hôp. Cheikh Zaid](#)
Urologie
Endocrinologie et Maladies Métaboliques
Pédiatrie

Décembre 2001

Pr. BALKHI Hicham*
Pr. BENABDELJLIL Maria
Pr. BENAMAR Loubna
Pr. BENAMOR Jouda
Pr. BENELBARHDADI Imane
Pr. BENNANI Rajae
Pr. BENOUACHANE Thami
Pr. BEZZA Ahmed*
Pr. BOUCHIKHI IDRISSE Med Larbi
Pr. BOUMDIN El Hassane*
Pr. CHAT Latifa
Pr. EL HJRI Ahmed
Pr. EL MAAQILI Moulay Rachid
Pr. EL MADHI Tarik
Pr. EL OUNANI Mohamed
Pr. ETTAIR Said

Anesthésie-Réanimation
Neurologie
Néphrologie
Pneumo-phtisiologie
Gastro-Entérologie
Cardiologie
Pédiatrie
Rhumatologie
Anatomie
Radiologie
Radiologie
Anesthésie-Réanimation
Neuro-Chirurgie
Chirurgie-Pédiatrique [Directeur Hôp. Des Enfants Rabat](#)
Chirurgie Générale
Pédiatrie - [Directeur Hôp. Univ. International \(Cheikh Khalifa\)](#)

Pr. GAZZAZ Miloudi*
Pr. HRORA Abdelmalek
Pr. KABIRI EL Hassane*
Pr. LAMRANI Moulay Omar
Pr. LEKEHAL Brahim
Pr. MEDARHRI Jalil

Neuro-Chirurgie
Chirurgie Générale [Directeur Hôpital Ibn Sina](#)
Chirurgie Thoracique
Traumatologie Orthopédie
Chirurgie Vasculaire Périphérique [V-D chargé Aff Acad. Est.](#)
Chirurgie Générale

****Enseignant militaire***

Pr. MIKDAME Mohammed*
Pr. MOHSINE Raouf
Pr. NOUINI Yassine
Pr. SABBAH Farid
Pr. SEFIANI Yasser
Pr. TAOUFIQ BENCHEKROUN Soumia

Hématologie Clinique
Chirurgie Générale
Urologie
Chirurgie Générale
Chirurgie Vasculaire Périphérique
Pédiatrie

Décembre 2002

Pr. AMEUR Ahmed*
Pr. AMRI Rachida
Pr. AOURARH Aziz*
Pr. BAMOU Youssef*
Pr. BELMEJDOUB Ghizlene*
Pr. BENZEKRI Laila
Pr. BENZZOUBEIR Nadia
Pr. BERNOUSSI Zakiya
Pr. CHOHO Abdelkrim*
Pr. CHKIRATE Bouchra
Pr. EL ALAMI EL Fellous Sidi Zouhair
Pr. FILALI ADIB Abdelhai
Pr. HAJJI Zakia
Pr. KRIOUILE Yamina
Pr. OUJILAL Abdelilah
Pr. RAISS Mohamed
Pr. SIAH Samir*
Pr. THIMOU Amal
Pr. ZENTAR Aziz*

Urologie
Cardiologie
Gastro-Entérologie
Biochimie-Chimie
Endocrinologie et Maladies Métaboliques
Dermatologie
Gastro-Entérologie
Anatomie Pathologique
Chirurgie Générale
Pédiatrie
Chirurgie Pédiatrique
Gynécologie Obstétrique
Ophtalmologie
Pédiatrie
Oto-Rhino-Laryngologie
Chirurgie Générale
Anesthésie Réanimation
Pédiatrie
Chirurgie Générale

Janvier 2004

Pr. ABDELLAH El Hassan
Pr. AMRANI Mariam
Pr. BENBOUZID Mohammed Anas
Pr. BENKIRANE Ahmed*
Pr. BOULAADAS Malik
Pr. BOURAZZA Ahmed*
Pr. CHAGAR Belkacem*
Pr. CHERRADI Nadia
Pr. EL FENNI Jamal*
Pr. EL HANCHI ZAKI
Pr. EL KHORASSANI Mohamed
Pr. HACHI Hafid
Pr. JABOUIRIK Fatima
Pr. KHARMAZ Mohamed
Pr. MOUGHIL Said
Pr. OUBAAZ Abdelbarre*
Pr. TARIB Abdelilah*
Pr. TIJAMI Fouad
Pr. ZARZUR Jamila

Ophtalmologie
Anatomie Pathologique
Oto-Rhino-Laryngologie
Gastro-Entérologie
Stomatologie et Chirurgie Maxillo-faciale
Neurologie
Traumatologie Orthopédie
Anatomie Pathologique
Radiologie
Gynécologie Obstétrique
Pédiatrie
Chirurgie Générale
Pédiatrie
Traumatologie Orthopédie
Chirurgie Cardio-Vasculaire
Ophtalmologie
Pharmacie Clinique
Chirurgie Générale
Cardiologie

****Enseignant militaire***

Janvier 2005

Pr. ABBASSI Abdellah
Pr. AL KANDRY Sif Eddine*
Pr. ALLALI Fadoua
Pr. AMAZOUZI Abdellah
Pr. BAHIRI Rachid
Pr. BARKAT Amina
Pr. BENYASS Aatif*
Pr. DOUDOUH Abderrahim*
Pr. HAJJI Leila
Pr. HESSISSEN Leila
Pr. JIDAL Mohamed*
Pr. LAAROUSSI Mohamed
Pr. LYAGOUBI Mohammed
Pr. SBIHI Souad
Pr. ZERAIDI Najia

Chirurgie Réparatrice et Plastique
Chirurgie Générale
Rhumatologie
Ophtalmologie
Rhumatologie [Directeur Hôp. Al Ayachi Salé](#)
Pédiatrie
Cardiologie
Biophysique
Cardiologie (mise en disponibilité)
Pédiatrie
Radiologie
Chirurgie Cardio-vasculaire
Parasitologie
Histo-Embryologie Cytogénétique
Gynécologie Obstétrique

AVRIL 2006

Pr. ACHEMLAL Lahsen*
Pr. BELMEKKI Abdelkader*
Pr. BENCHEIKH Razika
Pr. BOUHAFS Mohamed El Amine
Pr. BOULAHYA Abdellatif*
Pr. CHENGUETI ANSARI Anas
Pr. DOGHMI Nawal
Pr. FELLAT Ibtissam
Pr. FAROUDY Mamoun
Pr. HARMOUCHE Hicham
Pr. IDRIS LAHLOU Amine*
Pr. JROUNDI Laila
Pr. KARMOUNI Tariq
Pr. KILI Amina
Pr. KISRA Hassan
Pr. KISRA Mounir
Pr. LAATIRIS Abdelkader*
Pr. LMIMOUNI Badreddine*
Pr. MANSOURI Hamid*
Pr. OUANASS Abderrazzak
Pr. SAFI Soumaya*
Pr. SOUALHI Mouna
Pr. TELLAL Saida*
Pr. ZAHRAOUI Rachida

Rhumatologie
Hématologie
O.R.L
Chirurgie - Pédiatrique
Chirurgie Cardio - Vasculaire. [Directeur Hôpital Ibn Sina Marr.](#)
Gynécologie Obstétrique
Cardiologie
Cardiologie
Anesthésie Réanimation
Médecine Interne
Microbiologie
Radiologie
Urologie
Pédiatrie
Psychiatrie
Chirurgie - Pédiatrique
Pharmacie Galénique
Parasitologie
Radiothérapie
Psychiatrie
Endocrinologie
Pneumo - Phtisiologie
Biochimie
Pneumo - Phtisiologie

Octobre 2007

Pr. ABIDI Khalid
Pr. ACHACHI Leila
Pr. AMHAJJI Larbi*
Pr. AOUIFI Sarra
Pr. BAITE Abdelouahed*

Réanimation médicale
Pneumo phtisiologie
Traumatologie orthopédie
Parasitologie
Anesthésie réanimation

****Enseignant militaire***

Pr. BALOUCH Lhousaine*
 Pr. BENZIANE Hamid*
 Pr. BOUTIMZINE Nourdine
 Pr. CHERKAoui Naoual*
 Pr. EL BEKKALI Youssef*
 Pr. EL ABSI Mohamed
 Pr. EL MOUSSAOUI Rachid
 Pr. EL OMARI Fatima
 Pr. GHARIB Nouredine
 Pr. HADADI Khalid*
 Pr. ICHOU Mohamed*
 Pr. ISMAILI Nadia
 Pr. KEBDANI Tayeb
 Pr. LOUZI Lhoussain*
 Pr. MADANI Naoufel
 Pr. MARC Karima
 Pr. MASRAR Azlarab
 Pr. OUZZIF Ez zohra*
 Pr. SEFFAR Myriame
 Pr. SEKHSOKH Yessine*
 Pr. SIFAT Hassan*
 Pr. TACHFOUTI Samira
 Pr. TAJDINE Mohammed Tariq*
 Pr. TANANE Mansour*
 Pr. TLIGUI Houssain
 Pr. TOUATI Zakia

Mars 2009

Pr. ABOUZAHER Ali*
 Pr. AGADR Aomar*
 Pr. AIT ALI Abdelmounaim*
 Pr. AKHADDAR Ali*
 Pr. ALLALI Nazik
 Pr. AMINE Bouchra
 Pr. ARKHA Yassir
 Pr. BELYAMANI Lahcen*
 Pr. BJIJOU Younes
 Pr. BOUHSAIN Sanae*
 Pr. BOUI Mohammed*
 Pr. BOUNAIM Ahmed*
 Pr. BOUSSOUGA Mostapha*
 Pr. CHTATA Hassan Toufik*
 Pr. DOGHMI Kamal*
 Pr. EL MALKI Hadj Omar
 Pr. EL OUENNASS Mostapha*
 Pr. ENNIBI Khalid*
 Pr. FATHI Khalid
 Pr. HASSIKOU Hasna*
 Pr. KABBAJ Nawal

Biochimie-chimie
 Pharmacie clinique
 Ophtalmologie
 Pharmacie galénique
 Chirurgie cardio-vasculaire
 Chirurgie générale
 Anesthésie réanimation
 Psychiatrie
 Chirurgie plastique et réparatrice
 Radiothérapie
 Oncologie médicale
 Dermatologie
 Radiothérapie
 Microbiologie
 Réanimation médicale
 Pneumo phtisiologie
 Hématologie biologique
 Biochimie-chimie
 Microbiologie
 Microbiologie
 Radiothérapie
 Ophtalmologie
 Chirurgie générale
 Traumatologie-orthopédie
 Parasitologie
 Cardiologie

Médecine interne
 Pédiatrie
 Chirurgie Générale
 Neuro-chirurgie
 Radiologie
 Rhumatologie
 Neuro-chirurgie [Directeur Hôp.des Spécialités](#)
 Anesthésie Réanimation
 Anatomie
 Biochimie-chimie
 Dermatologie
 Chirurgie Générale
 Traumatologie-orthopédie
 Chirurgie Vasculaire Périphérique
 Hématologie clinique
 Chirurgie Générale
 Microbiologie
 Médecine interne
 Gynécologie obstétrique
 Rhumatologie
 Gastro-entérologie

****Enseignant militaire***

Pr. KABIRI Meryem
 Pr. KARBOUBI Lamya
 Pr. LAMSAOURI Jamal*
 Pr. MARMADE Lahcen
 Pr. MESKINI Toufik
 Pr. MESSAOUDI Nezha*
 Pr. MSSROURI Rahal
 Pr. NASSAR Ittimade
 Pr. OUKERRAJ Latifa
 Pr. RHORFI Ismail Abderrahmani*

Pédiatrie
 Pédiatrie
 Chimie Thérapeutique
 Chirurgie Cardio-vasculaire
 Pédiatrie
 Hématologie biologique
 Chirurgie Générale
 Radiologie
 Cardiologie
 Pneumo-Phtisiologie

Octobre 2010

Pr. ALILOU Mustapha
 Pr. AMEZIANE Taoufiq*
 Pr. BELAGUID Abdelaziz
 Pr. CHADLI Mariama*
 Pr. CHEMSI Mohamed*
 Pr. DAMI Abdellah*
 Pr. DARBI Abdellatif*
 Pr. DENDANE Mohammed Anouar
 Pr. EL HAFIDI Naima
 Pr. EL KHARRAS Abdennasser*
 Pr. EL MAZOUZ Samir
 Pr. EL SAYEGH Hachem
 Pr. ERRABIH Ikram
 Pr. LAMALMI Najat
 Pr. MOSADIK Ahlam
 Pr. MOUJAHID Mountassir*
 Pr. ZOUAIDIA Fouad

Anesthésie réanimation
 Médecine Interne **Directeur ERSSM**
 Physiologie
 Microbiologie
 Médecine Aéronautique
 Biochimie- Chimie
 Radiologie
 Chirurgie Pédiatrique
 Pédiatrie
 Radiologie
 Chirurgie Plastique et Réparatrice
 Urologie
 Gastro-Entérologie
 Anatomie Pathologique
 Anesthésie Réanimation
 Chirurgie Générale
 Anatomie Pathologique

Decembre 2010

Pr. ZNATI Kaoutar

Anatomie Pathologique

Mai 2012

Pr. AMRANI Abdelouahed
 Pr. ABOUELALAA Khalil*
 Pr. BENCHEBBA Driss*
 Pr. DRISSI Mohamed*
 Pr. EL ALAOUI MHAMDI Mouna
 Pr. EL OUAZZANI Hanane*
 Pr. ER-RAJI Mounir
 Pr. JAHID Ahmed

Chirurgie pédiatrique
 Anesthésie Réanimation
 Traumatologie-orthopédie
 Anesthésie Réanimation
 Chirurgie Générale
 Pneumophtisiologie
 Chirurgie Pédiatrique
 Anatomie Pathologique

Février 2013

Pr. AHID Samir
 Pr. AIT EL CADI Mina
 Pr. AMRANI HANCHI Laila
 Pr. AMOR Mourad
 Pr. AWAB Almahdi

Pharmacologie
 Toxicologie
 Gastro-Entérologie
 Anesthésie-Réanimation
 Anesthésie-Réanimation

****Enseignant militaire***

Pr. BELAYACHI Jihane	Réanimation Médicale
Pr. BELKHADIR Zakaria Houssain	Anesthésie-Réanimation
Pr. BENCHEKROUN Laila	Biochimie-Chimie
Pr. BENKIRANE Souad	Hématologie
Pr. BENSGHIR Mustapha*	Anesthésie Réanimation
Pr. BENYAHIA Mohammed*	Néphrologie
Pr. BOUATIA Mustapha	Chimie Analytique et Bromatologie
Pr. BOUABID Ahmed Salim*	Traumatologie orthopédie
Pr. BOUTARBOUCH Mahjouba	Anatomie
Pr. CHAIB Ali*	Cardiologie
Pr. DENDANE Tarek	Réanimation Médicale
Pr. DINI Nouzha*	Pédiatrie
Pr. ECH-CHERIF EL KETTANI Mohamed Ali	Anesthésie Réanimation
Pr. ECH-CHERIF EL KETTANI Najwa	Radiologie
Pr. ELFATEMI NIZARE	Neuro-chirurgie
Pr. EL GUERROUJ Hasnae	Médecine Nucléaire
Pr. EL HARTI Jaouad	Chimie Thérapeutique
Pr. EL JAOUDI Rachid*	Toxicologie
Pr. EL KABABRI Maria	Pédiatrie
Pr. EL KHANNOUSSI Basma	Anatomie Pathologique
Pr. EL KHLouFI Samir	Anatomie
Pr. EL KORAICHI Alae	Anesthésie Réanimation
Pr. EN-NOUALI Hassane*	Radiologie
Pr. ERRGUIG Laila	Physiologie
Pr. FIKRI Meryem	Radiologie
Pr. GHFIR Imade	Médecine Nucléaire
Pr. IMANE Zineb	Pédiatrie
Pr. IRAQI Hind	Endocrinologie et maladies métaboliques
Pr. KABBAJ Hakima	Microbiologie
Pr. KADIRI Mohamed*	Psychiatrie
Pr. LATIB Rachida	Radiologie
Pr. MAAMAR Mouna Fatima Zahra	Médecine Interne
Pr. MEDDAH Bouchra	Pharmacologie
Pr. MELHAOUI Adyl	Neuro-chirurgie
Pr. MRABTI Hind	Oncologie Médicale
Pr. NEJJARI Rachid	Pharmacognosie
Pr. OUBEJJA Houda	Chirurgie Pédiatrique
Pr. OUKABLI Mohamed*	Anatomie Pathologique
Pr. RAHALI Younes	Pharmacie Galénique Vice-Doyen à la Pharmacie
Pr. RATBI Ilham	Génétique
Pr. RAHMANI Mounia	Neurologie
Pr. REDA Karim*	Ophtalmologie
Pr. REGRAGUI Wafa	Neurologie
Pr. RKAIN Hanan	Physiologie
Pr. ROSTOM Samira	Rhumatologie
Pr. ROUAS Lamiaa	Anatomie Pathologique
Pr. ROUIBAA Fedoua*	Gastro-Entérologie
Pr. SALIHOUN Mouna	Gastro-Entérologie
Pr. SAYAH Rochde	Chirurgie Cardio-Vasculaire

***Enseignant militaire**

Pr. SEDDIK Hassan*
Pr. ZERHOUNI Hicham
Pr. ZINE Ali*

Gastro-Entérologie
Chirurgie Pédiatrique
Traumatologie Orthopédie

AVRIL 2013

Pr. EL KHATIB MOHAMED KARIM*

Stomatologie et Chirurgie Maxillo-faciale

MARS 2014

Pr. ACHIR Abdellah
Pr. BENCHAKROUN Mohammed*
Pr. BOUCHIKH Mohammed
Pr. EL KABBAJ Driss*
Pr. EL MACHTANI IDRISSE Samira*
Pr. HARDIZI Houyam
Pr. HASSANI Amale*
Pr. HERRAK Laila
Pr. JEAIDI Anass*
Pr. KOUACH Jaouad*
Pr. MAKRAM Sanaa*
Pr. RHISSASSI Mohamed Jaafar
Pr. SEKKACH Youssef*
Pr. TAZI MOUKHA Zakia

Chirurgie Thoracique
Traumatologie- Orthopédie
Chirurgie Thoracique
Néphrologie
Biochimie-Chimie
Histologie- Embryologie-Cytogénétique
Pédiatrie
Pneumologie
Hématologie Biologique
Génécologie-Obstétrique
Pharmacologie
CCV
Médecine Interne
Génécologie-Obstétrique

DECEMBRE 2014

Pr. ABILKACEM Rachid*
Pr. AIT BOUGHIMA Fadila
Pr. BEKKALI Hicham*
Pr. BENAZZOU Salma
Pr. BOUABDELLAH Mounya
Pr. BOUCHRIK Mourad*
Pr. DERRAJI Soufiane*
Pr. EL AYOUBI EL IDRISSE Ali
Pr. EL GHADBANE Abdedaim Hatim*
Pr. EL MARJANY Mohammed*
Pr. FEJJAL Nawfal
Pr. JAHIDI Mohamed*
Pr. LAKHAL Zouhair*
Pr. OUDGHIRI NEZHA
Pr. RAMI Mohamed
Pr. SABIR Maria
Pr. SBAI IDRISSE Karim*

Pédiatrie
Médecine Légale
Anesthésie-Réanimation
Chirurgie Maxillo-Faciale
Biochimie-Chimie
Parasitologie
Pharmacie Clinique
Anatomie
Anesthésie-Réanimation
Radiothérapie
Chirurgie Réparatrice et Plastique
O.R.L
Cardiologie
Anesthésie-Réanimation
Chirurgie Pédiatrique
Psychiatrie
Médecine préventive, santé publique et Hyg.

AOUT 2015

Pr. MEZIANE Meryem
Pr. TAHIRI Latifa

Dermatologie
Rhumatologie

**Enseignant militaire*

PROFESSEURS AGREGES :

JANVIER 2016

Pr. BENKABBOU Amine
Pr. EL ASRI Fouad*
Pr. ERRAMI Nouredine*
Pr. NITASSI Sophia

Chirurgie Générale
Ophtalmologie
O.R.L
O.R.L

JUIN 2017

Pr. ABI Rachid*
Pr. ASFALOU Ilyasse*
Pr. BOUAITI El Arbi*
Pr. BOUTAYEB Saber
Pr. EL GHISSASSI Ibrahim
Pr. HAFIDI Jawad
Pr. MAJBAR Mohammed Anas
Pr. OURAINI Saloua*
Pr. RAZINE Rachid
Pr. SOUADKA Amine
Pr. ZRARA Abdelhamid*

Microbiologie
Cardiologie
Médecine préventive, santé publique et Hyg.
Oncologie Médicale
Oncologie Médicale
Anatomie
Chirurgie Générale
O.R.L
Médecine préventive, santé publique et Hyg.
Chirurgie Générale
Immunologie

MAI 2018

Pr. AMMOURI Wafa
Pr. BENTALHA Aziza
Pr. EL AHMADI Brahim
Pr. EL HARRECH Youness*
Pr. EL KACEMI Hanan
Pr. EL MAJJAOUI Sanaa
Pr. FATIHI Jamal*
Pr. GHANNAM Abdel-Ilah
Pr. JROUNDI Imane
Pr. MOATASSIM BILLAH Nabil
Pr. TADILI Sidi Jawad
Pr. TANZ Rachid*

Médecine interne
Anesthésie-Réanimation
Anesthésie-Réanimation
Urologie
Radiothérapie
Radiothérapie
Médecine Interne
Anesthésie-Réanimation
Médecine préventive, santé publique et Hyg.
Radiologie
Anesthésie-Réanimation
Oncologie Médicale

NOVEMBRE 2018

Pr. AMELLAL Mina
Pr. SOULY Karim
Pr. TAHRI Rajae

Anatomie
Microbiologie
Histologie-Embryologie-Cytogénétique

NOVEMBRE 2019

Pr. AATIF Taoufiq*
Pr. ACHBOUK Abdelhafid*
Pr. ANDALOUSSI SAGHIR Khalid
Pr. BABA HABIB Moulay Abdellah*
Pr. BASSIR RIDA ALLAH
Pr. BOUATTAR TARIK
Pr. BOUFETTAL MONSEF
Pr. BOUCHENTOUF Sidi Mohammed*

Néphrologie
Chirurgie réparatrice et plastique
Radiothérapie
Gynécologie-Obstétrique
Anatomie
Néphrologie
Anatomie
Chirurgie-Générale

**Enseignant militaire*

Pr. BOUZELMAT HICHAM*
 Pr. BOUKHRIS JALAL*
 Pr. CHAFRY BOUCHAIB*
 Pr. CHAHDI HAFSA*
 Pr. CHERIF EL ASRI ABAD*
 Pr. DAMIRI AMAL*
 Pr. DOGHMI NAWFAL*
 Pr. ELALAOUI SIDI-YASSIR
 Pr. EL ANNAZ HICHAM*
 Pr. EL HASSANI MOULAY EL MEHDI*
 Pr. EL HJOUJI ABDERRAHMAN*
 Pr. EL KAOUI HAKIM*
 Pr. EL WALI ABDERRAHMAN*
 Pr. EN-NAFAA ISSAM*
 Pr. HAMAMA JALAL*
 Pr. HEMMAOUI BOUCHAIB*
 Pr. HJIRA NAOUFAL*
 Pr. JIRA MOHAMED*
 Pr. JNIENE ASMAA
 Pr. LARAQUI HICHAM*
 Pr. MAHFOUD TARIK*
 Pr. MEZIANE MOHAMMED*
 Pr. MOUTAKI ALLAH YOUNES*
 Pr. MOUZARI YASSINE*
 Pr. NAOUI HAFIDA*
 Pr. OBTEL MAJDOULINE
 Pr. OURRAI ABDELHAKIM*
 Pr. SAOUAB RACHIDA*
 Pr. SBITTI YASSIR*
 Pr. ZADDOUG OMAR*
 Pr. ZIDOUH SAAD*

Cardiologie
 Traumatologie-Orthopédie
 Traumatologie-Orthopédie
 Anatomie pathologique
 Neuro-chirurgie
 Anatomie Pathologique
 Anesthésie-Réanimation
 Pharmacie-Galénique
 Virologie
 Gynécologie-Obstétrique
 Chirurgie Générale
 Chirurgie Générale
 Anesthésie-Réanimation
 Radiologie
 Stomatologie et Chirurgie Maxillo-faciale
 O.R.L
 Dermatologie
 Médecine interne
 Physiologie
 Chirurgie-Générale
 Oncologie Médicale
 Anesthésie-Réanimation
 Chirurgie Cardio-Vasculaire
 Ophtalmologie
 Parasitologie-Mycologie
 Médecine préventive, santé publique et Hyg.
 Pédiatrie
 Radiologie
 Oncologie Médicale
 Traumatologie-Orthopédie
 Anesthésie-Réanimation

***Enseignant militaire**

2 - ENSEIGNANTS-CHERCHEURS SCIENTIFIQUES

PROFESSEURS DE L'ENSEIGNEMENT SUPERIEUR :

Pr. ABOUDRAR Saadia	Physiologie
Pr. ALAMI OUHABI Naima	Biochimie-chimie
Pr. ALAOUI KATIM	Pharmacologie
Pr. ALAOUI SLIMANI Lalla Naïma	Histologie-Embryologie
Pr. ANSAR M'hammed	Chimie Organique et Pharmacie Chimique
Pr. BARKIYOU Malika	Histologie-Embryologie
Pr. BOUHOUCHE Ahmed	Génétique Humaine
Pr. BOUKLOUZE Abdelaziz	Applications Pharmaceutiques
Pr. DAKKA Taoufiq	Physiologie Vice-Doyen chargé de la Rech. et de la Coop.
Pr. FAOUZI Moulay El Abbès	Pharmacologie
Pr. IBRAHIMI Azeddine	Biologie moléculaire/Biotechnologie
Pr. OULAD BOUYAHYA IDRISSE Mohammed	Chimie Organique
Pr. RIDHA Ahlam	Chimie
Pr. TOUATI Driss	Pharmacognosie
Pr. ZAHIDI Ahmed	Pharmacologie

PROFESSEURS HABILITES :

Pr. BENZEID Hanane	Chimie
Pr. CHAHED OUAZZANI Lalla Chadia	Biochimie-chimie
Pr. DOUKKALI Anass	Chimie Analytique
Pr. EL JASTIMI Jamila	Chimie
Pr. KHANFRI Jamal Eddine	Histologie-Embryologie
Pr. LYAHYAI Jaber	Génétique
Pr. OUADGHIRI Mouna	Microbiologie et Biologie
Pr. RAMLI Youssef	Chimie
Pr. SERRAGUI Samira	Pharmacologie
Pr. TAZI Ahnini	Génétique
Pr. YAGOUBI Maamar	Eau, Environnement

Mise à jour le 05/03/2021

KHALED Abdellah

Chef du Service des Ressources Humaines

FMPR

***Enseignant militaire**

DEDICACES





A mes parents

*Aucun texte ou dédicace ne saurait exprimer l'amour et la
profonde reconnaissance que je vous porte.*

*Puisse Dieu le tout puissant vous protéger de tout mal et vous
procurer une bonne santé et une longue vie pleine de bonheur.
Merci pour votre amour, et votre amitié. Vous étiez toujours là
pour me soutenir et m'écouter. La réalisation de ce travail est le
fruit de vos vœux et prières.*

*Je tiens à exprimer mes plus vifs remerciements à ma famille,
mon père et ma mère bien-aimés, qui m'ont aidé à surmonter les
difficultés et m'ont soutenu tout au long de ma vie.*



A ma sœur Farida

Je ne peux exprimer à travers ses lignes tous mes sentiments d'amour et de gratitude envers toi. Puisse Dieu vous assister en toute chose et vous protège et puisse l'amour et la fraternité nous unissent éternellement.



A mes chers/chères ami(e)s

*Hamid, Bouchra, Achraf, Mohamed, Amine, Youssef, Soukaina,
Nabil, Mohamed, Lamiaa, Omar et tous les autres.*

*En souvenir de notre profonde amitié et des moments agréables
que nous avons passés ensemble.*

*Merci à tous ceux qui, de près ou de loin, ont cru en moi et m'ont
encouragé dans mes moments d'échec.*

Avec tous mes vœux de bonheur et santé.





A mes amis de la Faculté de Médecine et de Pharmacie de Rabat

*Amina, Mehdi, Ali, Wahiba, Brahim et tous les autres... Merci
pour les bons moments partagés.*

*C'était un plaisir de vous rencontrer et de partager des moments
de joie et de travail.*

Je vous souhaite une vie pleine de bonheur et de succès.





A mes oncles Jamaa, Mohamed et Larbi

Que ce travail soit le témoignage de mon respect avec mes souhaits de bonheur pour vous tous. Que dieu vous procure santé, bonheur et succès.

A Dr. Omar Ousaid,

Pour m'avoir si bien accueilli et m'encadré dans votre officine, je vous remercie.





***Aux collègues de Pfizer El Jadida, et d'ARWA MEDIC et
toutes les personnes qui ont participé de près et de loin à
l'élaboration de ce travail et à tous ceux que j'ai omis de citer.***



REMERCIEMENTS





A notre maître et président de thèse
Monsieur le professeur M'hamed ANSAR
Professeur de chimie organique

*Vous nous avez accordé un immense honneur et un grand
privilège en acceptant la présidence de notre jury de thèse.*

*Veillez accepter, cher professeur, l'assurance de notre estime et
notre profond respect.*





A notre maître et rapporteur de thèse
Monsieur le professeur Azeddine IBRAHIMI
Professeur de biotechnologie

Nous vous remercions pour la gentillesse et la bienveillance avec lesquelles vous avez bien voulu diriger ce travail. Votre compétence, votre sérieux, et votre rigueur sont pour nous le meilleur exemple à suivre.

Vous nous avez toujours impressionnés par vos qualités humaines et professionnelles. Merci pour le temps que vous m'avez accordé malgré votre planning chargé.

Veuillez, cher professeur, trouver dans ce modeste travail l'expression de notre haute considération et notre profond respect.





A notre maître et juge de thèse
Monsieur le professeur Jaouad EL HARTI
Professeur de chimie thérapeutique

Merci pour vos incroyables conférences et discussions et merci pour votre soutien. J'ai beaucoup appris et bénéficié de votre cours et vos conseils.

Vous nous faites un grand honneur en acceptant de juger notre travail.

Pour l'honneur que vous me faites de siéger parmi les membres du jury, veuillez trouver ici l'expression de ma sincère reconnaissance.





A notre maître et juge de thèse
Monsieur le professeur Youssef RAMLI
Professeur de chimie

*Durant notre formation de travaux pratiques, nous avons eu le
privilège de bénéficier de votre enseignement et d'apprécier votre
rigueur et votre sens professionnel.*

*Vous nous faites un grand honneur en acceptant de juger ce
travail.*

*Veuillez croire, cher professeur, en l'expression de notre
gratitude et notre grande estime.*



LISTE DES ABREVIATIONS

Abréviations

ADN	: Acide désoxyribonucléique
AQUA	: Absolute quantification
ARNm	: Acide ribonucléique messenger
ATP	: Adénosine triphosphate
BCCP	: Biotin carboxyl carrier protein
CBP	: Calmodulin-binding peptide
CDK1	: Cyclin-dependent kinase 1
CKAP2	: Cytoskeleton associated protein 2
FDA	: Food and Drug Administration
FT-ICR-MS	: Fourier-transform ion cyclotron resonance mass spectrometry
GLUT4	: Glucose Transporter 4
GTP	: Guanosine triphosphate
IRC	: Insuffisance rénale chronique
IRS	: Insulin receptor substrate
iTRAQ	: Isobaric tag for relative and absolute quantification
LTQ	: Linear trap quadrupole
MPT	: Modifications post-traductionnelles
MRM	: Multiple reaction monitoring
PSAQ	: Protein standard absolute quantification
QconCAT	: Quantitative concatenation
REP	: Rab escort protein
RPLC	: Reversed-phase liquid chromatography
SRM	: Selected reaction monitoring
SWATH	: Sequential Window Acquisition of all THEoretical Mass Spectra
TEV	: Tobacco etch virus
TMT	: Tandem mass tagging
TPST	: Tyrosylprotein sulfotransferase
UBD	: Ubiquitin-binding domain
VHC	: Virus de l'hépatite C
XIC	: Extracted-ion chromatogram

LISTE DES ILLUSTRATIONS

Liste des figures

Figure 1: Situation hiérarchique de la protéomique au sein des sciences « omiques »	3
Figure 2: Résultats de la recherche documentaire sur les tendances de la publication sur la protéomique quantitative par spectrométrie de masse au cours des 5 dernières années dans trois bases de données.	7
Figure 3: Schéma de la structure secondaire d'une protéine (hélice α et feuillet β)	10
Figure 4: Phases structurales aboutissant à la formation d'une structure quaternaire	11
Figure 5: Le cycle de vie d'une protéine	13
Figure 6: Éléments de la transcription de l'ARNm	14
Figure 7: Les codons et leur acides aminés correspondants. (à noter que les codons UAA, UAG et UGA ne codent pas d'acide aminé)	16
Figure 8: Les différents types de modifications des protéines responsables de la génération de protéoformes.....	19
Figure 9: Les principaux types de réactions de modification post-traductionnelle	20
Figure 10: Schéma représentant la complémentarité des réactions de phosphorylation et de déphosphorylation.....	21
Figure 11: Schéma de l'acétylation de la protéine N-terminale et de la lysine	24
Figure 12: Mécanismes de régulation des facteurs de transcription acétylation-dépendante.....	25
Figure 13: Schéma du métabolisme des sulfates	26
Figure 14: Les trois types de réactions de méthylation protéique sur les résidus de l'arginine.....	29
Figure 15: Les trois voies de modification du nucléosome	30
Figure 16: Cibles directes de la transcription du HIF	31
Figure 17: Contrôle de la transcription HIF-1 dépendante de l'oxygène	32
Figure 18: Principales classes de glycanes et de glycoconjugués du glycocalyx du spermatozoïde humain.....	35

Figure 19: Schéma montrant la réaction catalysée par la biotine-protéine ligase (BPL) et le mécanisme général commun des carboxylases biotin-dépendantes	37
Figure 20: Schéma simplifié des 3 réactions de l'ubiquitination d'un substrat protéique	41
Figure 21: Les trois mécanismes de ligation de l'ubiquitine	42
Figure 22: Structure de l'ubiquitine mettant en évidence les huit sites de polyubiquitination	43
Figure 23: Les 3 étapes enzymatiques aboutissant aux différents types d'ubiquitination	43
Figure 24: Les chaînes de polyubiquitine homotypique (à gauche) et de polyubiquitine hétérotypique (les deux à droite)	44
Figure 25: Structure chimique de l'ADP-ribose, l'unité de base de l'ADP-ribosylation	46
Figure 26: Les 3 approches d'analyse de protéines par spectrométrie de masse en tandem	49
Figure 27: Résumé du processus et des approches utilisées dans l'analyse protéomique	52
Figure 28: Déroulement d'une analyse protéomique basée sur la spectrométrie de masse (MS et MS/MS)	62
Figure 29: Schéma d'une interface ESI-MS standard	64
Figure 30: Processus d'ionisation et de désorption dans le MALDI	65
Figure 31: Protocole de préparation des puces ProteinChip avec des échantillons biologiques et d'analyse des protéines retenues par SELDI-TOF-MS	67
Figure 32: Représentation schématique du fonctionnement d'un analyseur de type Orbitrap	71
Figure 33: Principe de fonctionnement d'un réflectron dans un spectromètre de masse à temps de vol.	72
Figure 34: Schéma du spectromètre de masse MALDI-TOF	73
Figure 35: Représentation schématique d'un analyseur quadripolaire (Flèche en rouge : trajectoire instable ; flèche en bleu : trajectoire stable)	74
Figure 36: Représentation schématique d'un channeltron (une dynode continue) : l'amplification du courant se fait par une cascade d'émission secondaire d'électrons	75
Figure 37: Aperçu des protéines de signalisation cellulaire ciblées par Tax1.....	77

Figure 38: Représentation schématique du système Y2H. (POI : protéine d'intérêt ; AD : domaine d'activation ; BD : domaine liant l'ADN)	80
Figure 39: Aperçu schématique de la procédure TAP. (Bait : protéine appât)	82
Figure 40: Comparaison entre la purification par affinité en tandem ou TAP (gauche) et en une seule étape (droite)	83
Figure 41: Modèle d'application de la technique BioID	84
Figure 42: Représentation schématique de la technique Split BioID	86
Figure 43: Vue d'ensemble des techniques d'identification des interactions protéine-protéine.	87
Figure 44: Résumé des stratégies de quantification basées sur la spectrométrie de masse	88
Figure 45: Principes de la quantification par l'approche sans marquage	91
Figure 46: Comparaison de la quantification par SILAC in vitro (A) et in vivo (B)	92
Figure 47: Structures chimiques des réactifs iTRAQ (a) 4-plex et (b) 8-plex.....	94
Figure 48: Structure chimique générale d'un réactif ICAT. (X = hydrogène ou deutérium).....	95
Figure 49: Structure générale d'une protéine chimérique QconCAT	97
Figure 50: Schéma des différentes étapes du développement des biomarqueurs	103
Figure 51: Spectres de masse exemplaires de cellules intactes d'une variété d'espèces bactériennes cliniques courantes.....	111
Figure 52: Aperçu des approches protéomiques dans l'identification et la caractérisation des bactéries	118
Figure 53: Vue d'ensemble du processus d'analyse protéomique du plasma des patients atteint de COVID-19	122

Liste des tableaux

Tableau 1: Exemples de protéines et de leur fonction	8
Tableau 2: Exemples des PTMs effectuées sur certains acides aminés	18
Tableau 3: Les classes des modifications post-traductionnelles	20
Tableau 4: Caractéristiques des protéines prényltransférases humaines	40
Tableau 5: Classification des détergents	55
Tableau 6: Résumé des différentes caractéristiques des méthodes d'ionisation actuelles utilisées en spectrométrie de masse	69
Tableau 7: Comparaison des caractéristiques et des spécifications des analyseurs de masse couramment utilisés en protéomique	74
Tableau 8: Caractéristiques et applications des méthodes quantitatives basées sur MS	95
Tableau 9: Résumé des approches pour la quantification absolue de l'abondance des protéines par spectrométrie de masse	98
Tableau 10: Comparaison des techniques de découverte des biomarqueurs usuelles	104

SOMMAIRE



PREMIERE PARTIE (I) : INTRODUCTION ET GENERALITES.....	1
1. Introduction	2
2. Définition de la protéine et de la protéomique	4
3. Histoire de la protéomique	4
4. Pourquoi la protéomique ?	5
5. Rôles de protéines dans l'organisme	7
6. Structure des protéines	9
6.1. Les acides aminés	9
6.2. Structure primaire	9
6.3. Structure secondaire	10
6.4. Structure tertiaire	10
6.5. Structure quaternaire	11
7. Classification des protéines	12
8. Biosynthèse des protéines	12
8.1. La transcription	14
8.1.1 Les modifications post-transcriptionnelles	15
8.2. La traduction	15
8.2.1. Le ribosome	16
8.2.2. Les étapes de la traduction	17
9. Les modifications post-traductionnelles	18
9.1. Les grands types de modifications post-traductionnelles	20
9.2. Les réactions de phosphorylation	21
9.3. Les réactions de l'acétylation	23
9.4. Les réactions de sulfatation	25
9.5. Les réactions de méthylation	27

9.6. Les réactions d'hydroxylation	30
9.7. Les réactions de glycosylation	33
9.8. Les réactions de biotinylation	36
9.9. Les réactions de prénylation	38
9.10. Les réactions d'ubiquitination	40
9.11. Les réactions de l'ADP-ribosylation	46
DEUXIEME PARTIE (II) : METHODES ET OUTILS D'ETUDE PROTEOMIQUE.....	47
1. Introduction	48
2. Les bases de l'analyse protéomique	49
2.1. La protéomique descendante	50
2.2. La protéomique ascendante	50
2.3. La protéomique « middle-down »	51
2.4. Lyse cellulaire et extraction de protéines	53
2.4.1 Les techniques mécaniques de lyse cellulaire	54
2.4.2 Les techniques chimiques de lyse cellulaire	54
2.5. Décomplexification et enrichissement des protéines	56
2.6. Purification des protéines	56
2.7. Séparation de protéines ou peptides	56
3. L'électrophorèse bidimensionnelle (2DE)	57
4. La chromatographie liquide à haute performance (HPLC)	59
5. Analyse des protéines par spectrométrie de masse	60
5.1. Les sources d'ionisation	63
5.1.1. Ionisation par électronébulisation	63
5.1.2. Désorption-ionisation laser assistée par matrice (MALDI)	64
5.1.3. SELDI	66
5.1.4. Autres techniques d'ionisation	68

5.2. Les analyseurs	69
5.2.1. L'analyseur à trappe orbitale	69
5.2.2. L'analyseur à temps de vol	71
5.2.3. L'analyseur quadripolaire	73
5.3. Les détecteurs	75
6. Techniques d'analyse d'interactions protéine-protéine (IPP)	76
6.1. Définition des interactions protéine-protéine	78
6.2. Méthodes conventionnelles d'analyse d'IPP	79
6.2.1. Technique du double hybride de la levure « Y2H »	79
6.2.1.1. Principe de la technique	79
6.2.1.2. Limites de la technique	80
6.2.2. Technique AP-MS	80
6.2.2.1. Purification par affinité en une seule étape	81
6.2.2.2. Purification par affinité en tandem (TAP)	81
6.3. Méthodes récentes d'analyse d'IPP	84
6.3.1. Technique Split BioID	85
6.3.2. Technique BioID APEX et APEX2	86
7. La protéomique quantitative	87
7.1. La quantification relative des protéines	89
7.1.1. L'approche sans marquage	89
7.1.2. L'approche avec marquage	91
7.2. La quantification absolue des protéines	96
8. La bioinformatique	99
TROISIEME PARTIE (III) : APPLICATIONS EN PROTEOMIQUE.....	100
1. La protéomique fonctionnelle	101
2. La protéomique différentielle	101

3. Applications dans la recherche médicale et biomédicale	102
3.1. Relation protéome-pathologie	102
3.2. Découverte de nouveaux biomarqueurs	102
3.3. Surveillance de la progression du cancer	106
3.4. Diagnostic biologique de pathologies	107
3.5. Suivi thérapeutique	107
3.6. Elucidation de mécanismes physiopathologiques	108
4. Applications en microbiologie	109
4.1. Applications en bactériologie	109
4.2. Applications en virologie	112
5. Applications dans la recherche pharmaceutique	113
5.1. Détermination de mécanismes d'action des médicaments	113
5.2. Découverte de nouvelles cibles thérapeutiques	114
6. Applications en toxicologie	115
7. Applications extra-médicales	117
7.1. Taxonomie des organismes microscopiques	117
7.2. Applications en agroalimentaire	119
8. Applications dans le contexte du COVID-19	120
8.1. Découverte de cibles thérapeutiques	120
8.2. Recherche de biomarqueurs	121
8.3. Identification du SARS-CoV-2 par MS	122
CONCLUSION	124
RESUMES	126
REFERENCES	130

Première partie (I) : INTRODUCTION ET GENERALITES

1. Introduction :

Le corps humain, comme celui des autres êtres vivants, est composé de substances dites organiques qui sont indispensables à son bon fonctionnement. Parmi ces substances, on trouve les protéines, qui sont des macromolécules qui assurent plusieurs fonctions métaboliques et physiologiques cruciales pour la maintenance et la survie des êtres vivants. Afin que les cellules humaines puissent respirer, le corps a besoin d'une fonction assurée par l'hémoglobine, une protéine sanguine indispensable au transport de l'oxygène vers les tissus et cellules utilisateurs. Cette protéine n'est qu'un exemple parmi plus de 10 000 autres protéines humaines [1].

Vu les diverses fonctions vitales et implications de protéines dans plusieurs pathologies, leur étude structurale et fonctionnelle promet une meilleure connaissance de l'ensemble de ces processus.

Comme la génomique (l'étude des gènes) s'est avérée inadéquate pour prédire la structure et les propriétés dynamiques de toutes les protéines, un nouveau champ — la protéomique — s'est développé. La protéomique a connu ses débuts en 1994 avec la première définition du protéome comme étant l'ensemble des protéines codées par un génome, ce qui peut se concevoir à l'échelle d'une cellule, d'un tissu ou d'un organisme entier [2]. La figure suivante (Figure 1) dépeint la situation hiérarchique de la protéomique au sein des sciences omiques.

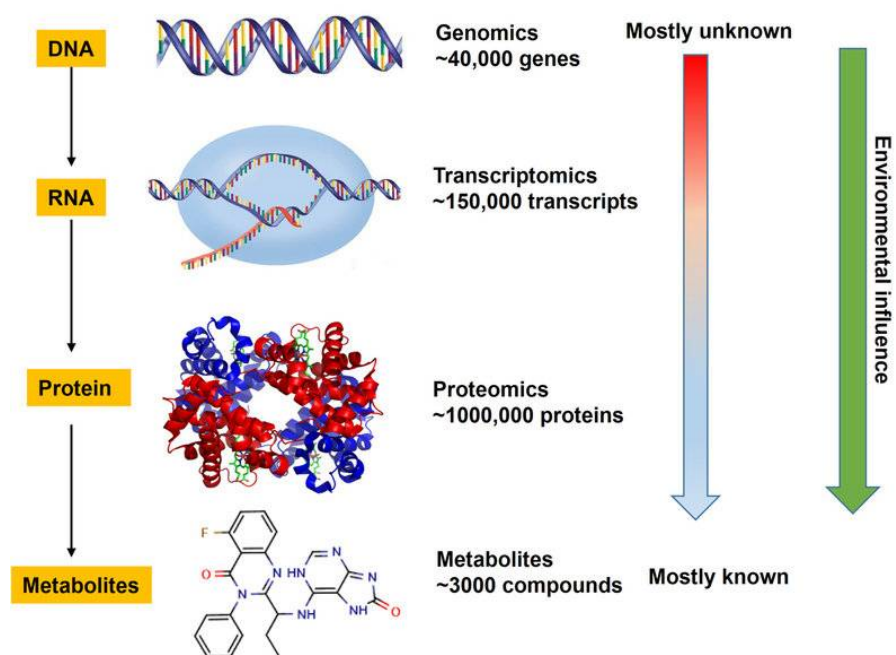


Figure 1: Situation hiérarchique de la protéomique au sein des sciences « omiques » [3]

Néanmoins, la protéomique et les technologies génomiques doivent être considérées comme complémentaires et il est bien évident que certaines questions sont mieux adressées par l'une ou l'autre approche. L'étude des protéines est clairement très importante pour obtenir de nouvelles connaissances sur la fonction cellulaire, car elles sont presque toujours les effecteurs des fonctions biologiques.

De plus, la majeure avancée scientifique et technologique dans le domaine de la bio-informatique et l'émergence de « Data Science » permet d'utiliser et analyser les données protéomiques d'une façon plus performante et pertinente [4], sans oublier la révolution de l'intelligence artificielle qui a permis de prédire la forme spatiale de protéines grâce au développement de systèmes à savoir le système AlphaFold développé par Google [5].

2. Définition de la protéine et de la protéomique :

La protéine est définie comme une biomolécule composée d'un enchainement de plusieurs acides aminés. Par convention, une chaîne comportant moins de 50 acides aminés est appelée « peptide », alors que celle ayant plus de 50 acides aminés serait désignée par le nom « protéine ».

Le terme « protéome », conçu par Marc R. Wilkins en 1994, désigne l'ensemble des peptides et/ou protéines présentes dans un organisme, dans un tissu, ou dans un liquide biologique. Ce mot hybride associe « protéine » et le suffixe « -ome » désignant un ensemble ou un système. La protéomique est donc la science ou la discipline qui étudie un ensemble de peptides et/ou de protéines d'un organisme.

En recherche clinique, la « protéomique clinique » consiste à analyser le protéome d'un échantillon biologique dans le but de découvrir et d'identifier des marqueurs biologiques ayant une valeur diagnostique ou qui peuvent être employés dans le suivi thérapeutique des diverses pathologies, ou des facteurs pathologiques qui peuvent être exploités comme cible thérapeutique [6].

3. Histoire de la protéomique :

Pour que la protéomique ait vu sa naissance, certains événements ont dû prendre place à priori. Ces progrès en protéomique ont eu lieu au cours d'une décennie qui a vu d'énormes progrès dans tous les domaines de la génomique fonctionnelle (comprendre la fonction des gènes et de leurs protéines correspondantes globalement). Le séquençage d'Edman de l'extrémité amino-terminale de protéines intactes ou de fragments digérés par voie enzymatique était l'approche la plus courante pour identifier les protéines évaluées sur gel.

Pendant des années, la spectrométrie de masse a été l'outil parfait du chimiste pour l'analyse des petites molécules. La grande précision des mesures de spectrométrie de masse permet de distinguer des espèces étroitement apparentées, et la spectrométrie de masse en tandem ou « MS/MS » peut fournir des informations structurales sur les ions moléculaires qui peuvent être isolés et fragmentés dans l'instrument. Pour mesurer la masse ou, plus précisément, le rapport masse/charge (m/z) d'une molécule dans un spectromètre de masse,

l'analyte doit d'abord être ionisé et transféré dans le système à vide poussé de l'instrument. Les peptides et les protéines, comme d'autres grosses molécules, se sont avérés difficiles à ioniser dans des conditions qui ne détruisent pas la molécule. À la fin des années 1980, deux méthodes ont été mises au point pour permettre l'ionisation des peptides et des protéines avec une grande sensibilité et sans fragmentation excessive. Ces percées sont l'ionisation par électronébulisation (ESI) et l'ionisation par désorption laser assistée par matrice (MALDI), qui ont suivi de près le développement de la désorption laser.

Le succès de ces méthodes d'ionisation des protéines a conduit au développement de spectromètres de masse commerciaux équipés d'instruments robustes de type ESI ou MALDI, qui ont rapidement pénétré la communauté de la chimie des protéines. Les sources d'ions MALDI étaient le plus souvent couplées à des analyseurs de masse à temps de vol (TOF), tandis que l'ESI était le plus souvent couplée à des spectromètres MS/MS à piège à ions ou à triple-quadripôle. Bien que les spectromètres de masse MALDI-TOF puissent déterminer la masse d'une protéine ou d'un peptide avec un haut degré de précision, la masse intrinsèque d'une protéine eucaryote n'est pas une caractéristique d'identification unique. Il a cependant été rapidement reconnu que les masses des différents peptides générés par la fragmentation d'une protéine isolée avec une enzyme dont la spécificité de clivage est connue pouvaient identifier une protéine de manière unique.

4. Pourquoi la protéomique ?

Grâce aux efforts du projet « Human Genome Project », on sait maintenant que le génome humain comporte environ 23 000 gènes, et donc il a théoriquement une capacité d'exprimer environ 23 000 protéines différentes (hormis les modifications transcriptionnelles) si on considère que l'ensemble des gènes sont protéogènes, alors que le protéome humain comporte plus d'un million de protéoformes, un nombre très loin de celui de gènes existant chez l'homme [7]. Une protéine donnée porte un ou plusieurs résidus aminoacides qui peuvent être sujets à une ou plusieurs modifications co- ou post-traductionnelles. Cependant, les différentes combinaisons de ces modifications sur une même protéine sont une source cruciale de la diversification des protéoformes.

La structure et la fonction d'une protéine seront donc définies par sa séquence d'acides aminés qui peuvent être a posteriori modifiés ou non, et non par la séquence du gène responsable de sa biosynthèse elle-même.

Les protéomes sont intrinsèquement plus complexes et dynamiques que les génomes, car chaque gène peut potentiellement donner naissance à plus d'une protéine. Pour ajouter à cette complexité, l'abondance de chaque protéine est régulée, et les protéines sont généralement traitées et modifiées chimiquement, et s'engagent souvent dans des interactions complexes avec d'autres protéines ou de petites molécules (par exemple des métabolites, des lipides, des ions métalliques et des cofacteurs).

La protéomique est une technique puissante qui permet l'identification et la quantification du protéome, à grande échelle, grâce à laquelle les chercheurs peuvent acquérir des connaissances importantes sur la composition moléculaire qualitative et quantitative (Figure 2) et la physiologie d'un échantillon biologique.

Il est également important de noter que les protéines sont des cibles privilégiées pour les agents thérapeutiques et les tests de diagnostic, la protéomique fournissant des informations sur la composition, les fonctions, les interactions médicamenteuses et la régulation des protéines qui ne sont pas prévisibles sur la base des séquences du génome.

La vaste majorité des médicaments agissent sur des enzymes, des protéines, des canaux transporteurs ou des récepteurs afin d'exercer leur effet pharmacologique. Or, les études en protéomique pourraient identifier de nouvelles cibles thérapeutiques potentielles et donc prodiguer plus d'espoir à la prise en charge pharmacologique de nombreuses maladies, comme c'est le cas pour certains cancers et maladies auto-immunes traités par des anticorps monoclonaux à l'instar de Adcetris® (brentuximab védotine) et Humira® (adalimumab).

Les interactions protéine-protéine (l'interactomique) sont aussi à l'origine de la diversité des conséquences fonctionnelles des protéines. Ces interactions sont exploitées par quelques médicaments comme le maraviroc [8].

En outre, l'analyse des protéines et de leurs modifications post-traductionnelles (PTM) est particulièrement cruciale pour l'étude de diverses affections cardiovasculaires, du diabète, du cancer, et des maladies neurodégénératives comme la maladie d'Alzheimer.

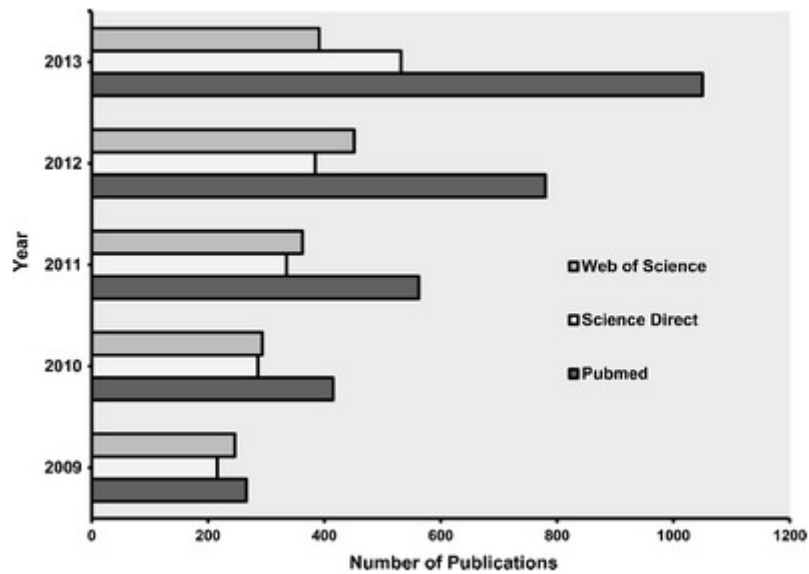


Figure 2: Résultats de la recherche documentaire sur les tendances de la publication sur la protéomique quantitative par spectrométrie de masse au cours des 5 dernières années dans trois bases de données. [9]

5. Rôles de protéines dans l'organisme :

A côté d'autres biomolécules importantes pour l'organisme humain à l'instar des glucides et des lipides, les protéines peuvent être assimilées à la fois à des maçons et des protecteurs des organismes vivants. Elles constituent la majeure partie de la matière sèche des cellules.

Le tableau 1 présente les types de fonctions qu'assurent les protéines.

FONCTION	DESCRIPTION	EXEMPLES
Anticorps	Les anticorps se lient à des particules exogènes spécifiques, comme les virus, les bactéries, pour protéger le corps.	L'immunoglobuline G (IgG) L'immunoglobuline M (IgM)
Protéines messagères	Les protéines messagères comme certaines hormones protéiques, transmettent des signaux pour réguler et coordonner les processus biologiques entre les cellules.	Hormone de croissance (hGH)
Enzyme	Les enzymes assurent l'accélération de réactions biochimiques dans les cellules.	HMG-CoA réductase Ornithine décarboxylase
Transport/stockage	Ces protéines se lient et transportent des atomes et des petites molécules dans les cellules et dans tout l'organisme.	Lipoprotéines, ferritine, hémoglobine
Composition structurale	Ce type de protéines sert comme support cellulaire et fait partie de l'intégrité de composants structuraux de l'organisme.	Actine, élastine

Tableau 1: Exemples de protéines et de leur fonction

6. Structure des protéines :

6.1. Les acides aminés :

Les protéines sont des molécules structurellement complexes et fonctionnellement sophistiquées. Elles sont assemblées principalement à partir d'un ensemble minimal de 20 acides aminés différents, chacun avec des propriétés chimiques différentes. Les acides aminés qui composent une protéine sont maintenus ensemble par des liaisons covalentes appelées « liaisons peptidiques » (Figure 3). Dans chaque type de protéine, les acides aminés sont présents dans un ordre unique, qui est exactement le même d'une molécule de ce type de protéine à l'autre. Une molécule d'insuline humaine, par exemple, a la même séquence d'acides aminés que toutes les autres molécules d'insuline humaine. Plusieurs milliers de protéines différentes ont été identifiées, chacune avec sa propre et unique séquence d'acides aminés.

Les acides aminés sont des molécules amphotères composées d'une fonction amine, une fonction carboxylique et une chaîne latérale. La composition chimique de cette dernière détermine les propriétés physico-chimiques des acides aminés. Il existe environ 20 acides aminés protéinogènes chez les êtres vivants. Les interactions qui se produisent entre les chaînes latérales des acides aminés dans une protéine déterminent la forme tridimensionnelle de cette protéine.

La structure d'une protéine ne s'arrête pas avec des hélices alpha et des feuillets beta ; il y a d'autres niveaux d'organisation. Ces niveaux ne sont pas indépendants, mais sont construits l'un sur l'autre pour établir la structure tridimensionnelle complète de la protéine.

6.2. Structure primaire :

La simple séquence d'acides aminés composant un polypeptide ou une protéine est qualifiée de structure primaire. C'est à partir de cette structure primaire, grâce aux interactions physico-chimiques entre les groupements des acides aminés, que se développent les autres structures tridimensionnelles progressivement.

6.3. Structure secondaire :

La structure secondaire des protéines se développe à partir de la structure primaire, grâce à la formation d'hélices α (alpha) et les feuillets β (bêta) – selon les règles de Corey-Pauling –, assurée et maintenue principalement par des liaisons hydrogène établies entre l'atome d'oxygène du carbonyle d'un acide aminé et l'hydrogène de la fonction amine d'un autre (Figure 3).

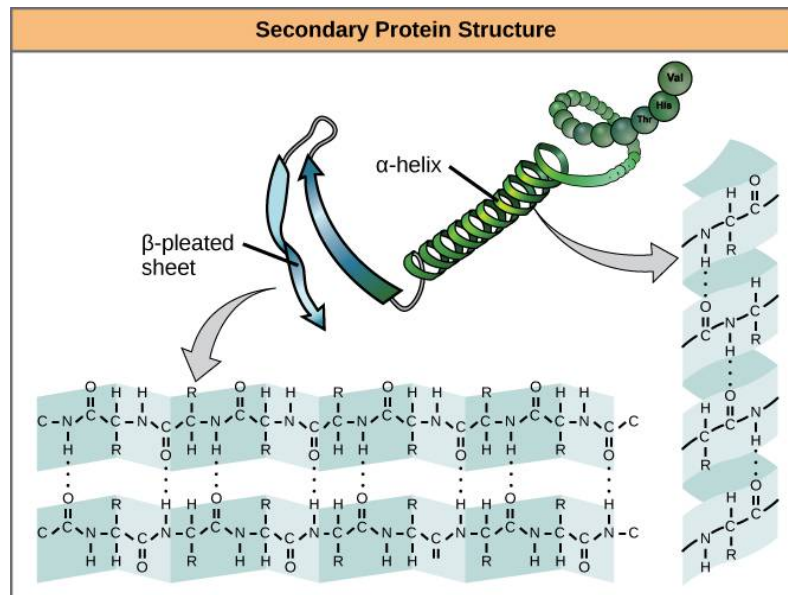


Figure 3: Schéma de la structure secondaire d'une protéine (hélice α et feuillet β)

6.4. Structure tertiaire :

La conformation tridimensionnelle formée entièrement par une chaîne de polypeptides, y compris les hélices α (alpha), les feuillets β (bêta), les coudes aléatoires et les autres boucles et plis qui se forment entre le N et le C-terminaux, est parfois appelée la structure tertiaire des protéines.

6.5. Structure quaternaire :

On parle de la structure quaternaire d'une protéine lorsqu'il y a assemblage et interaction de deux ou plusieurs sous-protéines de structure tertiaire pour former une nouvelle protéine tridimensionnelle et compacte. L'hémoglobine par exemple est une protéine comportant quatre sous-unités protéiques (deux chaînes alpha et deux chaînes bêta) associées entre elles (Figure 4).

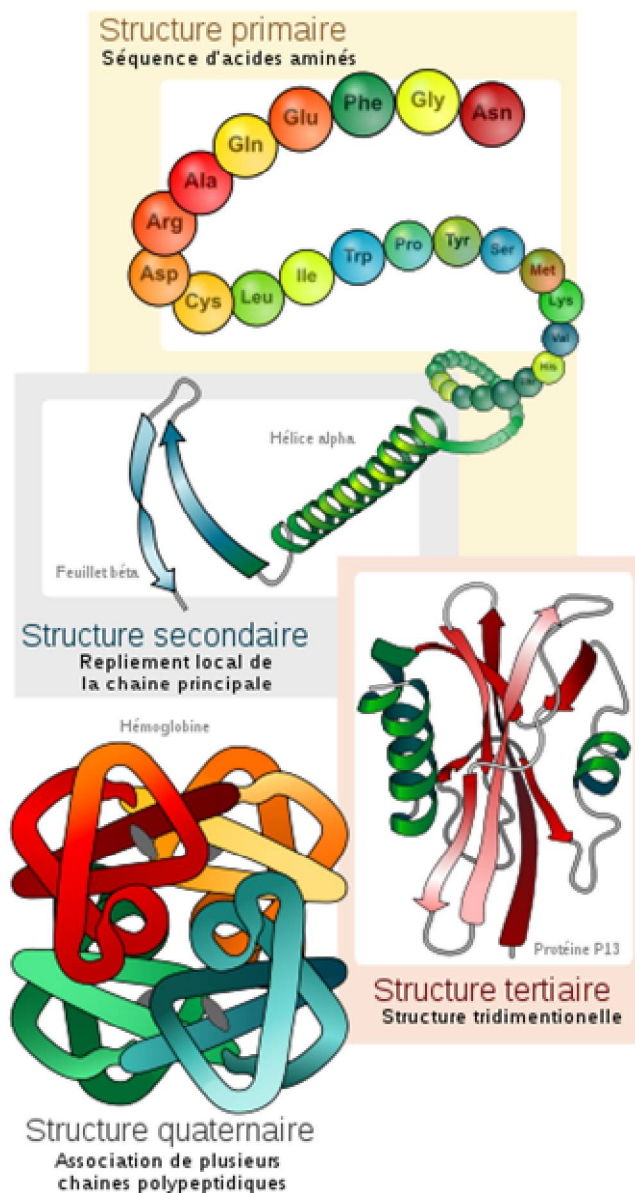


Figure 4: Phases structurales aboutissant à la formation d'une structure quaternaire

7. Classification des protéines :

Les protéines peuvent être classées en deux grandes catégories, selon leur état conjugué ou non : les protéines simples et les protéines conjuguées. Les protéines simples sont des protéines composées uniquement d'un enchainement d'acides aminés. En revanche, les protéines conjuguées sont des protéines simples additionnées d'une partie non protéique.

De plus, les protéines peuvent être classées selon leur forme tridimensionnelle en : protéines fibreuses (ou scléroprotéines) et protéines globulaires.

8. Biosynthèse des protéines :

La vie d'une protéine commence grâce à l'ADN, et ce par des processus intra-cellulaires effectués à chaque étape à l'aide des enzymes et protéines spécialisées. La séquence des acides aminés d'une protéine particulière est spécifiée par la séquence des nucléotides dans un segment particulier de l'acide désoxyribonucléique (ADN). Le processus de synthèse des protéines se déroule en deux étapes. Premièrement, l'ADN est transcrit en un acide ribonucléique (ARN) intermédiaire, l'ARN messenger (ARNm), qui possède une séquence ribonucléotidique complémentaire de celle de la séquence désoxyribonucléotidique d'un des brins de l'ADN servant de matrice (transcription).

L'ARNm s'attache aux particules ribonucléoprotéiques cytoplasmiques (ribosomes) qui sont les sites de synthèse des protéines, et y détermine l'ordre de liaison des acides aminés en une protéine spécifique (traduction). L'ARNm est traduit dans la direction 5' à 3'. La synthèse d'une protéine est initiée au niveau de l'acide aminé amino-terminal et se poursuit vers l'acide aminé carboxy-terminal. Pendant la traduction, un groupe de trois nucléotides adjacents dans l'ARNm (codon) spécifie quel acide aminé doit être lié à la chaîne peptidique en croissance. Une fois que la protéine est biosynthétisée, elle peut subir davantage de modifications dites « post-traductionnelles » (Figure 5).

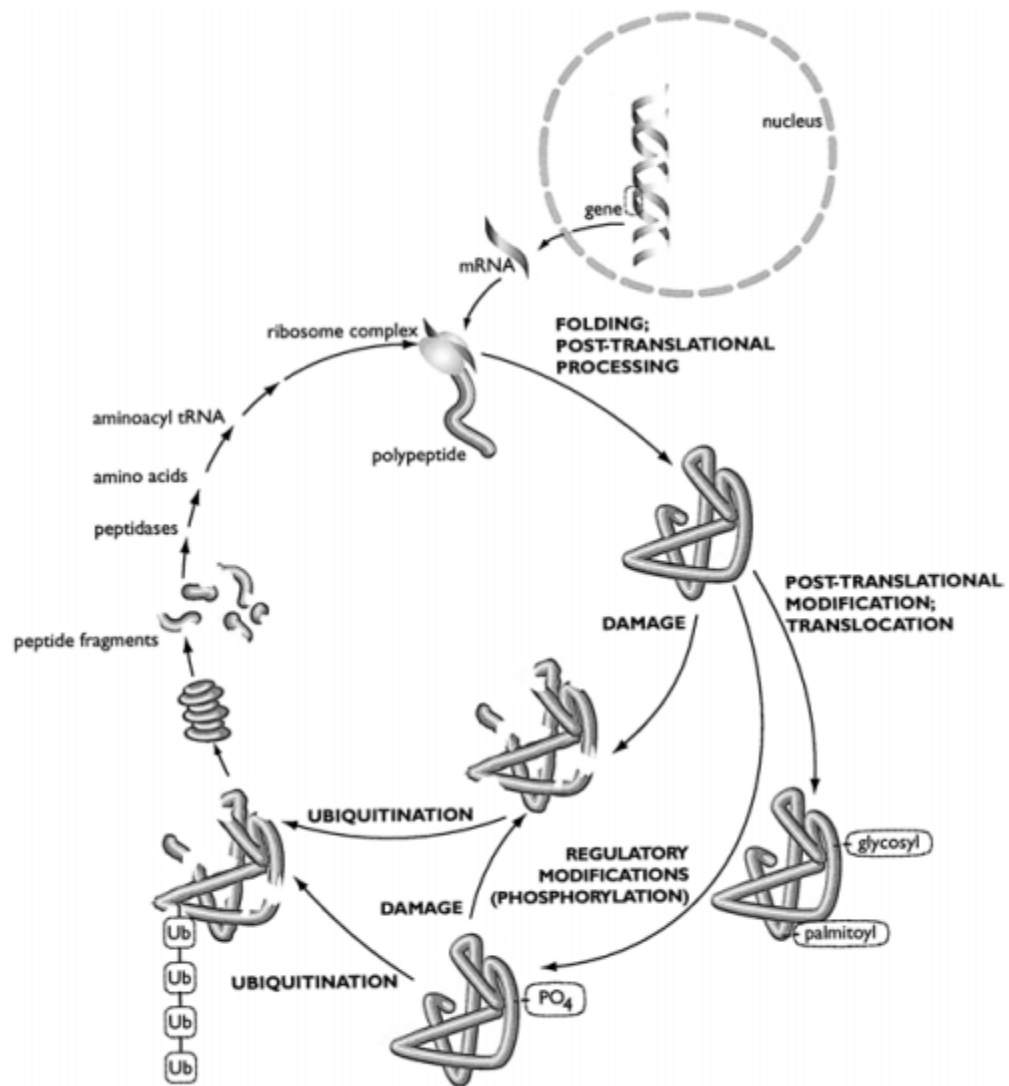


Figure 5: Le cycle de vie d'une protéine [10]

8.1. La transcription :

Chez les eucaryotes, la transcription est le premier pas de l'expression génique, assurée par une enzyme appelée « ARN polymérase », et qui se déroule au niveau du noyau cellulaire. Lors de ce processus, l'ADN est recopié en ARNm, une version plus spéciale et plus convenable pour la synthèse de protéines. A la différence de l'ADN, la structure de l'ARNm est composée de l'uracile à la place de la thymine (Figure 6).

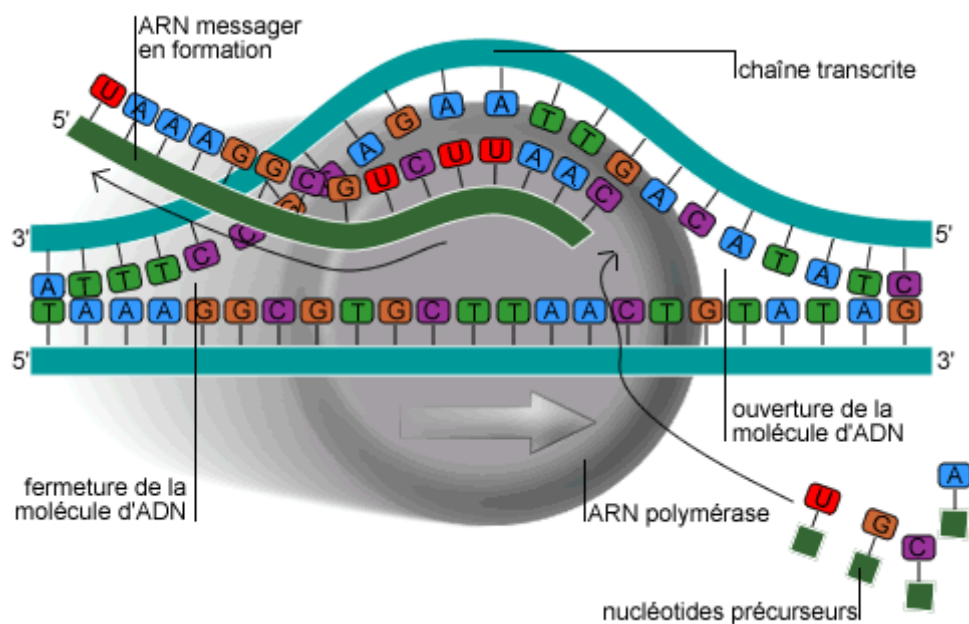


Figure 6: Éléments de la transcription de l'ARNm

La particularité de la transcription génique dans le cycle de vie des protéines est sa capacité de générer une ou plusieurs molécules d'ARNm à partir d'un seul gène, grâce à des modifications post-transcriptionnelles qui assurent la maturation de l'ARN pré-messager et la mise en œuvre de son épissage alternatif.

8.1.1 Les modifications post-transcriptionnelles :

L'épissage alternatif est un phénomène biologique qui existe presque chez tous les organismes eucaryotes et est catalysé par un large complexe ribonucléoprotéique connu sous le terme « splicéosome » [11]. C'est un processus par lequel différentes régions d'exons et d'introns d'un ARN pré-messager sont réunies pour produire des molécules d'ARN messenger matures, qui conduisent souvent à des protéines ou des isoformes uniques et fonctionnellement différents. Etant donné que plus de 90% des gènes codant pour des protéines subissent de l'épissage alternatif, ce dernier paraît incontestablement indispensable à une partie de la diversification du protéome humain [12].

8.2. La traduction :

Pour donner naissance aux protéines, le message contenu dans l'ARNm doit être traduit en langage protéique. Pour ce faire, les cellules utilisent une machinerie ribosomale qui assure l'appariement correcte des codons de l'ARNm et les acides aminés correspondants.

Un codon est un triplet de nucléotides de l'ARNm qui est reconnu et traduit en un acide aminé par les ribosomes. Chaque codon correspond à l'un des 20 acides aminés protéinogènes (Figure 7).

	U		C		A		G		
U	UUU	phénylalanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine	UCA		UAA	stop	UGA	stop	A
	UUG		UCG		UAG		UGG	tryptophane	G
C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	isoleucine	ACU	thréonine	AAU	asparagine	AGU	sérine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lysine	AGA	arginine	A
	AUG	méthionine	ACG		AAG		AGG		G
G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	acide glutamique	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 7: Les codons et leur acides aminés correspondants. (à noter que les codons UAA, UAG et UGA ne codent pas d'acide aminé)

Les codons UAA, UAG et UGA sont qualifiés de « codons stop », car ils marquent la fin du processus de traduction et la fin de la chaîne polypeptidique.

8.2.1. Le ribosome :

Le ribosome est un complexe macromoléculaire ribonucléoprotéique présent dans toutes les cellules vivantes, qui sert de site de biosynthèse des protéines (traduction). Les ribosomes relient les acides aminés dans l'ordre spécifié par les molécules d'ARN messenger (ARNm). Ils se composent de deux sous-unités principales : la petite sous-unité ribosomale, qui lit et traduit les codons de l'ARNm, et la grande sous-unité, qui a pour rôle d'enchaîner les acides aminés entre eux afin de former une chaîne polypeptidique.

La grande sous-unité 60S du ribosome eucaryotique (50S en bactéries) se compose de trois molécules de rARN (25S, 5.8S et 5S) et de 46 protéines, tandis que la petite sous-unité 40S (30S pour les bactéries) comprend une chaîne rRNA (18S) et 33 protéines [13].

8.2.2. Les étapes de la traduction :

La traduction est généralement divisée en trois étapes : l'initiation, l'élongation et la terminaison. Lors de l'initiation de la traduction, la petite sous-unité ribosomique (40S) se lie à l'initiateur spécifique, l'ARNt-méthionyle (Met-ARNt), et à un ARNm. L'étape d'initiation est terminée lorsque la petite sous-unité sélectionne un codon de départ et que la grande sous-unité ribosomique (60S) se joint pour former un ribosome fonctionnel. Un certain nombre de protéines non ribosomiques spécifiques sont également nécessaires pour les différentes étapes du processus de traduction.

Après la formation du complexe d'initiation, la traduction se poursuit par l'élongation de la chaîne polypeptidique. Le mécanisme d'élongation dans les cellules procaryotes et eucaryotes est très similaire. Le ribosome possède trois sites de liaison de l'ARNt, appelés sites P (peptidyle), A (aminoacyl) et E (exit ou sortie). L'ARNt méthionyle initiateur est lié au site P. La première étape de l'élongation est la liaison de l'ARNt aminoacyl suivant au site A par appariement avec le second codon de l'ARNm. L'ARNt aminoacyl est escorté vers le ribosome par un facteur d'élongation (EF-Tu chez les procaryotes, eEF-1 α chez les eucaryotes), qui est complexé au GTP. Après l'hydrolyse du GTP, l'EF-Tu (complexé avec le GDP) quitte le ribosome, avec l'ARNt aminoacyl inséré dans le site A. Une liaison peptidique est alors formée, entraînant le transfert de la méthionine vers l'ARNt de l'aminoacyl au site A. Le ribosome déplace alors trois nucléotides le long de l'ARNm. Ce mouvement transloque l'ARNt peptidyle (Met-AA) vers le site P et l'ARNt non chargé vers le site E, laissant un site A vide prêt pour l'ajout de l'acide aminé suivant. La translocation est médiée par l'EF-G, couplée à l'hydrolyse du GTP [14].

L'élongation de la chaîne polypeptidique se poursuit jusqu'à ce qu'un codon stop (UAA, UAG ou UGA) soit transloqué dans le site A du ribosome. Lorsqu'un codon stop pénètre dans le site A, les facteurs de terminaison se lient au ribosome et favorisent l'hydrolyse de l'ARNt peptidyle pour libérer le polypeptide terminé. Les ribosomes sont ensuite recyclés par des interactions avec un certain nombre de facteurs protéiques pour générer des sous-unités ribosomiques capables de subir un autre cycle de synthèse protéique.

9. Les modifications post-traductionnelles :

De nombreuses protéines subissent des modifications post-traductionnelles qui modifient considérablement leurs propriétés physico-chimiques, y compris leur pliage et conformation spatiale, leur stabilité et, par conséquent, leur activité et leur fonction biologique. Ces modifications post-traductionnelles sont un ensemble de transformations chimiques réversibles ou irréversibles, qui commencent dès la fin de la traduction de l'ARNm en polypeptide. Elles aident à faire diversifier le génome limité des organismes et sont majoritairement catalysées par des enzymes (transférases, ligases, hydroxylases etc...) qui ont la capacité de reconnaître certaines zones spécifiques et particulières de ces protéines. Le tableau 2 présente une liste des modifications post-traductionnelles et les résidus d'acides aminés sur lesquels elles se produisent. Néanmoins, il est important de mentionner qu'il n'y a actuellement pas de modifications connues pour les chaînes latérales de l'alanine, l'isoleucine, la leucine, la phénylalanine et de la valine.

Table 1 Posttranslational modifications by amino acid side chain modified

<i>Amino acid</i>	<i>Reaction</i>	<i>Amino acid</i>	<i>Reaction</i>
Arg	N-methylation	His	Phosphorylation
	N-ADP ribosylation		Aminocarboxypropylation
Asn	N-glycosylation		N-methylation
	N-ADP ribosylation	Lys	N-methylation
	Protein splicing		N-acylation by acetyl, biotinyl, lipoyl, ubiquitinyl groups
Asp	Phosphorylation		C-hydroxylation
	Isomerization to isoAsp	Met	Oxidation to sulfoxide
Cys	S-hydroxylation (S-OH)	Pro	C-hydroxylation
	Disulfide bond formation	Ser	Phosphorylation
	Phosphorylation		O-glycosylation
	S-acylation		Phosphopantetheinylation
	S-prenylation		Autocleavages
	Protein splicing	Thr	Phosphorylation
Gln	Transglutamination	Trp	C-mannosylation
Glu	Methylation	Tyr	Phosphorylation
	Carboxylation		Sulfation
	Polyglycination		Ortho-nitration
	Polyglutamylolation		TOPA quinone
Gly	C-hydroxylation		

No known modifications on Ala, Ile, Leu, Phe, and Val side chains.

Tableau 2: Exemples des PTMs effectuées sur certains acides aminés

Les modifications les plus courantes sont le clivage spécifique des protéines précurseurs ; la formation de ponts disulfure ; l'ajout voire l'élimination des groupes de faible poids moléculaire, conduisant ainsi à des modifications telles que l'acétylation, l'amidation, la biotinylation, la déamidation, la prénylation, la glycosylation, l'hydroxylation, la méthylation, la phosphorylation etc... ; ou la protéolyse. (Figure 9)

Avec la prise de conscience des PTMs, leur liste ne cesse pas de s'allonger et de plus en plus les recherches sont centrées sur leur fonction biologique et leur interaction dans le contexte de biologie des systèmes.

Tant qu'il existe plus de 100 types de modifications post-traductionnelles de protéines enregistrés dans la base de données UniProtKB [15], seuls ceux les plus fréquents et les plus importants sont discutés en général dans cette section. De plus, le terme « protéoforme » a été inventé pour couvrir les variantes moléculaires de protéines issues de modifications post-traductionnelles et d'autres transformations (Figure 8). Lorsque ce terme est employé en pluriel, il désigne toutes les molécules de protéines hautement apparentées provenant de toutes les sources de variation donnant naissance à des produits pouvant être remontés à un gène commun. Il s'agit notamment de produits qui diffèrent en raison de variations génétiques, d'ARN alternativement épissé et de modifications post-traductionnelles [16].

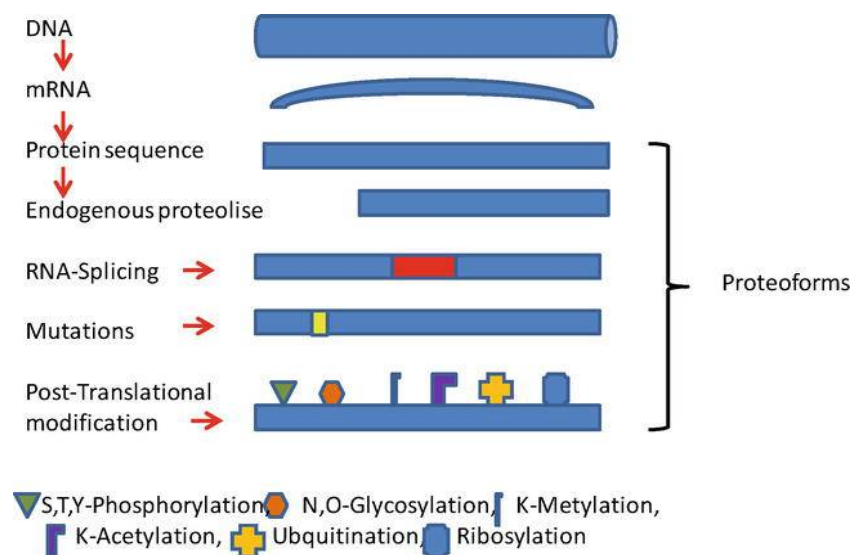


Figure 8: Les différents types de modifications des protéines responsables de la génération de protéoformes.

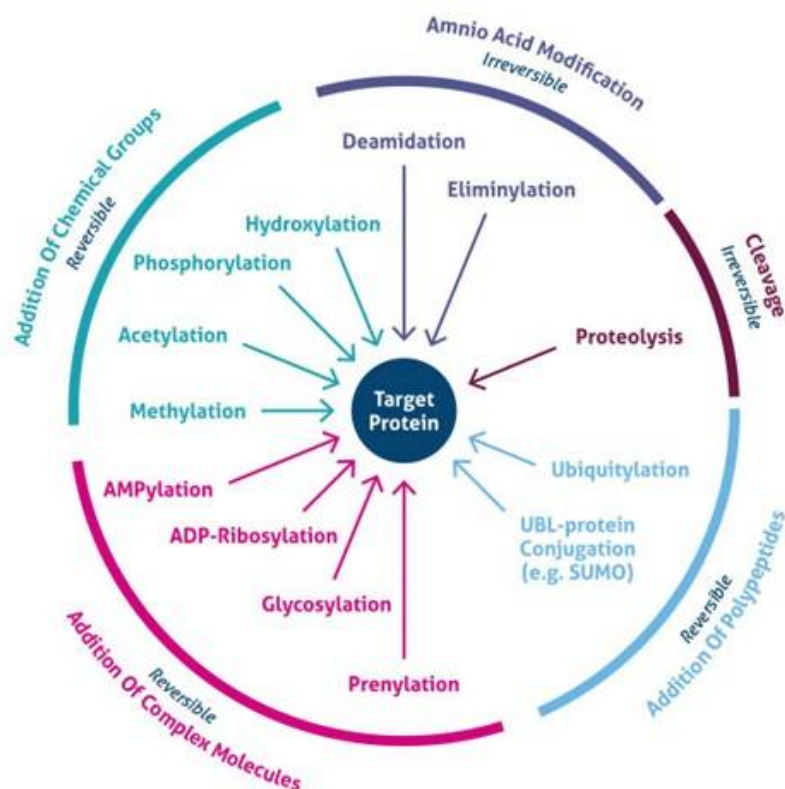


Figure 9: Les principaux types de réactions de modification post-traductionnelle

9.1. Les grands types de modifications post-traductionnelles :

Les modifications post-traductionnelles peuvent être réparties en 3 classes. Le tableau suivant représente ces types avec leurs exemples.

Classe des PTMs	Clivage de protéines	Modifications chimiques des protéines	La protéolyse (dégradation de protéines)
Exemple(s)	Proenzymes comme pepsinogène, activation d'hormones comme pour la préproinsuline.	Méthylation, acétylation, phosphorylation....	Il s'agit d'une transformation irréversible. L'ubiquitination facilite la dégradation...

Tableau 3: Les classes des modifications post-traductionnelles

9.2. Les réactions de phosphorylation :

La phosphorylation est une réaction essentielle dans la régulation de nombreux processus biochimiques. Elle s'agit de la modification post-traductionnelle des protéines la plus investiguée, et est physiologiquement la plus importante car elle est effectuée par une vaste famille de kinases regroupées sous le terme « kinome », qui inclus plus de 500 kinases. C'est une modification chimique assurée par des protéines kinases, et par laquelle une molécule acquiert un groupement phosphate inorganique. La déphosphorylation est le processus inverse de la phosphorylation, et est assurée par un groupe d'enzymes appelées phosphatases.

La phosphorylation consiste à transférer un groupement phosphate inorganique d'un cosubstrat riche en énergie (ATP ou GTP), vers une molécule contenant des acides aminés ayant un groupement hydroxyle libre, d'une façon covalente. Les acides aminés les plus fréquemment concernés par ce transfert sont la sérine, la threonine et la tyrosine, dont la sérine compte pour la majorité des réactions de phosphorylation.

La nature complémentaire de la phosphorylation et de la déphosphorylation (Figure 10) peut être assimilée à la dichotomie marche/arrêt d'un interrupteur électrique. Dans la majeure partie, les réactions de phosphorylation induisent une activation de protéines ou d'enzymes, alors que les réactions de déphosphorylation amorcent leur inactivation, et ceci permet la régulation de certains processus biologiques tels que la transduction ou la propagation de signal et le contrôle de l'apoptose, de la prolifération et de la croissance des cellules [17].

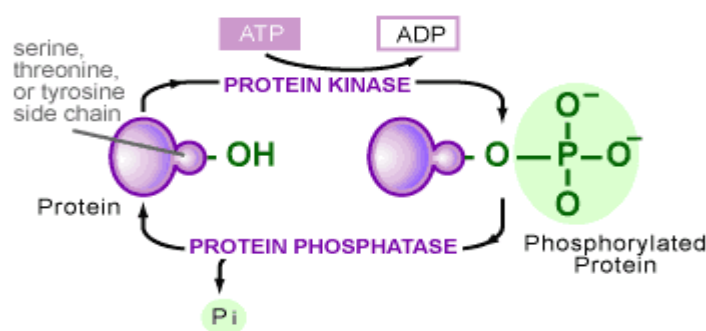


Figure 10: Schéma représentant la complémentarité des réactions de phosphorylation et de déphosphorylation

La régulation de la glycémie par l'insuline s'appuie sur la stimulation du récepteur membranaire de l'insuline qui est un récepteur à activité tyrosine kinase (RTK) ayant une structure hétérotétramérique $\alpha 2\beta 2$. Lorsque l'insuline s'attache au récepteur de l'insuline (RI), la partie cytoplasmique de la sous-unité β de ce dernier s'autophosphoryle grâce à l'activation de la tyrosine kinase intrinsèque du RI, ce qui déclenche deux principales voies de signalisation cellulaire, la voie PI3K/Akt ayant des effets métaboliques et la voie Ras/MAPK ayant un effet sur la croissance cellulaire. Dans la voie PI3K/Akt, les protéines IRS rejoignent le RI à l'aide de leur domaine de liaison de phosphotyrosine (PTB), et sont ensuite phosphorylées sur les résidus tyrosine par le RI. Lors de la phosphorylation, l'IRS-1 devient actif et active à son tour la phosphoinositide 3-kinase (PI3K) après son recrutement via le domaine « Src homology 2 » (SH2) de la sous-unité régulatrice p85 [18]. La sous-unité catalytique p110 de la PI3K assure la phosphorylation de PIP2 en PIP3 qui active ensuite la kinase 3-phosphoinositide-dépendante 1 et 2 (PDK1 et PDK2). Ces deux dernières enzymes activent à leur tour la protéine kinase B (PKB ou Akt).

A la fin de cette signalisation, il y a une translocation de récepteurs GLUT4 vers la membrane plasmique [19] et la phosphorylation de la glycogène synthase kinase 3 (GSK3) par la PKB entraîne sa désactivation, conduisant ainsi à l'activation de la glycogène synthase et à la biosynthèse hépatique du glycogène. La GSK3 était l'objet d'intérêt des recherches pharmacologiques sur le diabète de type 2 et le cancer depuis les deux dernières décennies [20].

Les protéines kinases ayant une relation avec certains néoplasmes constituent des molécules thérapeutiquement ciblées par une classe de médicaments antinéoplasiques appelés les inhibiteurs de la tyrosine kinase, à l'instar du sunitinib et de l'imatinib [21]. Dans la leucémie myéloïde chronique par exemple, le gène hybride BCR-ABL1 aboutit à l'expression de la protéine chimérique p210^{BCR-ABL1}, qui possède une activité tyrosine-kinase constitutive [22]. L'inhibition de l'activité de cette protéine assure un ralentissement de la leucémogénèse et constitue donc un mécanisme antinéoplasique efficace [23].

Le virus d'Ebola possède la protéine virale VP35 qui est un cofacteur de la polymérase virale. La phosphorylation de cette protéine dans les cellules hôtes est indispensable à la réplication de particules virales de ce virus [24].

En ce qui concerne la division cellulaire, la phosphorylation des résidus C-terminaux T561, T579, T603 et S608 de la protéine CKAP2 par le complexe CyclinB1–CDK1 joue un rôle crucial dans la stabilisation des microtubules et l'intégrité des centrosomes [25].

9.3. Les réactions de l'acétylation :

L'acétylation constitue l'une des modifications post-traductionnelles majeures chez les eucaryotes, à côté de la phosphorylation. Elle désigne un transfert d'un groupement acétyle ($\text{CH}_3\text{-CO-}$) d'une molécule donatrice, notamment l'acétyl-CoA, vers une autre molécule. Ce transfert peut avoir lieu catalytiquement par l'acétyltransférase ou par voie non-enzymatique. L'azote (N), le soufre (S), et l'oxygène (O) sont les atomes principaux pouvant recevoir une acétylation, mais c'est la N-acétylation qui constitue la majorité des réactions d'acétylation chez l'homme. Deux types principaux de la N-acétylation d'acides aminés d'une protéine peuvent avoir lieu : la $\text{N}\epsilon$ -lysine-acétylation (lysine-acétylation) et la $\text{N}\alpha$ -acétylation (Figure 11). Il est important de noter que l'acétylation des résidus lysine est réversible et est régulée grâce aux déacétylases.

● La $\text{N}\epsilon$ -lysine-acétylation :

Comme son nom l'indique, c'est une acétylation qui se produit sur la fonction amine de la chaîne latérale des résidus de la lysine. C'est une réaction généralement réversible, catalysée par la lysine acétyltransférase (KAT). Le processus inverse de cette réaction est effectué par la lysine déacétylase (KDAC). Néanmoins, il existe aussi des KAT qui agissent sur les résidus de lysine des histones appelées « histone acétyltransférases (HATs) », dont l'action est réversible grâce aux histones désacétylases (HDACs) [26].

La $\text{N}\epsilon$ -lysine-acétylation est l'une des principales modifications post-traductionnelles des histones. L'acétylation de multiples résidus de lysine dans les extrémités N-terminales des histones ou à l'extrémité C-terminale du facteur de transcription p53 est considérée cruciale dans la régulation épigénétique de la transcription de certains gènes et au remodelage de la chromatine [27] (Figure 12).

L'altération de la structure de la chromatine suite à l'acétylation de l'histone H4 résulte en une activation de la transcription des gènes *in vivo* et *in vitro* [28].

• La N α -acétylation :

Il s'agit d'une acétylation qui se produit sur la fonction α -amine terminale d'une chaîne polypeptidique. Cette réaction est catalysée par la N-acétyltransférase (NAT), et est considérée irréversible étant donné que les déacétylases correspondantes n'ont pas été découvertes jusqu'à présent [29].

Bien que ce type des MPT ait lieu dans la majeure partie au cours des modifications co-traductionnelles, son rôle dans le contexte des modifications post-traductionnelles n'est pas négligeable.

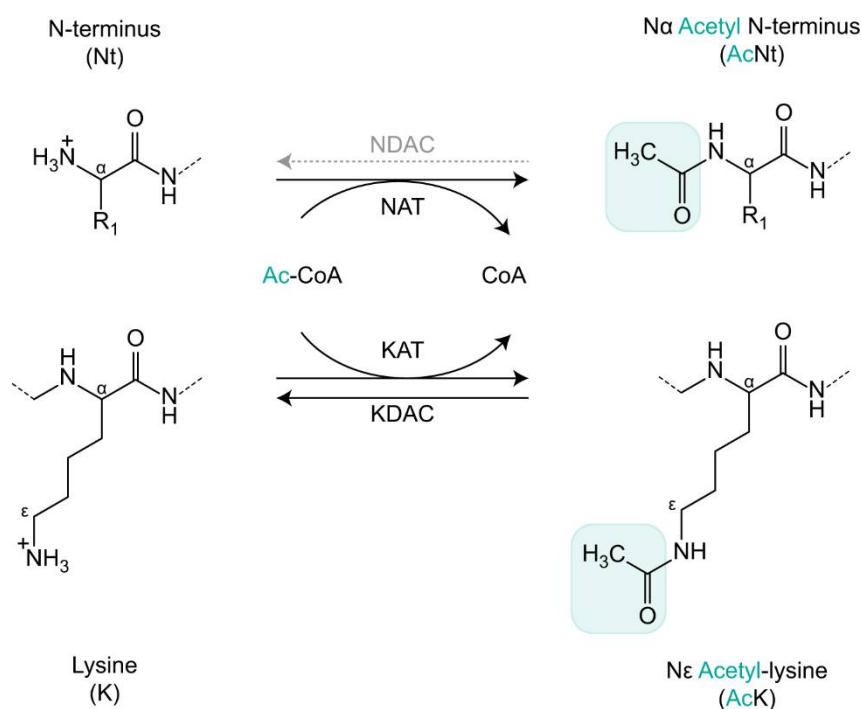


Figure 11: Schéma de l'acétylation de la protéine N-terminale et de la lysine [30]

La lysine-acétylation joue un rôle crucial dans la régulation de l'expression génique. L'acétylation des résidus de lysine localisés dans le site actif de certaines enzymes inhibe leur activité catalytique grâce à l'altération de la charge de ces résidus et de la morphologie du site actif [31].

Chez les bactéries de l'espèce *Pseudomonas aeruginosa*, un total de 117 protéines N α -acétylée ont été identifiées [32].

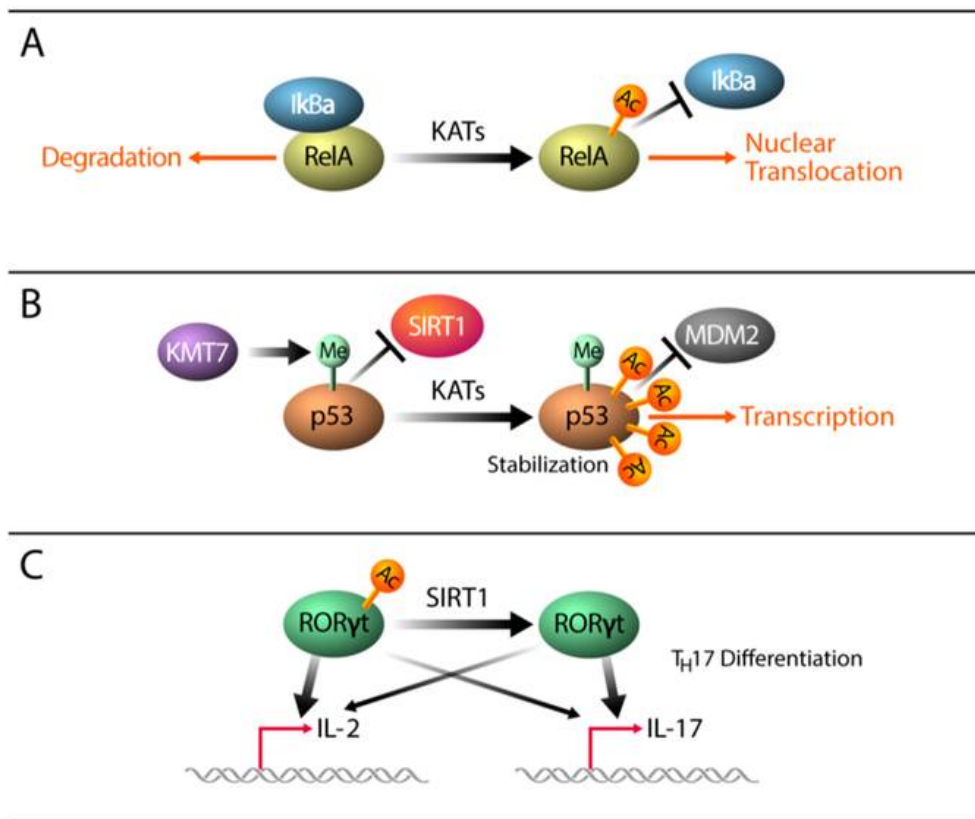


Figure 12: Mécanismes de régulation des facteurs de transcription acétylation-dépendante.

9.4. Les réactions de sulfatation :

La sulfatation des protéines désigne le transfert du groupement sulfate à une protéine ou un polypeptide. Ce transfert est assuré par les tyrosylprotéine-sulfotransférases (TPSTs), qui permettent la fixation d'un groupement sulfate sur le groupement hydroxyle des résidus de tyrosine, et qui se présentent chez l'homme et les mammifères sous deux isoformes : TPST1 et TPST2 [33].

Un autre type de sulfotransférases, le carbohydre sulfotransférase, assure le transfert du groupement sulfate à un résidu glucidique des glycoprotéines ou des carbohydrates. Les carbohydre sulfotransférases pourraient participer à certains mécanismes régulateurs comme la régulation de l'adhésion leucocytaire et la mobilisation des lymphocytes, comme c'est le cas pour les L-sélectines qui se lient spécifiquement aux récepteurs CD15s sulfatés [34].

Néanmoins, c'est la sulfatation des résidus de tyrosine par les TPSTs qui cible plus sélectivement les protéines et les polypeptides, et il est à noter que cette transformation est physiologiquement irréversible [35].

Le sulfate inorganique est une molécule inerte qui doit être tout d'abord activée avant son emploi dans un processus métabolique. L'activation du sulfate inorganique implique la formation du PAPS en deux étapes (Figure 13). Dans la première étape, le sulfate inorganique réagit avec l'ATP dans une réaction catalysée par l'ATP sulfurylase, qui aboutit à la formation du pyrophosphate et de l'APS. Dans la deuxième étape, l'APS réagit de nouveau avec l'ATP et aboutit à la formation du PAPS et de l'ADP. Cette réaction étant catalysée par l'APS kinase [36].

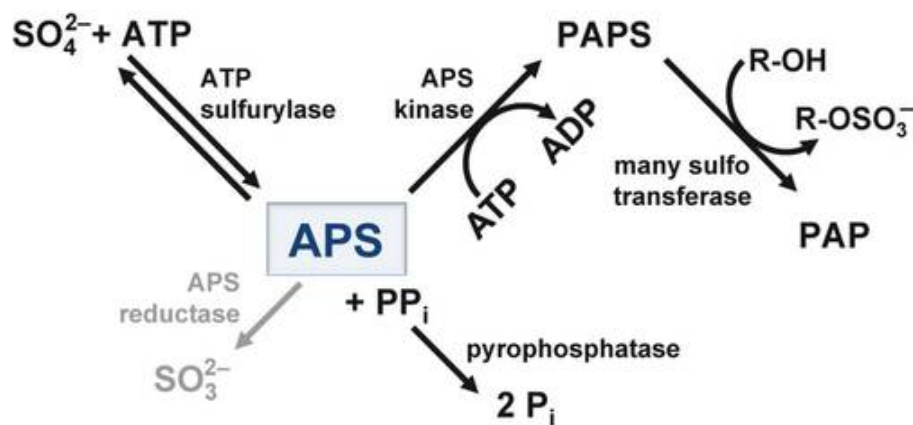


Figure 13: Schéma du métabolisme des sulfates [36]

Les souris naissantes dont les gènes TPST1 et TPST2 sont inactivés, ont une survie post-natale sévèrement affectée, principalement à cause d'une insuffisance cardiopulmonaire. Ces souris ont également développé une hypothyroïdie primaire [37].

La sulfatation des résidus de tyrosine des récepteurs des chimiokines paraît indispensable à leur fonctionnement. L'infection et la pénétration des cellules sanguines par le VIH-1 et l'agent du paludisme *Plasmodium vivax*, et la mobilisation des leucocytes sont favorisés par des résidus de tyrosine sulfatée dans les parties extra-cellulaires des récepteurs de chimiokines. La sulfatation des résidus de tyrosine du récepteur/antigène Duffy pour les chimiokines (DARC) est cruciale pour l'invasion des érythrocytes par le parasite du paludisme *plasmodium vivax* [38].

De plus, un anticorps humain (E51) dirigé contre le virus de VIH-1 possède un peptide pE51 qui porte des résidus de tyrosine sulfatée dans la région CDR3, ce qui suggère l'importance de cette sulfatation dans la reconnaissance de la glycoprotéine virale gp120 [39].

9.5. Les réactions de méthylation :

La méthylation est une réaction qui cible une panoplie de biomolécules à savoir l'ADN, les protéines et les métabolites secondaires. Quant à la méthylation des protéines, elle consiste à additionner un groupement méthyle à la fonction amine de la chaîne latérale d'un résidu aminoacide d'une chaîne polypeptidique. Ce transfert se produit à l'aide de méthyltransférases qui dépendent d'un co-substrat porteur et donneur du groupement méthyle – le S-Adénosylméthionine ou SAM. La méthylation des protéines se produit majoritairement sur les résidus de lysine.

L'arginine est un acide aminé qui subit aussi une méthylation catalysée par une famille d'enzymes appelées « arginine méthyltransférases » (PRMTs) et qui cible notamment les protéines cytoplasmiques et nucléaires dont les histones et les protéines non-histones [40]. Cet acide aminé est très basique et unique car il porte le groupe guanidine qui contient cinq donneurs potentiels de liaisons hydrogène [41], bien positionnés pour des interactions avec des accepteurs de liaisons hydrogène à savoir les acides nucléiques et les protéines. Pour chaque addition de groupement méthyle, un donneur potentiel d'hydrogène est enlevé, de sorte que la méthylation des résidus d'arginine dans les protéines modulera facilement leurs interactions avec d'autres biomolécules et donc leurs fonctions physiologiques.

La méthylation des résidus protéiques de l'arginine est impliquée dans diverses activités en relation avec l'expression génique, à savoir la transduction de signal, la régulation transcriptionnelle, la maturation de l'ARN, et la réparation de l'ADN [42]. Récemment, ce processus s'est avéré souvent aberrant et dérégulé dans certains néoplasmes, et les inhibiteurs de PRMTs sont investigués dans l'espoir de découvrir un nouveau traitement anti-néoplasique ciblant les PRMTs [43].

On compte à ce jour 9 variations de PRMTs chez les mammifères, qui sont classifiées en 3 types. Le type I est composé essentiellement de PRMTs 1, 2, 3, 4 (CARM1), 6 et 8, et catalyse les réactions de monométhylation et de diméthylation asymétrique, alors que le type II est représenté par les PRMT5 et la PRMT9 qui catalysent les réactions de monométhylation et de diméthylation symétrique, et finalement, le type III inclue seulement la PRMT7 et catalyse uniquement les réactions de monométhylation [44]. Il est à noter que l'activité catalytique majeure de la méthylation des résidus de l'arginine est assurée par la PRMT1 [45].

Il existe trois types de réactions de méthylation sur les résidus de l'arginine, qui aboutissent à trois types de produits : la ω -N^G-monométhylarginine (MMA), la ω -N^G,N^G-diméthylarginine asymétrique (ADMA) et la ω -N^G,N^G-diméthylarginine symétrique (SDMA) (Figure 14).

- **La ω -N^G-monométhylarginine :**

Les MMA sont retrouvées notamment dans les histones, et sont obtenues suite à l'action des arginines méthyltransférases de type I, II et III comme catalyseurs enzymatiques. Comme le nom du produit l'indique, il s'agit d'une addition d'un seul méthyle.

- **La ω -N^G,N^G-diméthylarginine asymétrique :**

Les ADMA sont obtenues suite à l'action des arginines méthyltransférases de type I comme catalyseurs enzymatiques. La double méthylation obtenue forme une asymétrie de part et d'autre de l'atome de carbone « ω ».

- **La ω -N^G,N^G-diméthylarginine symétrique:**

Les SDMA sont obtenues suite à l'action des arginines méthyltransférases de type II (PRMT5 et PRMT9) comme catalyseurs enzymatiques. Par contraste, la double méthylation obtenue cette fois-ci forme une symétrie de part et d'autre de l'atome de carbone « ω ».

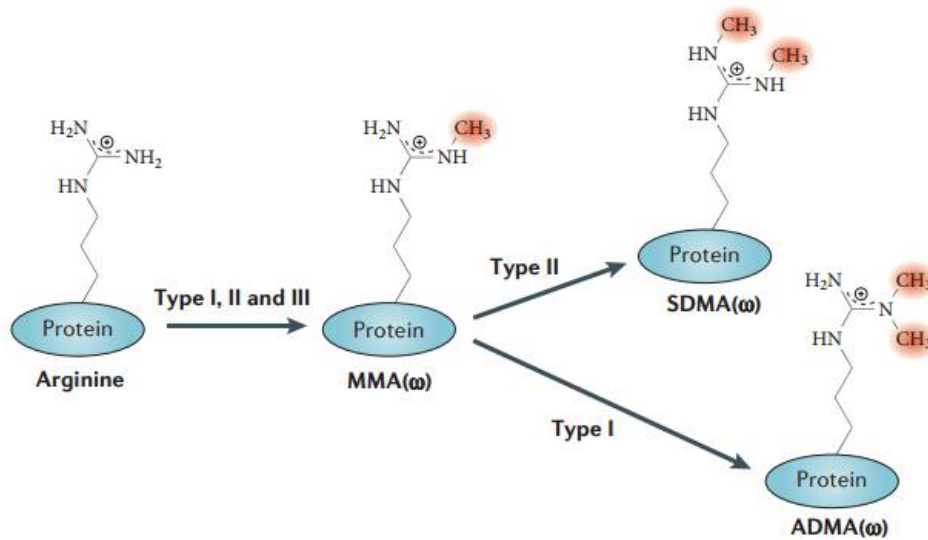


Figure 14: Les trois types de réactions de méthylation protéique sur les résidus de l'arginine

La méthylation des résidus de l'arginine et de lysine constitue l'une des plus importantes modifications post-traductionnelles des protéines histones, à côté des réactions de phosphorylation et d'acétylation [46]. Elle joue un rôle essentiel dans la régulation de la transcription. Le nucléosome est le constituant de base de la chromatine et joue un rôle essentiel dans les grandes fonctions nucléaires (transcription, réplication, réparation de l'ADN). Pour moduler son activité, la cellule met en œuvre trois voies principales: le recrutement de complexes de remodelage du nucléosome, les modifications post-traductionnelles des histones et l'incorporation des variants d'histones. Parmi ces modifications post-traductionnelles, on trouve la méthylation. La méthylation de la lysine peut avoir de multiples effets sur la fonction chromatiniennne, en fonction du résidu de lysine spécifique et du niveau de modification (mono-, di- ou triméthylation d'une seule lysine). Par exemple, la diméthylation de H3-K9 et la triméthylation de H3-K27 sont toutes deux largement associées à l'extinction des gènes et à la formation de l'hétérochromatine, tandis que la méthylation de H3-K4, H3-K36 ou H3-K79 est associée à une chromatine active [47].

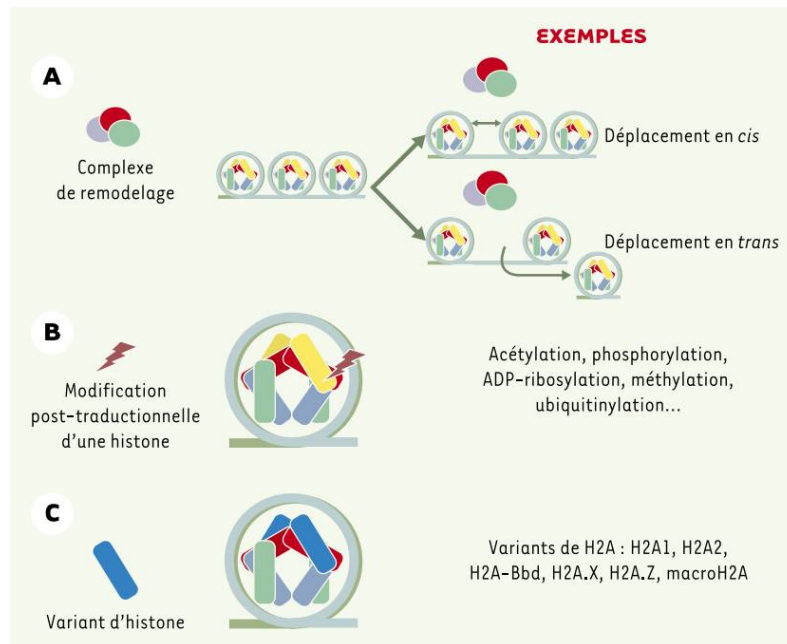


Figure 15: Les trois voies de modification du nucléosome [48].

(A. Déplacement en cis ou en trans des nucléosomes par un complexe de remodelage. B. Modification post-traductionnelle des extrémités des histones. C. Substitution d'une histone conventionnelle par un variant.)

(B.

9.6. Les réactions d'hydroxylation :

L'hydroxylation post-traductionnelle implique la transformation oxydative d'une liaison C-H en un groupe C-OH sur la chaîne latérale d'un acide aminé. L'hydroxylation transforme les molécules hydrophobes en molécules hydrophiles et donc les molécules hydroxylées interagissent plus facilement avec les fonctions polaires d'autres biomolécules.

Les représentants les plus connus de la famille des hydroxylases Fe(II)/ α -cétoglutarate dépendantes sont les prolyl et lysyl hydroxylases, qui servent comme catalyseurs principaux de l'hydroxylation post-traductionnelle des protéines. Ces enzymes utilisent le fer ferreux comme co-facteur et l' α -cétoglutarate comme co-substrat [49].

• Le rôle d'hydroxylation dans l'adaptation aux conditions hypoxiques :

L'hydroxylation de résidus de proline joue un rôle important dans la régulation de l'activité du facteur de transcription HIF (hypoxia inducible factor), un hétérodimère composé de la sous-unité HIF-1 α et la sous-unité HIF-1 β , dont le rôle a été élucidé dans les travaux de trois chercheurs qui ont été attribué le prix Nobel de médecine de l'année précédente (2019) [50], et qui ont pu identifier le mécanisme moléculaire qui régule l'expression de certains gènes en réponse aux variations des niveaux d'oxygène dans les tissus.

En cas d'hypoxie, l'hydroxylation des résidus de proline est inhibée. La sous-unité HIF-1 α forme un dimère avec HIF-1 β qui se lie aux gènes cibles sensibles au facteur HIF, tels que l'érythropoïétine (EPO) ou le facteur de croissance de l'endothélium vasculaire (VEGF), et plus d'autres (Figure 16).

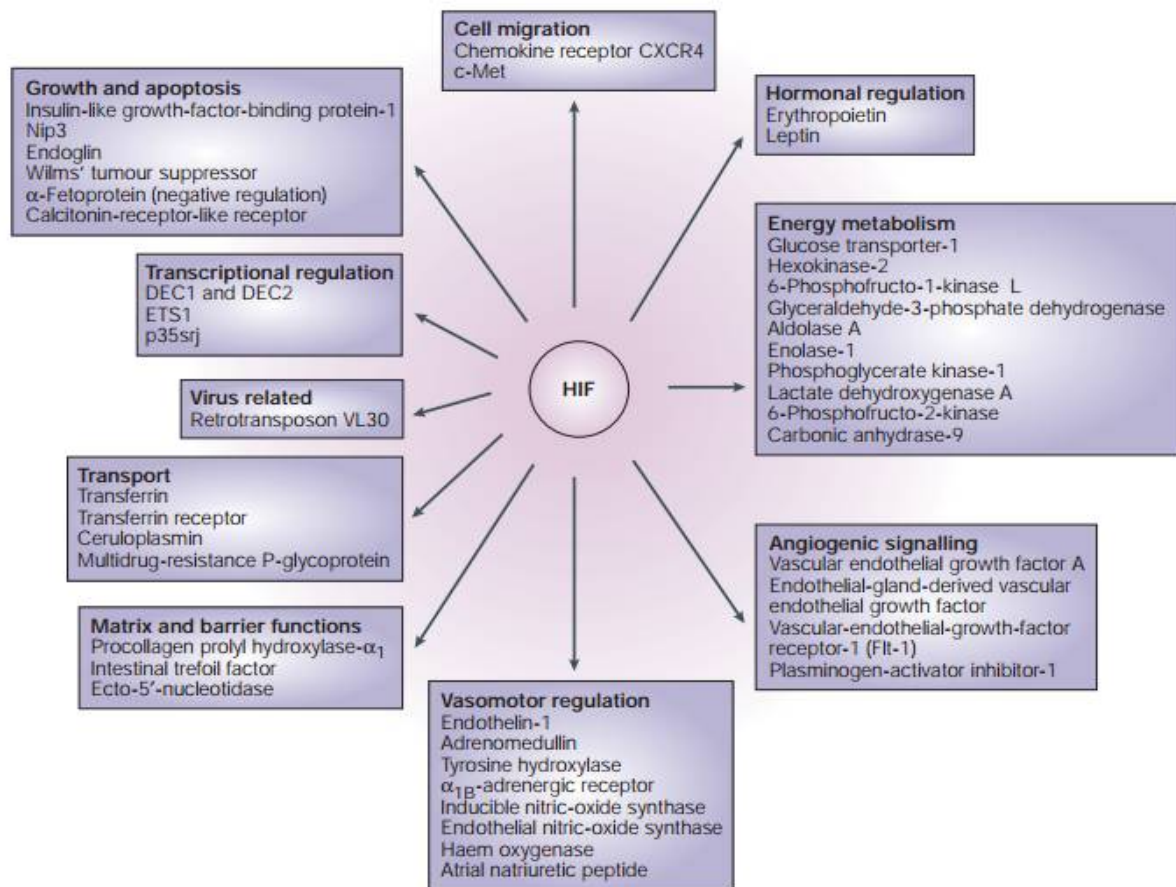


Figure 16: Cibles directes de la transcription du HIF [51]

Dans les conditions normoxiques, le HIF-1 (Factor inhibiting HIF 1) hydroxyle le résidu asparaginyle 803 (N) dans le domaine de transactivation (TAD) C-terminal de la sous-unité HIF-1 α , empêchant le recrutement de CBP/p300 dans le complexe HIF-1 (Figure 17).

Les PHDs (prolyl hydroxylase domain protein) hydroxylent deux résidus de prolyle (P402 et P564) situés au niveau du domaine de dégradation dépendant de l'oxygène (ODD) de la sous-unité HIF-1 α , et ceci augmente considérablement l'affinité de HIF-1 α pour le pVHL (protéine du gène suppresseur de tumeur de von Hippel-Lindau), permettant ainsi la liaison de pVHL à l'unité HIF-1 α . Par conséquent, le complexe formé est rapidement dégradé par les protéosomes suite à une ubiquitination préalable résultant de la reconnaissance du domaine ODD hydroxylé par l'ubiquitine ligase E3 [52]. Comme résultat final, l'activité transcriptionnelle de HIF-1 se trouve très réduite.

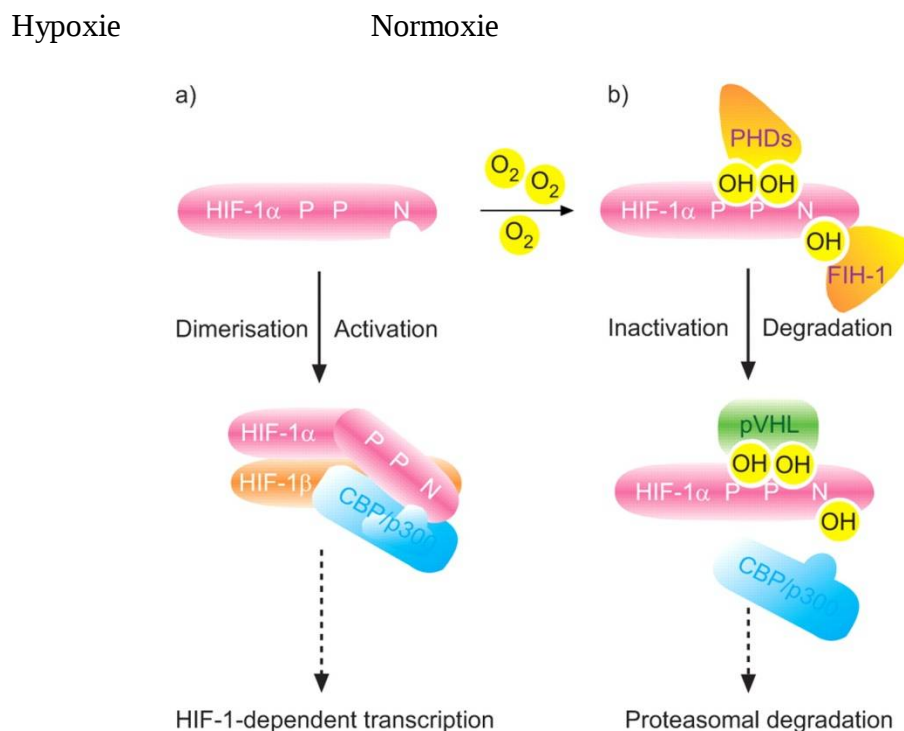


Figure 17: Contrôle de la transcription HIF-1 dépendante de l'oxygène [53]

De plus, l'induction du VEGF stimule la néovascularisation, ce qui peut renforcer la prolifération des cellules tumorales, d'où la nécessité de l'hydroxylation du facteur HIF-1.

- **Le rôle d'hydroxylation dans la motilité cellulaire :**

L'actine est une protéine intracellulaire biglobulaire présente dans toutes les cellules eucaryotes. Elle est l'unité structurelle des microfilaments qui font partie du cytosquelette. Les vertébrés expriment trois isoformes d'actine principales, dont l'isoforme α des muscles squelettiques, cardiaques et lisses et les isoformes β et γ exprimées dans les cellules non musculaires et musculaires [54]. Il a été démontré que les β -actines et γ -actines non-musculaires sont des substrats spécifiques du PHD3, et leur hydroxylation au niveau des résidus de proline par la PHD3 assure une régulation négative de la polymérisation de l'actine et de la motilité cellulaire. [55]

- **Le rôle d'hydroxylation dans la stabilité du collagène :**

Le collagène est la protéine la plus abondante chez les mammifères, et est un composant principal du tissu conjonctif. Il s'assemble en fibres qui forment une matrice structurelle et mécanique des os, de la peau, des tendons, des parois des vaisseaux sanguins et d'autres tissus conjonctifs. La 4-hydroxylation des résidus de proline qui composent le collagène renforce la stabilité de la structure en triple hélice du tropocollagène et influence la formation de fibrilles de collagène [56].

9.7. Les réactions de glycosylation :

Dans le processus de glycosylation, les unités glucidiques ou les oligopolysaccharides peuvent être liés de manière covalente aux protéines et modifiés par diverses réactions enzymatiques qui sont catalysées par les glycosyltransférases (GT) et les glycosidases, et ceci se produit principalement dans le réticulum endoplasmique et l'appareil de Golgi. Il faut préciser que la glycation de protéines qui se produit considérablement chez les diabétiques est une réaction spontanée et n'est pas effectuée par des enzymes, comme c'est le cas pour les réactions de glycosylation classiques [57].

Les glycanes peuvent former des chaînes linéaires ou ramifiées par la formation de liaisons glycosidiques α ou β avec tout groupement hydroxyle libre d'un autre monosaccharide. La présence d'une variété non négligeable de ces glucides disponibles pour la glycosylation des protéines participe activement à l'expansion du protéome, car les

modifications structurelles qu'apporte la glycosylation peuvent moduler les propriétés biophysiques et les caractéristiques thermodynamiques et cinétiques des protéines comme la stabilité et la solubilité [58, 59].

Les surfaces des cellules de mammifères sont ornées d'une grande variété de glycoconjugués complexes, dont la diversité des structures est fonction à la fois de la cellule ou du tissu où le glycoconjugué est exprimé. Ces glycoconjugués se présentent sous forme de glycolipides et de glycoprotéines, et c'est grâce aux éléments glucidiques que se produisent les principales interactions entre les différentes cellules et que se déroulent les processus de communication, de reconnaissance cellulaire et de détection de pathogènes (Figure 18). L'O-GlcNAcylation est une forme de glycosylation qui peut déclencher des changements dans l'activité et la stabilité des protéines, facilitant ainsi la transduction du signal comme c'est le cas pour la réponse cellulaire au stress oxydatif [60].

Une étude a démontré que la mutation des sites de glycosylation d'une protéine (RI₃₃₂) *in vivo* réduit l'efficacité du repliement et augmente la dégradation des protéines associée au réticulum endoplasmique (ERAD) [61].

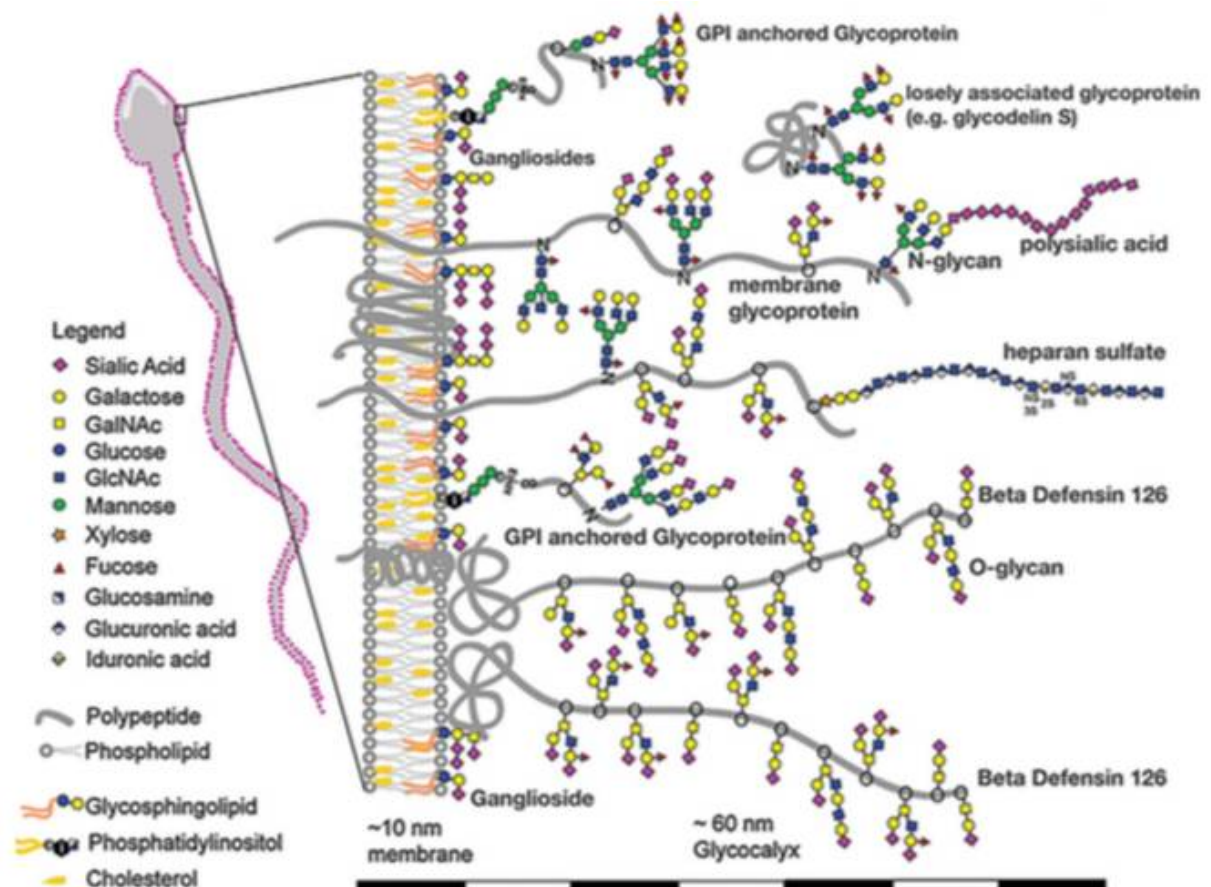


Figure 18: Principales classes de glycanes et de glycoconjugués du glycocalyx du spermatozoïde humain. [62]

• Les glycosyltransférases :

Les glycosyltransférases constituent une grande famille d'enzymes qui interviennent dans la biosynthèse des oligosaccharides et des polysaccharides. Ces enzymes sont généralement regroupées en familles en fonction du type de sucre qu'elles transfèrent (les galactosyltransférases, fucosyltransférases, sialyltransférases, etc...). Ces enzymes transfèrent les résidus de sucre d'un substrat donneur activé, généralement un nucléotide-sucre (UDP-glucose, UDP-GlcNAc etc...), à un accepteur qui peut être un lipide, une protéine ou un oligosaccharide en pleine croissance.

La perte de l'enzyme OGT (O-linked β -N-acetylglucosamine transferase) dans les cardiomyocytes des souris a permis le développement de multiples anomalies cardiaques (réduction de l'épaisseur de la paroi ventriculaire, développement anormal des vaisseaux coronaires...), voire une létalité post-natale précoce [63]. Ceci suggère que l'OGT joue un rôle essentiel dans la régulation du développement du tissu cardiaque en activant l'expression de l'angiopoïétine-1 (Angpt1) des cardiomyocytes.

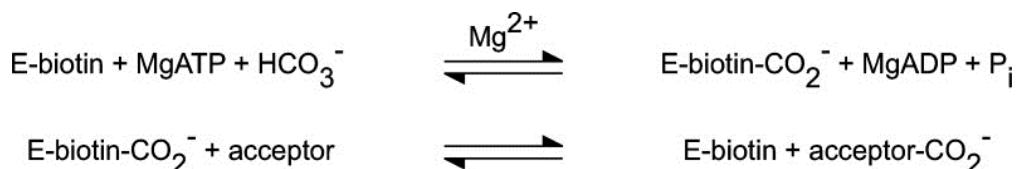
9.8. Les réactions de biotinylation :

La biotinylation des protéines est l'ancrage covalent de la biotine à une protéine, et ceci à l'aide d'une réaction catalysée par une biotine-protéine ligase (BPL), aussi connue sous le nom « holocarboxylase synthétase (HCS) » chez les mammifères.

La biotine (vitamine B8) est un membre du groupe des vitamines hydrosolubles du complexe B, et est synthétisée par les plantes supérieures et la plupart des bactéries et champignons. Elle est intégrée comme groupement prosthétique par quatre carboxylases biotine-dépendantes répandues dans les mitochondries et le cytoplasme des cellules humaines : l'acétyl-CoA carboxylase, la pyruvate carboxylase, la méthylcrotonyl-CoA carboxylase, et la propionyl-CoA carboxylase [64].

Ces enzymes sont cruciales dans plusieurs voies métaboliques, notamment la gluconéogenèse, le métabolisme des acides aminés et la biosynthèse des acides gras.

La biotine est activée par l'action ATP-dépendante de HCS qui conduit à la formation de la biotinyl-AMP. Cette dernière réagit avec une fonction nucléophile – le plus souvent la fonction ϵ -amine des résidus lysine – portée par l'apocarboxylase (forme inactive), et aboutit à la forme active de l'enzyme qui est l'holocarboxylase.



Les carboxylases biotine-dépendantes ont un mécanisme catalytique commun qui implique deux domaines à activité enzymatique. Le domaine de la biotine carboxylase (BC) fixe une molécule de CO₂ sur la protéine porteuse de biotine carboxyle (BCCP), et le domaine de carboxyltransférase (CT) transfère le CO₂ au substrat spécifique de chaque carboxylase (Figure 19).

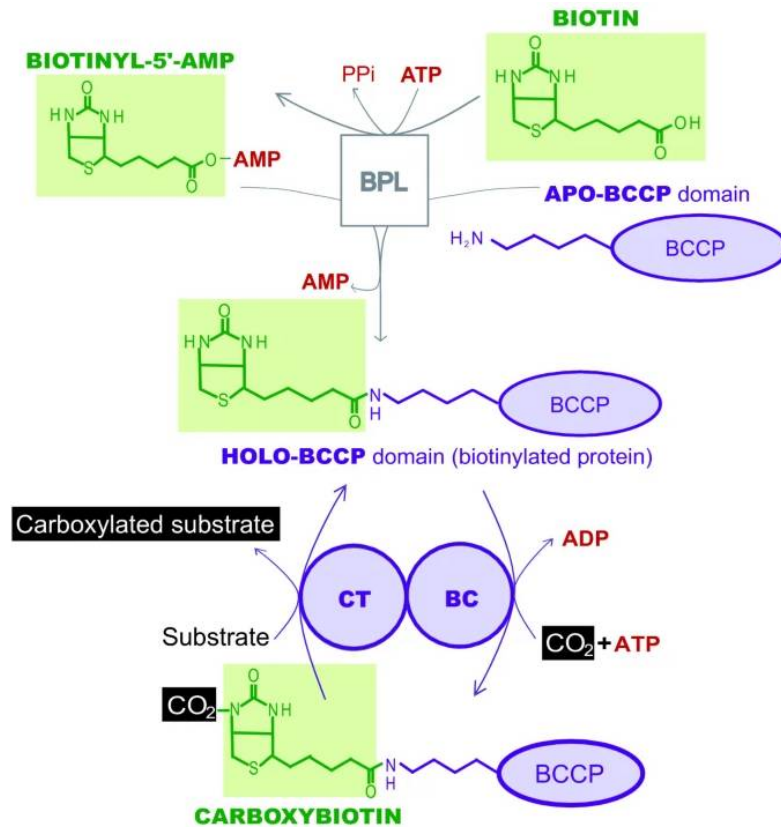


Figure 19: Schéma montrant la réaction catalysée par la biotine-protéine ligase (BPL) et le mécanisme général commun des carboxylases biotin-dépendantes [65]

- **Mortalité post-natale suite au déficit de l'HCS :**

Comme la liaison du dioxyde de carbone à l'apocarboxylase inactive se fait par l'intermédiaire de la biotine, l'absence de l'HCS dans l'organisme se traduit par une absence de l'activité des carboxylases. Le déficit multiple en carboxylase est une maladie néonatale rare mais potentiellement mortelle, caractérisée par l'absence d'activité des carboxylases en raison d'une déficience à caractère autosomique récessif de l'activité de l'HCS [66].

9.9. Les réactions de prénylation :

La prénylation (ou isoprénylation) des protéines est une modification post-traductionnelle covalente et irréversible, impliquant la fixation d'un isoprénoïde à 15 atomes de carbone (farnésyle) ou à 20 atomes de carbone (géranylgéranyle) au résidu cystéine d'un motif térapeptidique C-terminal de reconnaissance CaaX ("C" est une cystéine, "a" est un acide aminé aliphatique, et "X" est un des autres acides aminés). Il est à noter que la nature du résidu X détermine si une protéine sera farnésylée ou géranylgéranylée. Les protéines se terminant par la sérine, la méthionine, l'alanine ou la glutamine deviennent farnésylées, tandis que les protéines se terminant par la leucine ou l'isoleucine sont géranylgéranylées [67].

La prénylation se trouve dans toutes les cellules eucaryotes et est catalysée par trois prényltransférases différentes : la protéine farnesyltransférase (FTase), la géranylgéranyltransférase de type 1 (GGTase-I) et la géranylgéranyltransférase de type 2 (GGTase-II ou RabGGTase) (Tableau 4). Ces enzymes utilisent comme coenzyme le farnésyl-pyrophosphate et le géranylgéranyl-pyrophosphate, deux intermédiaires issus de la voie du mévalonate qui assure la biotsynthèse des isoprénoïdes chez les eucaryotes [68]. Un quatrième élément précédemment inconnu s'ajoute à ces prényltransférases, la géranylgéranyltransférase de type 3 (GGTase-III), qui a été récemment décrite comme enzyme effectuant la géranylgéranylation de la protéine FBLX2, et favorisant le ciblage membranaire des ubiquitines ligases. [69]

Les molécules qui subissent massivement les réactions de prénylation sont les GTPases des familles protéiques Ras, Rho et Rab.

● La géranylgéranylation :

Elle implique la fixation d'un résidu géranylgéranyl à une cystéine située à l'extrémité C-terminale d'une protéine cible. Ces résidus agissent en tant qu'un ancrage membranaire pour la protéine cible modifiée, et ce grâce à leur excellente hydrophobicité, déclenchant la translocation de la protéine du cytosol vers la membrane biologique et favorisant son interaction avec les autres zones hydrophobes de plusieurs protéines [70].

La géranylgéranlation des protéines dans les cellules et catalysée par deux enzymes, la GGTase-I et la GGTase-II. Contrairement à la FTase et la GGTase-I, la géranylgéranyltransférase de type II ne reconnaît pas directement une séquence C-terminale tétrapeptidique de type CaaX, et elle a besoin de l'assistance d'une protéine adaptatrice appelée protéine d'escorte de Rab (REP) pour reconnaître les motifs XCCXX, XXXCC, XXCXC, XXCCX, CCXXX, ou XCXXX [71]. La géranylgéranyltransférase de type II transfère deux groupes géranylgéranyles aux résidus de cystéine de ces motifs localisés à l'extrémité carboxyle des protéines Rab d'une façon dépendante de la protéine d'escorte de Rab [72].

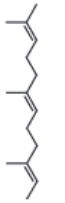
- **La farnésylation :**

La farnésylation est l'une des étapes importantes de la modification post-traductionnelle des protéines associées à la transduction des signaux intracellulaires. Elle est catalysée par une métalloenzyme hétérodimérique connue sous le nom de farnésyltransférase, et implique le transfert d'un groupe farnésyle du farnésylpyrophosphate au groupement sulfure de la cystéine C-terminale des protéines cibles, notamment Ras [73].

La protéine Ras est une GTPase monomérique nécessitant une farnésylation pour son activation. L'activation constitutive de Ras étant un facteur contributif majeur dans un certain nombre de tumeurs malignes, son inactivation en interférant avec l'étape de farnésylation a été largement étudiée comme stratégie de développement de nouveaux agents anticancéreux. Les inhibiteurs de la farnésyltransférase (ITF) sont une nouvelle classe d'agents anticancéreux (ex : Tipifarnib) développés pour inhiber l'enzyme farnésyltransférase qui est responsable du transfert d'un groupe farnésyl à la protéine Ras [74].

Table I

Features of human protein prenyltransferases

	Farnesyltransferase (FT)	Geranylgeranyltransferase I (GGT I)	Geranylgeranyltransferase II (GGT II or RabGGT)
Gene name (α subunit)	<i>FNTA</i>	<i>FNTA</i>	<i>RABGGTA</i>
Gene name (β subunit)	<i>FNTB</i>	<i>PGGT/B</i>	<i>RABGGTB</i>
Substrate motif	Carboxy-terminal -Ca ₁ a ₂ X box*	Carboxy-terminal -Ca ₁ a ₂ X box*	Carboxy-terminal motif such as -CC, -CXC, -CCX, -CCXX, -CCXXX, or -CXXX†
Lipid anchor	 Farnesyl	 Geranylgeranyl	 Geranylgeranyl (one or two moieties per substrate)

*C, cysteine; a₁, and a₂ are amino acids that are preferably aliphatic; the identity of a₁ is more flexible than a₂. X can be C, S, Q, A, M, T, H, V, N, F, G, or I for FT, or L, F, I, V, or M for GGT I (in decreasing order of affinity). †Requires an escort protein such as REP1.

Tableau 4: Caractéristiques des protéines prényltransférases humaines [75]

9.10. Les réactions d'ubiquitination :

A un moment donné, une protéine défective ou celle ayant accomplie la mission physiologique qui lui a été désignée, doivent disparaître définitivement, et ce grâce, au moins en partie, au mécanisme de contrôle qualité protéolytique qui est étroitement contrôlé par l'ubiquitine.

L'ubiquitine est une petite protéine composée de 76 acides aminés (~8 kDa) qui, après son expression, participe au marquage des protéines principalement pour leur dégradation intracellulaire assurée par le système ubiquitine/protéasome, une des voies majeures de la dégradation protéique chez les eucaryotes [76]. Cette protéine est conservée au cours de l'évolution chez tous les organismes eucaryotes, et est absente chez les procaryotes.

Au cours de l'ubiquitination, l'ubiquitine est attachée au résidu de lysine d'une protéine cible par une liaison isopeptidique qui se forme entre la fonction carboxyle C-terminale du résidu glycine de l'ubiquitine et la fonction amine de la chaîne latérale des résidus lysine. Ce processus s'effectue via une cascade de réactions enzymatiques réalisées par l'enzyme d'activation E1 (*ubiquitin-activating enzyme*), de conjugaison E2 (*ubiquitin-conjugating enzyme*) et de ligature E3 (*ubiquitin protein-ligase*), successivement.

L'enzyme d'activation E1 fixe d'une façon ATP-dépendante une ubiquitine par une liaison thioester, activant ainsi l'ubiquitine et attirant les enzymes E2. L'ubiquitine activée est ensuite délivrée à l'enzyme de conjugaison E2, et sera davantage transférée de l'E2 à un résidu de lysine d'une protéine cible par l'action de l'ubiquitine ligase (E3) (Figure 20). Cette dernière assure la reconnaissance spécifique du substrat via le domaine de liaison à l'ubiquitine (*Ubiquitin-binding domain*, UBD), ce qui reflète le grand nombre d'ubiquitines ligases qui dépasse les 500 [77].

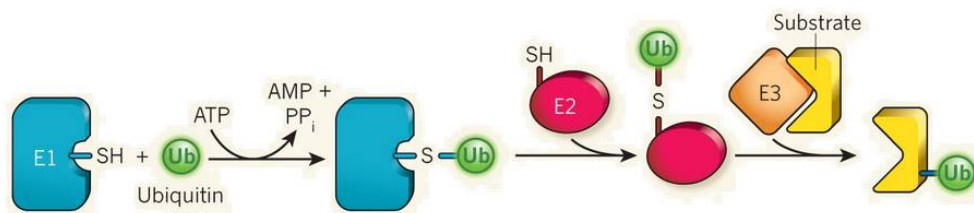


Figure 20: Schéma simplifié des 3 réactions de l'ubiquitination d'un substrat protéique [78]

Comme pour les E1s et les E2s, les E3s de type HECT (*homologous to E6AP C-terminus*) et RBR (*RING between RING*) peuvent également former une liaison thioester avec l'ubiquitine avant sa fixation sur le substrat. Contrairement, les E3s de type RING (*really interesting new gene*)/U-box forment un pontage entre une E2 activée et son substrat, et par conséquent, la fixation de l'ubiquitine sur le substrat se fait directement, sans passer par l'intermédiaire E3-ubiquitine lié par liaison thioester [79] (Figure 21).

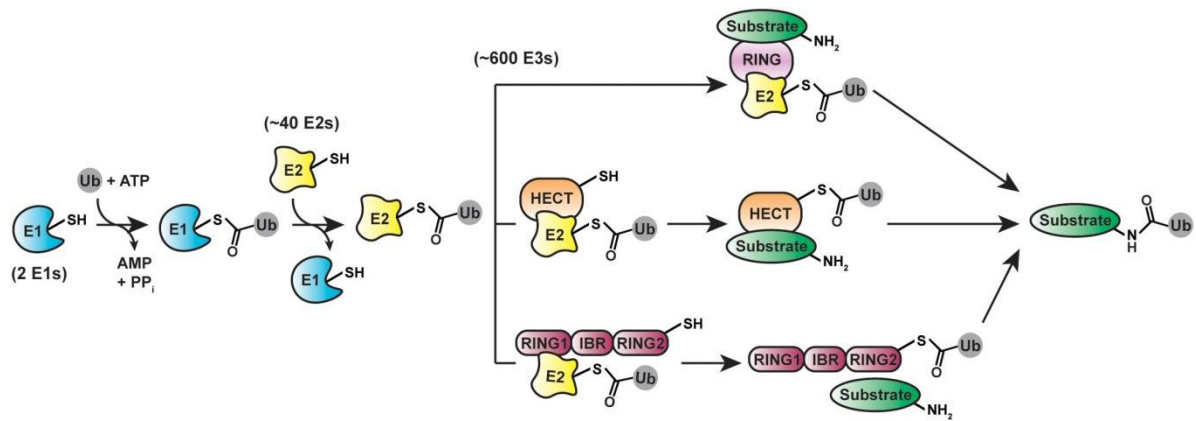


Figure 21: Les trois mécanismes de ligation de l'ubiquitine [80]

Une ou plusieurs ubiquitines peuvent être attachées sur un substrat protéique, via deux processus distincts appelées monoubiquitination et polyubiquitination.

• La monoubiquitination :

La monoubiquitination est le type d'ubiquitination le plus basique. Il s'agit tout simplement de l'attachement d'une seule ubiquitine à une protéine cible. Lorsque plusieurs molécules d'ubiquitine sont attachées individuellement aux différents résidus de lysine, on parle de **multi-mono-ubiquitination**.

Les monoubiquitines sont souvent impliquées dans la régulation de la réparation de l'ADN, la réplication de l'ADN, l'endocytose, la régulation des processus apoptotiques et la régulation de la transcription [81].

• La polyubiquitination :

L'extrémité C-terminale de l'ubiquitine peut être conjuguée à l'une des sept résidus lysines (K6, K11, K27, K29, K33, K48, K63) ou au résidu méthionine M1, portés par une autre ubiquitine liée à un substrat (Figure 22), formant ainsi une chaîne d'au moins 4 molécules d'ubiquitine de différentes liaisons et de différents branchements (Figure 23). Cette réaction de polymérisation est appelée « polyubiquitination ».

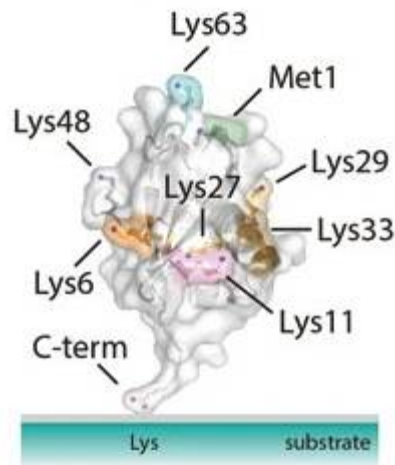


Figure 22: Structure de l'ubiquitine mettant en évidence les huit sites de polyubiquitination [82]

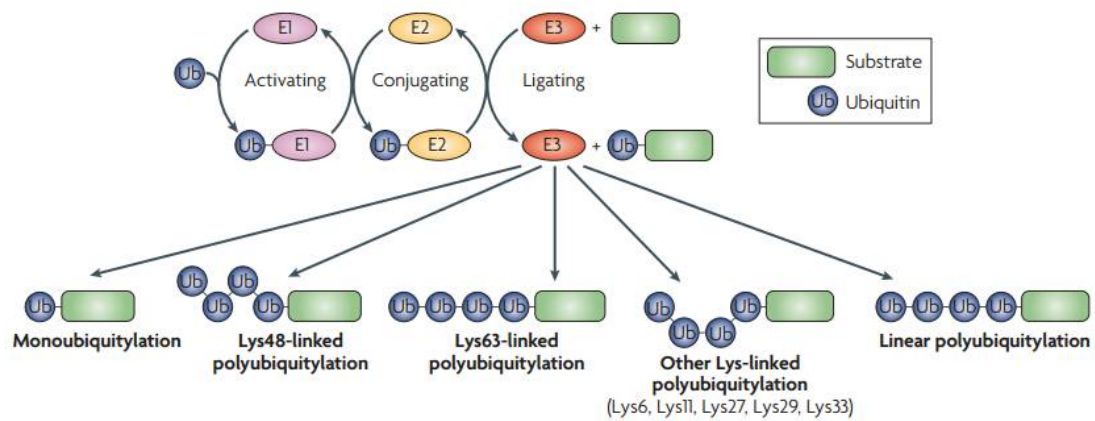


Figure 23: Les 3 étapes enzymatiques aboutissant aux différents types d'ubiquitination [83]

La polyubiquitination a un rôle fonctionnel notamment dans la détermination du destin et de la fonction de la protéine polyubiquitinée. A titre d'exemple, la polyubiquitination de type K48/K11 est généralement impliquée dans la préparation d'une protéine cible à la dégradation dépendante du protéasome 26S, alors que la polyubiquitination de type K63 est souvent impliquée dans des processus non-protéolytiques tel que la réparation de l'ADN, l'activation de protéines kinases, l'endocytose et la signalisation cellulaire [84].

La polyubiquitination est plus complexe que la monoubiquitination, car suivant la nature de branchement, on distingue deux principales catégories de polyubiquitines : les polyubiquitines homotypiques qui portent des ubiquitines régulièrement liées via le même résidu de lysine, et les polyubiquitines hétérotypiques dans lesquelles les liaisons se produisent à plusieurs positions de lysine (Figure 24). Cette dernière catégorie comporte les polyubiquitines ramifiées ayant une ubiquitination à plus de deux sites sur une seule molécule d'ubiquitine. Il a été démontré que les chaînes hétérotypiques K11/K48-polyUb fixées sur la cycline B1 forment une interaction de haute affinité avec le protéasome 26S plus que les chaînes homotypiques K11-polyUb, ce qui souligne l'utilité de la topologie des polyubiquitines dans la détermination du devenir du substrat protéique lié [85].

La fonction spécifique d'une protéine polyubiquitinée peut être précisée voire régulée via l'introduction de modifications post-traductionnelles à l'instar de la phosphorylation ou l'acétylation, qui rajoute davantage une complexité structurelle aux polyubiquitines. *In vitro*, les ubiquitines acétylées assurent l'inhibition de l'élongation de la polyubiquitine et stabilisent les monoubiquitines [86].

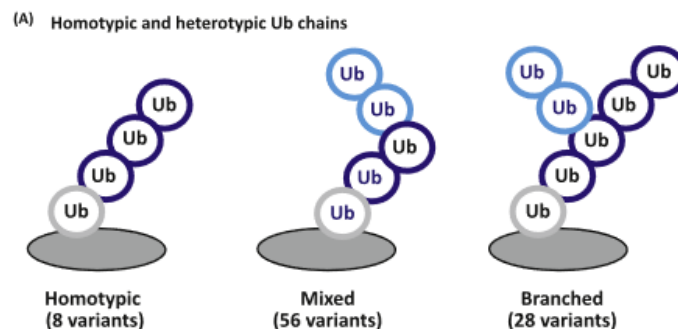


Figure 24: Les chaînes de polyubiquitine homotypique (à gauche) et de polyubiquitine hétérotypique (les deux à droite) [87]

• La désubiquitination :

La régulation rapide de l'ubiquitination est rendue possible par les enzymes de désubiquitination (DUBs) pouvant partiellement découper les chaînes de polyubiquitine voire éliminer l'ensemble des molécules d'ubiquitine portées par une protéine.

● La reconnaissance de l'ubiquitine :

L'interaction de protéines monoubiquitinylées ou polyubiquitinylées avec leur cible n'aura jamais lieu si les décodeurs du message porté par les résidus d'ubiquitine - les domaines de liaison à l'ubiquitine (UBD) - n'existaient pas. Jusqu'à ce jour, on dénombre plus de 20 familles des UBDs qui diffèrent selon leur repliement et leur topologie, convenablement orientés afin de détecter le type d'ubiquitination porté par le substrat [88].

Ces UBDs peuvent être regroupés en 4 différentes classes en fonction des caractéristiques structurales spécifiques : les doigts de zinc, le domaine PH (Pleckstrin homology), les hélices alpha, les autres UBDs non classifiées. Bien que structurellement très distincts, presque tous les UBDs interagissent par une liaison non covalente avec la zone hydrophobe de l'ubiquitine formée par Leu8, Ile44 et Val70 [89].

● Les protéines « ubiquitin-like » et l'autophagie sélective :

Les protéines « ubiquitin-like » ou Ubl, sont des protéines qui partagent la même structure tridimensionnelle – le pli β -grasp – avec l'ubiquitine. Chez l'homme, on dénombre 9 familles de protéines Ubl : NEDD8 (Rub1), SUMO, URM1, UFM1, ATG8, ATG12, FAT10, FUB1 et ISG15 [90].

La formation de l'autophagosome est réalisée par un ensemble de protéines ATG (*Autophagy-related*), hautement conservées de la levure jusqu'aux mammifères. Les ATG8 sont des protéines Ubl (*ubiquitin-like*) qui jouent un rôle essentiel dans la formation de l'autophagosome, sa maturation, son aptitude à recruter le matériel cellulaire à dégrader, et sa liaison avec les lysosomes [91]. Contrairement à l'ubiquitine, qui est conjuguée aux résidus de lysine des protéines cibles, les protéines ATG8 se fixent au groupement amine des lipides membranaires (surtout celui du phosphatidyléthanolamine) par la même triade enzymatique de l'ubiquitination.

● Rôle de l'ubiquitination dans l'activation de la voie NF- κ B :

Le facteur de transcription NF- κ B est un régulateur clé de plusieurs processus cellulaires, tels que l'apoptose, les réponses immunitaires innées et adaptatives, la migration et la prolifération cellulaires. L'activation de la voie NF- κ B est étroitement régulée par un

certain nombre de modifications post-traductionnelles, notamment la phosphorylation et l'ubiquitination. Dans les cellules non stimulées, NF- κ B se lie aux protéines inhibitrices de la famille κ B (I κ B) et est séquestré dans le cytoplasme. Lors de la stimulation, I κ B est phosphorylé par le complexe I κ B kinase (IKK), qui se compose de deux kinases IKK α et IKK β , ainsi que de la sous-unité régulatrice essentielle (NEMO, également appelée IKK γ ou IKKAP). L'I κ B phosphorylé est ensuite ubiquitiné et dégradé par le protéasome 26S, permettant ainsi au NF- κ B de se transloquer vers le noyau, où il régule l'expression de plusieurs gènes cibles [92].

9.11. Les réactions de l'ADP-ribosylation :

Parmi les modifications post-traductionnelles réversibles, à l'instar de l'ubiquitination et la phosphorylation, il y a l'ADP-ribosylation, qui est définie comme un transfert d'une ou plusieurs unités d'ADP-ribose du coenzyme nicotinamide adénine dinucléotide (NAD⁺) sur des substrats protéiques ou nucléiques cibles, en cédant une molécule de nicotinamide. L'ADP-ribose est composée de deux résidus osidiques reliés par deux phosphates, dont l'un est attaché à un résidu d'adénine (Figure 25).

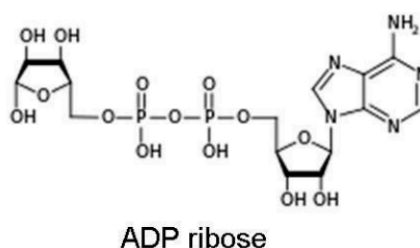


Figure 25: Structure chimique de l'ADP-ribose, l'unité de base de l'ADP-ribosylation [93]

Plusieurs exotoxines bactériennes célèbres, dont la toxine du choléra, la toxine diphtérique et la toxine pertussique sont des ADP-ribosyltransférases.

Le niveau d'intérêt scientifique suscité par l'ADP-ribosylation est lié à ses implications physiologiques cruciales notamment dans les réactions pro-inflammatoires, la réparation de l'ADN, et la division cellulaire.

Deuxième partie (II) : METHODES ET OUTILS D'ETUDE PROTEOMIQUE

1. Introduction :

La protéomique est devenue l'une des technologies de pointe à la disposition des chercheurs de l'ère post-génomique, en raison du rôle crucial et complexe des protéines et des interactions protéine-protéine dans la fonction cellulaire. En tant que discipline post-génomique, la protéomique englobe les efforts visant non seulement à identifier et à quantifier l'ensemble des protéines d'un protéome, mais aussi de déterminer la localisation cellulaire, l'expression, les interactions protéine-protéine, et les modifications post-traductionnelles (PTM), ce qui rend l'étude complète d'un protéome un défi plus exigeant que le séquençage d'un génome. Or, la protéomique offre des connaissances biologiques importantes que la génomique et la transcriptomique n'abordent pas si facilement. Au cours des dernières années, les technologies d'analyse protéomique se sont concentrées sur les analyses par spectrométrie de masse, l'électrophorèse bidimensionnelle sur gel, sans oublier l'importance primordiale des outils bio-informatiques qui ont permis le traitement aisé de gigantesques données protéiques. La protéomique à haut débit a été rendu possible grâce à des instruments modernes de haute résolution, capables par exemple, d'identifier plus de 2500 protéines en une durée de 90 minutes, par l'utilisation d'un spectromètre de masse hybride (Q Exactive) utilisant un filtre quadripolaire et un Orbitrap [94].

Jusqu'à la fin des années 1980, le séquençage des protéines était principalement réalisé par la méthode de dégradation d'Edman. Au cours de ce processus, l'isothiocyanate de phényle réagit avec un groupement amine N-terminal pour former un dérivé de phénylthiocarbamoyle qui est ensuite clivé pour produire un dérivé de thiazolinone et une nouvelle terminaison N. La thiazolinone AA libérée est stabilisée, puis identifiée par électrophorèse ou chromatographie. Ce processus est ensuite répété pour trouver la séquence peptidique. Le processus d'Edman est lent et il est limité aux peptides de quelques résidus par la dérivation cyclique, mais lorsqu'il est bien fait, il est précis avec une efficacité de >99% par AA. Il nécessite environ 100 pmol de peptide pur, mais il ne rend pas toujours fidèlement compte des PTM car la dégradation d'Edman ne fonctionne pas sans un groupe aminé libre α -sur le terminus N.

Pourtant, ce n'est qu'en 1988 que John Fenn a mis au point une méthode d'analyse particulière, nommée « ionisation par électrospray », qui permet d'appliquer la spectrométrie de masse à l'étude de grosses molécules, notamment les protéines. Le prix Nobel de chimie 2002 a été décerné pour “le développement de méthodes d'identification et d'analyse de la structure des macromolécules biologiques”, la moitié étant attribuée conjointement à John B. Fenn et Koichi Tanaka pour son “développement de méthodes d'ionisation par désorption douce pour les analyses par spectrométrie de masse des macromolécules biologiques”.

2. Les bases de l'analyse protéomique :

Deux approches différentes ont été initialement développées pour l'analyse d'un échantillon protéique. La grande majorité de l'analyse protéomique est basée sur l'approche dite « ascendante », où le mélange complexe de protéines d'une cellule ou d'un tissu est digéré enzymatiquement en un mélange de peptides, au contraire de l'autre approche complémentaire dite « descendante », dans laquelle les protéines ne subissent pas de digestion préalable avant son analyse par spectrométrie de masse. Néanmoins, une troisième approche appelée « *middle-down* » a été mise en place et commence à gagner sa place dans l'analyse protéomique (Figures 26, 27).

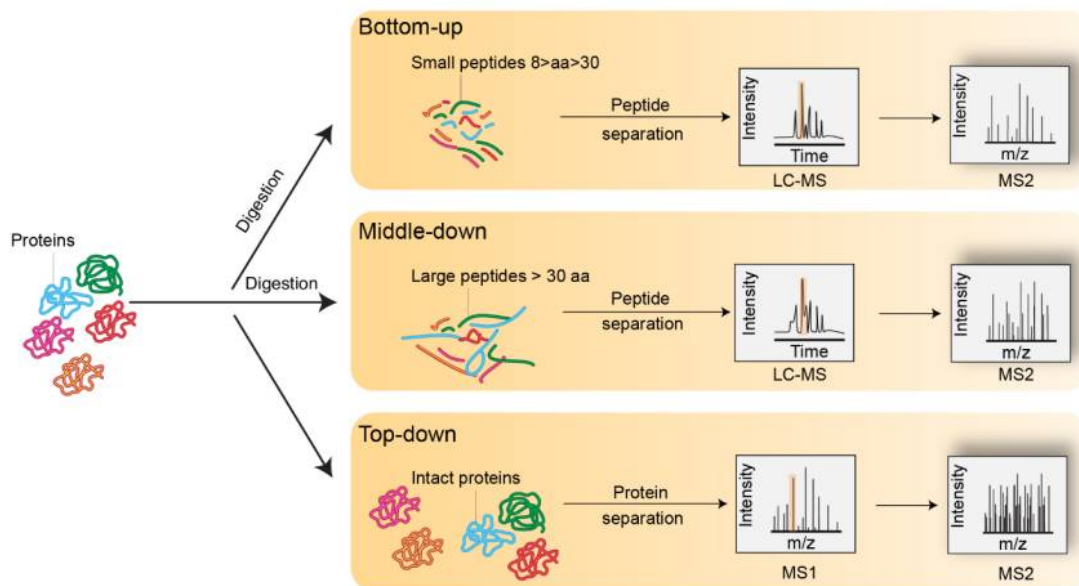


Figure 26: Les 3 approches d'analyse de protéines par spectrométrie de masse en tandem

2.1. La protéomique descendante :

La protéomique descendante ou *top-down* représente une forme directe d'analyse de protéines. Par rapport à la protéomique ascendante et à la protéomique *middle-down*, la protéomique descendante présente l'avantage de préserver des protéoformes intactes et elle est de plus en plus utilisée pour l'analyse protéomique.

Dans l'analyse *top-down* par spectrométrie de masse, les ions protéiques intacts sont introduits dans la phase gazeuse par ionisation par électronébuliseur (Electrospray ionization ou ESI) et sont ensuite fragmentés couramment par dissociation induite par collision (Collision-induced dissociation ou CID), par dissociation par capture d'électrons (ECD) ou dissociation par transfert d'électrons (ETD) dans le spectromètre de masse, ce qui donne les masses de la protéine intacte ionisée et des ions fragmentés.

2.2. La protéomique ascendante :

La protéomique ascendante ou *bottom-up* est la technique qui a connu l'intérêt et l'usage majeurs dans les analyses protéomiques que la protéomique *top-down/middle-down*. Elle représente une forme indirecte d'analyse protéomique car l'analyte original, qui est la protéine intacte, subit d'abord une digestion protéolytique aboutissant à plusieurs peptides de petites masses, dont les masses sont utilisées ultérieurement pour identifier la protéine source [95]. Cette digestion protéolytique se produit généralement au moyen d'une protéase, essentiellement une trypsine, ou moins souvent par clivage chimique, par exemple en utilisant du bromure de cyanogène, générant ainsi des peptides courts ayant moins de 30 acides aminés. Afin d'identifier la source protéique des peptides, les peptides générés agissent comme des empreintes et sont ensuite corrélés et tracés à leur origine protéique par inférence, et ce à l'aide de bases de données accessibles via des moteurs de recherche tels que Mascot, MaxQuant/Andromeda ou Sequest, qui contiennent un panel de peptides générés « *in silico* » [96].

La préparation des échantillons pour la protéomique ascendante comprend plusieurs étapes critiques :

(1) Extraction et solubilisation des protéines ;

(2) Dénaturation des protéines ;

(3) Digestion enzymatique (parfois chimique) ;

(4) Nettoyage des peptides, y compris l'élimination du détergent et le dessalage ;

(5) Séparation des peptides, normalement réalisée par chromatographie liquide

2.3. La protéomique « *middle-down* » :

La protéomique « *middle-down* » est un terme qui a été inventé pour la première fois en 2009, et constitue une approche intermédiaire qui a émané d'une petite modification de l'approche ascendante. Comme la protéomique ascendante, cette approche repose également sur la digestion enzymatique des protéines suivie d'une analyse spectrométrique et d'une identification par inférence. Par contre, l'approche « *middle-down* » est plus récente, et sa principale caractéristique est que la taille des peptides résultants serait supérieure à celle des plus petits peptides habituellement générés dans l'approche ascendante, car l'approche « *middle-down* » impliquerait la génération de peptides protéolytiques d'une longueur supérieure à environ 20 résidus d'acides aminés et allant jusqu'à environ 100 résidus d'acides aminés (des masses moléculaires entre 2.5 kDa et 10 kDa).

Le nombre de peptides générés dans un échantillon produit par l'approche « *middle-down* » serait inférieur au nombre de peptides produits par les protocoles typiques de la protéomique ascendante. Cela dit, il est donc plus probable de détecter un plus grand nombre de peptides uniques en utilisant cette approche.

Comme c'est le cas pour l'approche ascendante, l'approche « *middle-down* » peut utiliser également des protocoles qui emploient la trypsine comme enzyme de digestion protéolytique limitée. GluC, AspN, et LysC sont des protéases pouvant également être utilisées dans cette approche. De plus, l'approche « *middle-down* » est très convenable pour l'analyse des histones, car la queue N-terminale de l'histone intacte (50-60 acides aminés) qui porte la plupart des modifications post-traductionnelles peut être facilement clivée à l'aide d'enzymes telles que GluC (clivage après l'acide glutamique) ou AspN (clivage avant l'acide aspartique) [97].

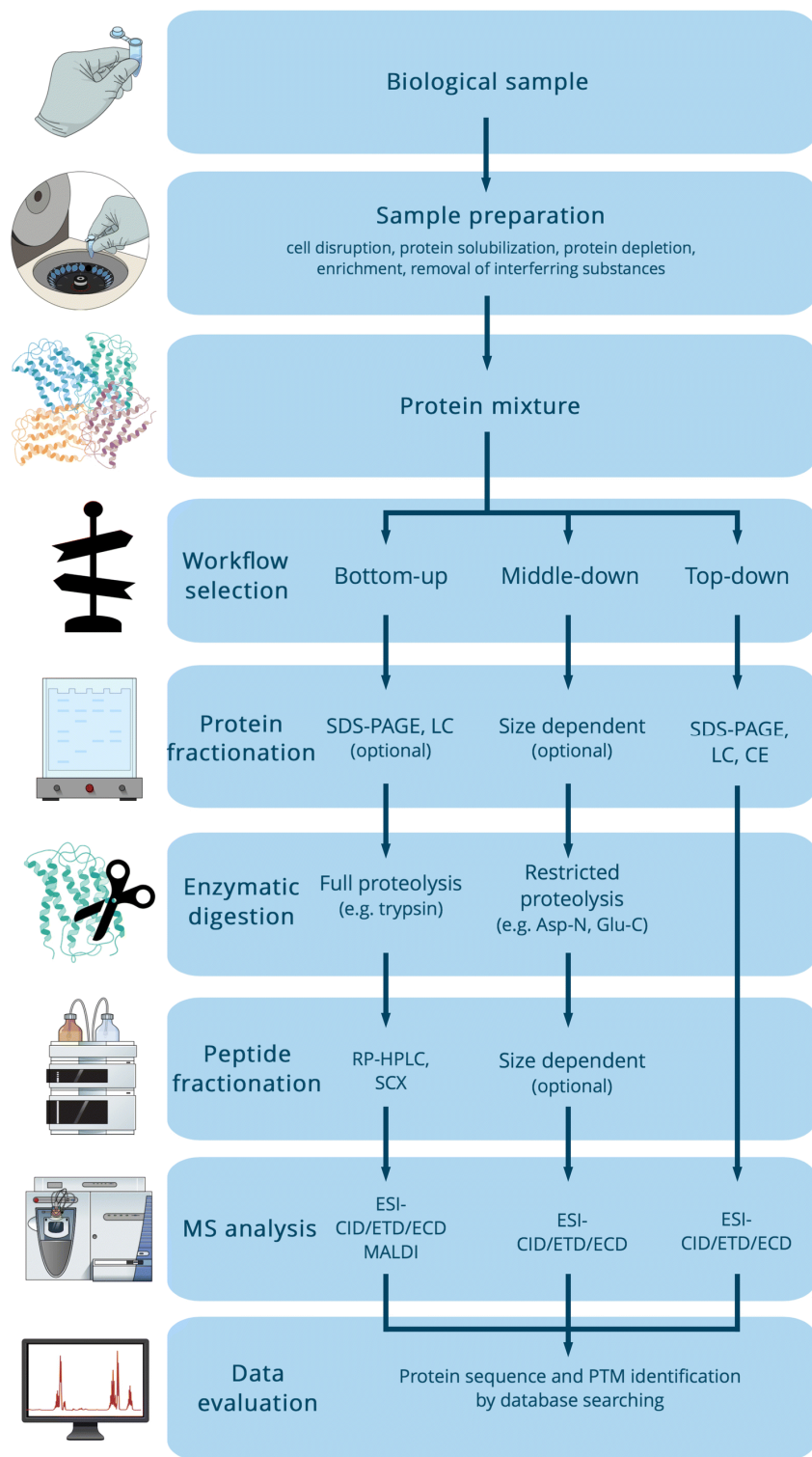


Figure 27: Résumé du processus et des approches utilisées dans l'analyse protéomique [98]

2.4. Lyse cellulaire et extraction de protéines :

Avant tout commencement d'une analyse protéomique, les protéines de l'échantillon biologique à analyser (un morceau de tissu animal, une prise de sang, un flacon de bactéries, une feuille ou une plaque de cellules cultivées...) sont extraites avec soin, dans l'objectif de solubiliser le plus grand nombre possible de protéines et d'éliminer le maximum possible des impuretés et contaminants à savoir les acides nucléiques, les lipides les métabolites non-protéiques etc.... Cette étape est cruciale et impose un contrôle strict afin de produire un échantillon de qualité et d'éviter les dégradations protéiques qui peuvent rendre l'analyse spectrométrique (qualitative ou quantitative) qui en suit inutile. Afin d'obtenir des protéines à partir d'un spécimen biologique, l'échantillon doit d'abord être prélevé sur l'organisme ou sur une culture cellulaire.

L'extraction de protéines commence souvent par une lyse ou désintégration des cellules, suivis d'une purification et d'un enrichissement spécifique et/ou de l'isolement d'une protéine d'intérêt particulier. La lyse cellulaire, est le processus de rupture de la paroi et/ou de la membrane cellulaire pour libérer le contenu intracellulaire contenant des molécules ou des particules d'intérêt, qui sont les protéines pour notre cas.

En outre, de nombreuses techniques ont été développées pour obtenir le rendement protéique le plus élevé possible pour différents types d'échantillons, organismes, ou fractions subcellulaires. Les équipements de lyse mécanique, tels que les homogénéisateurs, les batteurs de billes (*bead beaters*) et les sonicateurs, constituent des techniques « violentes » couramment utilisées pour désintégrer les tissus ou les cellules dont l'extraction des protéines est difficile. En revanche, les méthodes chimiques utilisent des dénaturants ou des détergents pour lyser les cellules et solubiliser les protéines. Du fait que la lyse cellulaire libère également des protéases et d'autres enzymes cataboliques, les « cocktails » d'inhibiteurs de protéases (EDTA, aprotinine, pepstatine A) [99] et de phosphatases (Cantharidine, calyculine A, orthovanadate de sodium) à large spectre sont généralement employés afin d'empêcher la protéolyse non spécifique et la perte de phosphorylation des protéines.

2.4.1 Les techniques mécaniques de lyse cellulaire :

- **Lyse cellulaire par battage de billes (*bead-beating*) :**

La lyse cellulaire par battage de bille a été employée pour l'extraction de l'ADN et de l'ARN, et son application en protéomique s'est avérée également utile. Dans cette technique, les billes de verre ou de céramique sont placées dans un compartiment cylindrique où le mouvement violent d'agitation à grande vitesse des billes engendré par un rotor endommage les structures cellulaires (membranes ou parois) et intercellulaires par impaction de l'échantillon contre les surfaces solides des billes [100].

- **Lyse cellulaire par sonication :**

La sonication est utilisée pour désintégrer les membranes cellulaires et libérer le contenu des cellules via un processus appelé « sonoporation ». La sonication est effectuée lors de la préparation d'extraits de protéines afin de briser la cellule. Bien que le tampon de lyse puisse être utilisé, la sonication peut aider à séparer la cellule.

2.4.2 Les techniques chimiques de lyse cellulaire :

Lorsque les méthodes mécaniques de lyse cellulaire peuvent détériorer les protéines significativement, les techniques chimiques ou « *soft* » qui utilisent les solutions de lyse cellulaire peuvent pallier ce problème. Les solutions de lyse cellulaire sont principalement des tampons et des réactifs à base de détergents optimisés pour des applications particulières de lyse cellulaire. L'extraction efficace de protéines pour différentes espèces d'organismes et différents types de cellules et de tissus nécessitent différentes formulations de tampons adaptées et préalablement testées.

Les détergents utilisés sont des molécules amphiphiles, contenant à la fois des régions hydrophiles et hydrophobes, qui permettent de rompre les associations lipide-lipide, protéine-protéine, protéine-lipide et de dénaturer les protéines et autres macromolécules. Ces détergents sont classifiés en détergents ioniques, non-ioniques, zwitterioniques et chaotropiques (Tableau 5). Les molécules de chaque classe possèdent des propriétés physicochimiques particulières.

Plusieurs tampons ont été formulés pour l'extraction de protéines suivant le type cellulaire, la nature de l'échantillon et de l'organisme. Le tampon RIPA (*radioimmunoprecipitation assay*) est un choix convenable et « *gold standard* » pour la lyse et l'extraction de protéines de cellules ou de tissus mous de mammifères. Les protéines extraites par cette méthode chimique peuvent être utilisées pour une variété d'applications et d'expérimentations en aval telles que le Western blot, l'immunoprécipitation, l'ELISA, l'électrophorèse SDS-PAGE (électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium) et les purifications de protéines par affinité. Le tampon RIPA possède de fortes capacités de dénaturation particulièrement utiles pour la rupture des structures membranaires, et il se compose de 50 mM de Tris-HCl (pH 7,4), 150 mM de NaCl, 1% de Triton X-100, 0,5% de désoxycholate de sodium, 0,1% de SDS et 5 mM d'EDTA [101]. D'autres formulations pour le tampon RIPA existent sur le marché.

Type de détergent	Molécules
Ionique	Dodécylsulfate de sodium (SDS), désoxycholate, cholate, sarkosyl
Non-ionique	Triton X-100, DDM, digitonine, tween 20, tween 80
Zwitterionique	CHAPS
Chaotropique	Urée

Tableau 5: Classification des détergents

Il est à noter que le choix de détergents doit prendre en compte la possibilité de débarrasser l'échantillon à analyser des détergents. Une fois les protéines extraites, il est nécessaire d'éliminer ces contaminants car les détergents interfèrent avec la digestion enzymatique, les séparations RPLC et la spectrométrie de masse, endommageant parfois les instruments et irréversiblement les colonnes.

2.5. Décomplexification et enrichissement des protéines :

Le but ultime de techniques d'enrichissement des protéines et la déplétion des protéines majoritaires et la concentration de protéines d'intérêt qui sont souvent minoritaires. La teneur en protéines du sérum est toutefois dominée par une poignée de protéines telles que l'albumine, la transferrine, l'haptoglobuline, les immunoglobulines et les lipoprotéines. Malheureusement, les protéines sériques sont présentes dans une extraordinaire plage dynamique de concentration qui s'étend probablement sur plus de 10 ordres de grandeur. Cette large gamme dynamique dépasse les capacités analytiques des méthodes protéomiques traditionnelles, ce qui rend la détection des protéines sériques de moindre abondance extrêmement difficile. La réduction de la complexité de l'échantillon (par exemple pour diminuer le niveau des protéines abondantes) est donc une première étape essentielle dans l'analyse du protéome sérique.

Différentes stratégies, telles que la précipitation de l'albumine, l'exclusion par la taille ou l'immunodéplétion, ont été développées pour éliminer certaines des protéines les plus abondantes du sérum ou du plasma sanguin afin de faciliter l'analyse des protéines les moins abondantes.

2.6. Purification des protéines :

Après l'étape d'enrichissement, la purification de la protéine ou des protéines d'intérêt peut être nécessaire. La purification des protéines est réalisée par des méthodes électrophorétiques ou non électrophorétiques. Parmi ces dernières, on trouve les méthodes chromatographiques et non chromatographiques (Affinity Precipitation, Three Phase Partitioning [102] et Macroaffinity Ligand-Facilitated Three-Phase Partitioning).

2.7. Séparation de protéines ou peptides :

La séparation des échantillons complexes de protéines est achevée par trois méthodes qui sont les plus employées en protéomique : l'électrophorèse sur gel de polyacrylamide dénaturant (PAGE) ou l'électrophorèse sur gel de polyacrylamide au dodécylsulfate de sodium (SDS-PAGE), l'électrophorèse sur gel bidimensionnelle et la chromatographie liquide

à haute performance (HPLC). Néanmoins, d'autres outils additionnels peuvent être employés à savoir la chromatographie d'affinité et la chromatographie à échange d'ions.

Les spectromètres de masse ont des limites de performance inhérentes. L'une de ces limitations concerne la suppression des ions, où les analytes à faible abondance ou avec une faible efficacité d'ionisation seront détectés avec une sensibilité fortement compromise lorsqu'ils sont analysés simultanément avec des analytes à plus forte abondance ou avec une meilleure efficacité d'ionisation.

De plus, dans un échantillon protéomique complexe, les peptides isobares ou quasi-isobares ne peuvent être résolus avec précision, ce qui interfère avec l'identification des protéines. Pour atténuer ces limitations, des techniques de séparation appropriées sont nécessaires avant l'analyse par MS afin de réduire la complexité de l'échantillon et d'améliorer l'ionisation pour une identification précise des protéines.

3. L'électrophorèse bidimensionnelle (2DE) :

Deux types d'électrophorèse sont couramment utilisés dans l'analyse protéomique : la focalisation isoélectrique (FIE) et la SDS-PAGE. La focalisation isoélectrique utilise un gradient de pH formé dans le gel pour séparer les protéines en fonction de leur pI. La SDS-PAGE sépare les protéines en fonction de leur poids moléculaire lorsqu'elles migrent à travers une matrice de gel de densité définie.

L'électrophorèse bidimensionnelle sur gel (2DE) est une technique à base de gel largement utilisée pour analyser la composition protéique des échantillons biologiques. Elle est capable de résoudre des mélanges complexes contenant plus de mille composants protéiques en taches protéiques individuelles par le couplage de deux techniques de séparation orthogonales : la focalisation isoélectrique (première dimension) et l'électrophorèse SDS-PAGE (deuxième dimension). Dans la première dimension, les protéines sont séparées par pI en utilisant une bande à gradient de pH. Dans la deuxième dimension, les protéines séparées par la focalisation isoélectrique sont séparées par poids moléculaire [103].

Les protéines sont généralement visualisées en colorant le gel avec des réactifs non spécifiques, visibles, fluorescents ou absorbant les UV. Le bleu de Coomassie peut être décoloré de manière réversible et est compatible avec la spectrométrie de masse, alors que la coloration à l'argent n'est pas compatible en raison de l'utilisation de formaldéhyde ou de glutaraldéhyde pendant l'étape de fixation et de sensibilisation qui entraîne une réticulation des résidus de lysine dans la chaîne protéique interférant avec l'analyse par MS et entravant ainsi la digestion de la trypsine [104].

De plus, les zones d'intérêt spécifiques du gel peuvent être excisées pour identification par digestion par la trypsine dans le gel, suivie d'une analyse par spectrométrie de masse. Les avantages de l'électrophorèse bidimensionnelle comprennent la capacité de résoudre des milliers de protéines dans un seul gel, et a été fréquemment appliquée dans les études de protéomique pour analyser les lysats cellulaires.

Malgré l'ampleur de l'application du 2-DE, la technique est extrêmement laborieuse, longue et plus sensible aux erreurs de reproductibilité technique. De plus, la technique du 2-DE ne parvient pas à résoudre les protéines peu abondantes et hydrophobes, à savoir les protéines membranaires [105].

Une percée dans l'électrophorèse bidimensionnelle sur gel est arrivée avec l'introduction de la capacité de multiplexer des protéines marquées par fluorescence sur le même gel. Cette technique est appelée l'électrophorèse différentielle bidimensionnelle sur gel (2D-DIGE). 2D-DIGE est une forme modifiée de 2DE qui permet de comparer jusqu'à trois échantillons de protéines simultanément sur le même gel. Les protéines de chaque échantillon sont marquées de manière covalente avec des colorants fluorescents de couleur différente. Les protéines qui sont communes aux échantillons apparaissent comme des taches avec un rapport fixe de signaux fluorescents, alors que les protéines qui diffèrent entre les échantillons ont des rapports de fluorescence différents. L'aspect important de cette technologie est l'élimination de la variabilité entre les gels grâce au multiplexage.

Les applications de l'électrophorèse bidimensionnelle comprennent l'analyse de l'expression des protéines, les études de modifications post-traductionnelles [106] et l'analyse des biomarqueurs.

4. La chromatographie liquide à haute performance (HPLC) :

Dans toute analyse protéomique, la tâche la plus importante et parfois la plus difficile est la séparation d'un mélange complexe de protéines ou de peptides. Pour séparer un mélange pareil, des techniques utilisant l'HPLC peuvent être employées.

Dans la chromatographie, les espèces moléculaires soumises à la séparation existent dans un échantillon qui est constitué d'analytes et de matrice. Les analytes étant les espèces moléculaires d'intérêt, et la matrice est le reste des composants de l'échantillon. L'échantillon est introduit dans une phase mobile en écoulement qui passe par une phase stationnaire. La phase stationnaire retient les différentes espèces moléculaires qui passent, plus ou moins fortement, et les libère séparément dans le temps, pour les renvoyer dans la phase mobile.

En protéomique, la chromatographie liquide en phase inversée ou RP-HPLC (où la phase stationnaire est plus apolaire que l'éluant) est la plus utilisée, et les colonnes sont souvent remplies ou monolithiques [107]. L'éluant utilisé dans la RP-HPLC est généralement composé d'un mélange d'eau et de solvants organiques miscibles, généralement de l'acétonitrile, du méthanol ou du tétrahydrofuranne.

Il existe quelques techniques de chromatographie liquide qui peuvent être utilisées selon l'objectif et la nature des composants peptidiques et/ou protéiques contenus dans l'échantillon à analyser. Par exemple, la chromatographie d'affinité sur ions métalliques immobilisés (IMAC) est une technique de HPLC utilisée pour l'enrichissement des phosphoprotéines, car elle utilise une résine de chromatographie à charge positive (Fe^{3+} , Ti^{4+} , Ga^{3+} , ...) qui se lie aux fragments de phosphoprotéines à charge négative. De plus, la chromatographie liquide à interaction hydrophile (HILIC) est utilisée pour la séparation des peptides hydrophiles et utilise des solvants miscibles à l'eau, contrairement à la chromatographie en phase normale. Le couplage de ces techniques dans des flux de travail de la chromatographie liquide multidimensionnelle en ligne ou hors ligne a augmenté la profondeur de l'analyse LC-MS des échantillons de protéines complexes.

À la suite du développement de la protéomique, l'UPLC commence à remplacer la HPLC classique. Contrairement à la HPLC conventionnelle, la UPLC utilise une colonne composée de particules fines (< 2 micromètres), ce qui réduit la consommation de solvant et permet de gagner du temps. Elle offre également une résolution et une sensibilité plus élevées, ce qui augmente l'efficacité de la séparation [108]. Le développement de l'UPLC était basé sur la taille des particules, qui est un facteur impliqué dans l'équation de Van Deemeter, qui décrit la relation entre le débit et le HETP (efficacité de la colonne).

- **La chromatographie liquide multidimensionnelle :**

La chromatographie liquide multidimensionnelle combine deux ou plusieurs colonnes chromatographiques pour augmenter le pouvoir de résolution, des séparations afin de mieux fractionner les peptides avant leur entrée dans le spectromètre de masse. La suppression de l'ionisation peut se produire si trop de peptides co-élutent ou s'il y a une grande différence dans l'abondance des peptides. Les séparations à haute résolution améliorent la détection des peptides présents en quantités minimales, ce qui entraîne une augmentation globale de la gamme dynamique. Les séparations multidimensionnelles sont donc un excellent choix pour les échantillons très complexes.

5. Analyse des protéines par spectrométrie de masse :

Au cours des deux dernières décennies, le vaste champ d'application de la protéomique commence à être réalisé, en grande partie grâce aux développements technologiques de la spectrométrie de masse.

La spectrométrie de masse est une technologie analytique puissante qui peut être utilisée pour identifier des composés organiques (protéines, enzymes, peptides...) ou inorganiques inconnus, déterminer la structure de molécules complexes ou quantifier des concentrations extrêmement faibles d'analytes connus. Elle est actuellement l'une des meilleures méthodes et les plus efficaces pour étudier et localiser les PTM.

Pour une analyse basée sur la spectrométrie de masse, les molécules doivent être ionisées, ou chargées électriquement, pour produire des ions individuels. Ainsi, l'analyse par spectrométrie de masse exige que l'atome ou la molécule d'intérêt ait la capacité d'être ionisé et d'être présent dans la phase gazeuse. Les instruments d'analyse par spectrométrie de masse analysent les molécules en établissant un rapport entre la masse de chaque molécule et la charge ; cette caractéristique d'identification est spécifique à chaque molécule et est appelée rapport masse/charge (m/z). Par conséquent, si la molécule a une charge unique ($z = 1$), le rapport m/z sera égal à la masse moléculaire (Figure 28).

Bien qu'il existe de nombreuses configurations d'instruments disponibles, le fonctionnement du système MS peut être organisé généralement en trois étapes principales :

(1) Génération d'ions à partir de molécules

(2) Séparation des ions en fonction de leur masse et de leur charge dans un analyseur de masse

(3) Détection des ions et sortie de l'instrument.

Selon le type d'ionisation utilisé, ces étapes se déroulent entièrement ou partiellement sous vide pour faire avancer le mouvement des ions dans l'instrument.

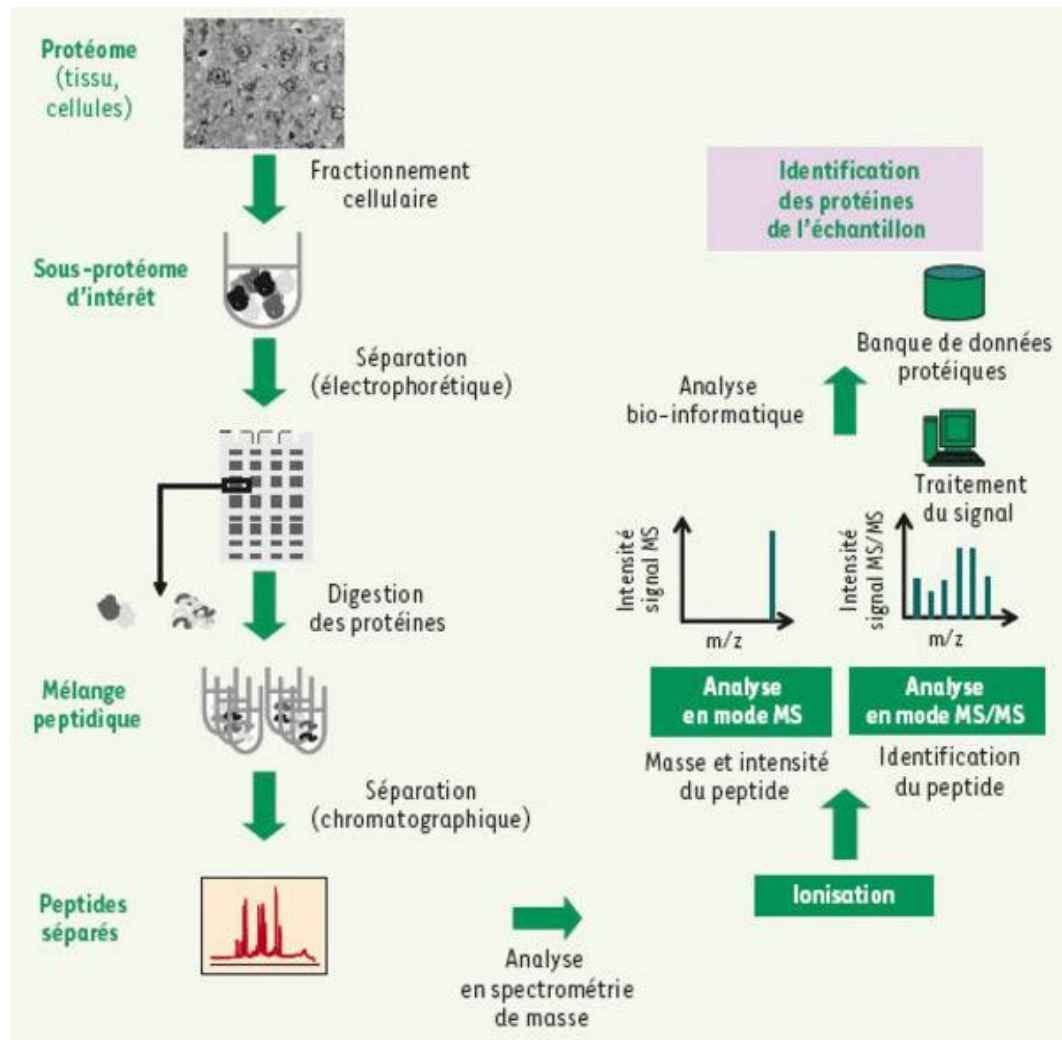


Figure 28: Déroulement d'une analyse protéomique basée sur la spectrométrie de masse (MS et MS/MS) [109].

5.1. Les sources d'ionisation :

Avant l'analyse des molécules d'un échantillon, celles-ci doivent être volatilisées et ionisées, et c'est les instruments d'ionisation qui assurent ces deux opérations. En protéomique, deux sources d'ionisation sont les plus fréquemment employées : MALDI (*matrix-assisted laser desorption-ionization*) et ESI (*electrospray ionization*).

5.1.1. Ionisation par électronébulisation :

La spectrométrie de masse utilisant l'ionisation par électronébuliseur (*electrospray ionization* ou ESI) fut publiée pour la première fois par John B. Fenn en 1988, ce qui lui a permis de partager le prix Nobel de chimie en 2002 avec deux autres chercheurs.

ESI est une méthode par laquelle un flux de liquide contenant les molécules d'intérêt est introduit à pression atmosphérique dans la source d'ions d'un spectromètre de masse par l'intermédiaire d'un capillaire recouvert de métal. Par l'application de tensions électriques appropriées (2-6 Kv) sur le capillaire (la tension peut être négative ou positive selon la nature de l'analyte), la contre-électrode et l'optique de focalisation, un aérosol de micro-gouttelettes fortement chargées est généré dans la source, qui traverse une série de « skimmers », rencontrant un gaz séchant (nitrogène) afin d'améliorer l'évaporation et l'élimination efficace de solvants de l'analyte chargé, dont l'effet net est la création d'ions gazeux désolvatés dont la distribution de charge est proportionnelle au nombre de groupes ionisables dans la molécule.

La taille initiale des gouttelettes peut varier de quelques micromètres à 60 μm de diamètre, en fonction du diamètre de la buse de nébulisation et du débit du liquide. La taille des gouttelettes diminue en raison de l'évaporation des molécules de solvant et, à mesure que le diamètre des gouttelettes diminue, la densité de charge augmente et les forces répulsives favorisent la désintégration électrohydrodynamique en de nombreuses gouttelettes plus petites (Figure 29).

En outre, les analyseurs de masse les plus utilisés pour l'ESI sont les quadrapôles, les analyseurs TOF et les pièges à ions. ESI-MS a été utilisé pour la détection, l'identification et la classification des microbes via l'étude des glucides, des lipides, de l'ADN et ARN, des protéines et des peptides.

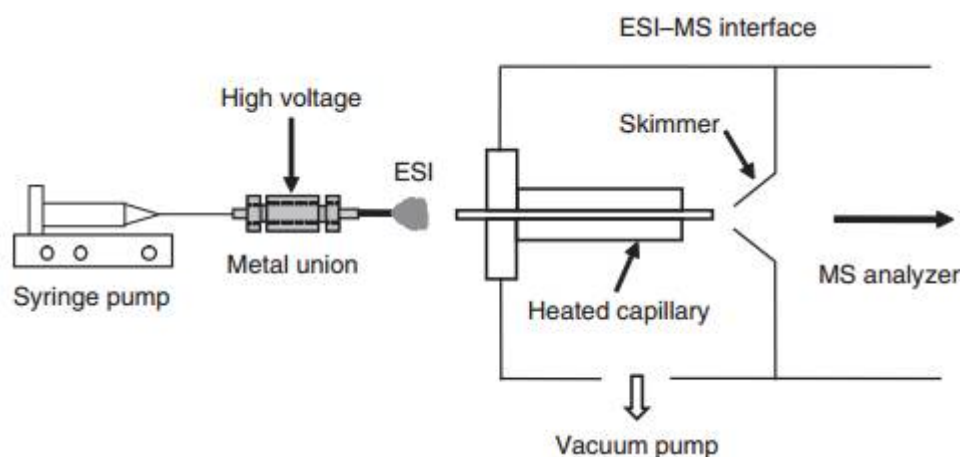


Figure 29: Schéma d'une interface ESI-MS standard [110]

5.1.2. Désorption-ionisation laser assistée par matrice (MALDI) :

La désorption/ionisation laser assistée par matrice est l'une des méthodes d'ionisation douce les plus importantes pour la spectrométrie de masse biologique. Elle est devenue l'une des approches les mieux établies en protéomique en raison de sa grande sensibilité et de sa faible consommation d'échantillons.

Dans cette technique, l'échantillon à analyser est préparé en le mélangeant ou en mettant un revêtement d'une solution d'un composé organique absorbant l'énergie appelée matrice. Lorsque cette matrice se cristallise par séchage, l'échantillon piégé dans la matrice se cristallise également. L'échantillon dans la matrice est ionisé automatiquement avec un rayon laser. La désorption et ionisation avec le rayon laser génère des ions monoprotonés à partir des analytes de l'échantillon (Figure 30). Les ions protonés sont ensuite accélérés, et se séparent les uns des autres sur la base de leur rapport masse/charge (m/z). Les analytes chargés sont alors détectés et mesurés à l'aide de différents types d'analyseurs de masse comme les analyseurs de pièges à ions, analyseurs de temps de vol (TOF) etc.

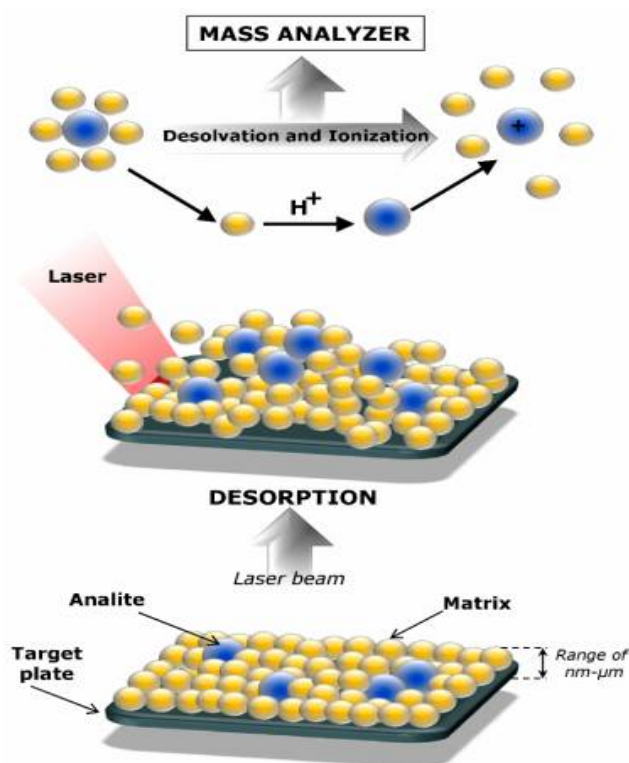


Figure 30: Processus d'ionisation et de désorption dans le MALDI [111]

La matrice idéale présente généralement les propriétés suivantes : forte absorption d'électrons à la longueur d'onde du laser utilisé ; meilleure stabilité dans le vide ; pression de vapeur plus faible ; faible masse moléculaire (pour faciliter la vaporisation) ; bonne miscibilité avec l'analyte à l'état solide.

L'ESI et le MALDI sont des techniques d'analyse très sensibles utilisant des concentrations d'analytes d'ordre de picomoles. Cependant, l'une des principales différences entre le MALDI et l'ESI est l'état dans lequel l'échantillon est introduit dans la source d'ions. L'ESI utilise un échantillon solvaté qui est infusé dans l'instrument, tandis que le MALDI utilise l'état solide. En outre, l'avantage du MALDI par rapport à la technique d'ionisation ESI est sa plus grande tolérance à la présence de détergents ou de sels dans les échantillons dans le cas de l'analyse de protéines pré-traitées et purifiées par des techniques électrophorétiques, par exemple l'électrophorèse bidimensionnelle [112].

5.1.3. SELDI :

Décrite à l'origine en 1993 (Hutchens et Yip 1993), le SELDI ou *surface-enhanced laser desorption/ionization* est variation de la technique d'ionisation MALDI. Dans le SELDI, les protéines d'intérêt dans un échantillon se lient sélectivement à une surface chimique avant l'analyse par spectrométrie de masse, contrairement à la technique MALDI, dans laquelle l'échantillon est mélangé avec la matrice et appliqué sur une plaque métallique.

La technique SELDI est employée en combinaison avec l'analyseur de masse TOF (SELDI-TOF-MS) et offre une faible consommation d'échantillon, une limite de détection femto-molaire, et une très grande sensibilité de détection, facilitant ainsi l'analyse des quantités infinitésimales d'échantillon. De plus, elle permet d'analyser les protéines de faible poids moléculaire (< 20 kDa).

Le SELDI implique la liaison des protéines et des peptides présents dans des échantillons biologiques complexes, tels que le sérum, les lysats cellulaires etc..., aux puces à protéines (ProteinChip® Arrays) (Figure 31). Les protéines non liées et les composés interférents de la spectrométrie de masse sont éliminés par lavage et les protéines qui sont retenues à la surface du réseau sont analysées et détectées par SELDI-TOF-MS.

Ces puces utilisent diverses surfaces de chromatographie ou surfaces biologiques pour capturer les protéines de mélanges biologiques complexes en fonction de leurs propriétés physico-chimiques. Les surfaces chromatographiques sont composées de substances chimiques hydrophobes, hydrophiles, d'échange d'ions, ou autres. Ces surfaces sont souvent utilisées pour le profilage de protéines provenant de mélanges biologiques, souvent pour la découverte de biomarqueurs. Les surfaces pré-activées sont utilisées pour immobiliser de manière covalente des molécules d'appât spécifiques telles que des anticorps, des récepteurs ou des peptides.

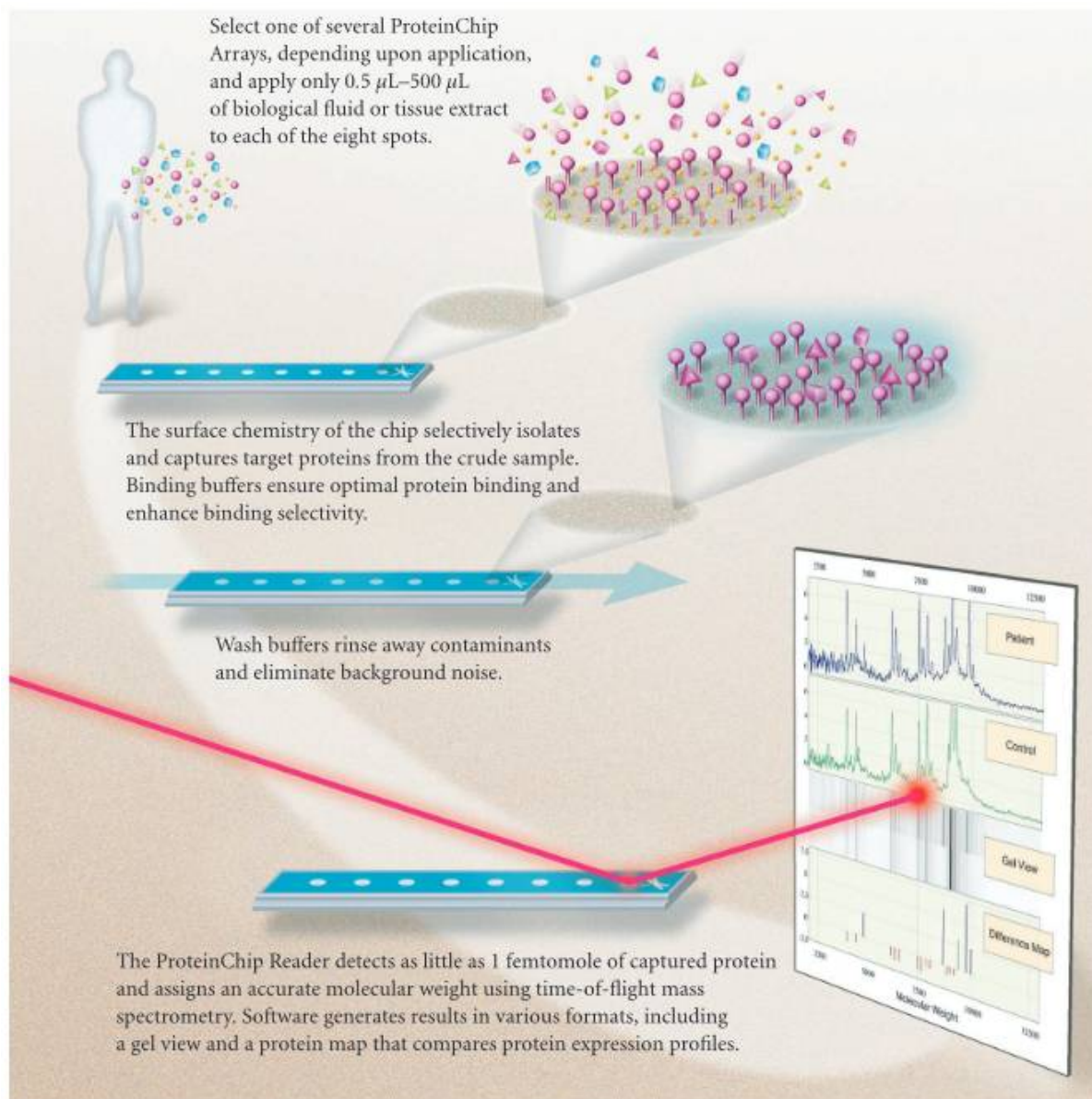


Figure 31: Protocole de préparation des puces ProteinChip avec des échantillons biologiques et d'analyse des protéines retenues par SELDI-TOF-MS [113].

5.1.4. Autres techniques d'ionisation :

A côté de MALDI et ESI, plusieurs autres techniques d'ionisation sont utilisées dans la spectrométrie de masse, à savoir l'ionisation électronique (EI), l'ionisation chimique (CI) et bombardement par atomes rapides (FAB). Néanmoins, ces techniques ne sont pas adaptées à l'échelle protéomique qui comprend de macromolécules ayant une masse moléculaire de plus de 5000 Da, dont l'ionisation n'est assurée convenablement que par le MALDI et l'ESI (Tableau 6).

De plus, la sensibilité était le principal problème avec ces techniques, car les quantités nécessaires pour effectuer une analyse étaient beaucoup plus élevées que celles nécessaires pour le séquençage d'Edman, ce qui rendait ces techniques d'ionisation inadaptées à l'identification des protéines en routine, au contraire de MALDI et ESI, qui ont permis pour la première fois l'acquisition de spectres de masse uniquement avec des quantités infinitésimales de peptides et de protéines [114].

Table 1. Summary of the various features of current ionization methods used in mass spectrometry

Ionization technique	Nature of analytes	Sample introduction	Mass range	Brief description
Electron impact (EI)	Volatile; thermally stable	Gas chromatography solid or liquid probe	<1000 Da	Hard method; mainly fragment ions
Chemical ionization (CI)	Volatile; thermally stable	Gas chromatography solid or liquid probe	<1000 Da	Soft method; Molecular ion
Fast atom bombardment FAB	Organometallic compounds	Liquid chromatography, direct injection	<5000 Da; optimal range 200–2000	Soft method; require matrix
Matrix-assisted laser desorption ionization (MALDI)	Biomolecules (proteins, DNA, glycoconjugates)	Sample is cocrystallized with a matrix	Can go beyond 500,000 Da	Very soft method singly charged ions
Electrospray ionization (ESI)	Organic and inorganic compounds	Sample in solution	From very low masses to extraordinarily high	Very soft method; multiply charged ions

Tableau 6: Résumé des différentes caractéristiques des méthodes d'ionisation actuelles utilisées en spectrométrie de masse [115]

5.2. Les analyseurs :

Un analyseur de masse est le composant du spectromètre de masse qui prend les masses ionisées et les sépare en fonction des rapports masse/charge (m/z) et les envoie au détecteur où elles sont détectées et ensuite converties en une sortie numérique.

5.2.1. L'analyseur à trappe orbitale :

Dans les années 90, les travaux d'Alexander A. Makarov et de ses collègues ont largement contribué au développement de cette technologie, et une première publication et description de cette technique a été faite en 2000 [116].

L'analyseur à trappe orbitale, appelé communément Orbitrap, fait partie des analyseurs de masse à piège d'ions qui utilisent une combinaison de champs électriques ou magnétiques pour capturer ou « piéger » les ions à l'intérieur de l'analyseur de masse. Il existe de multiples configurations de pièges à ions, notamment des pièges à ions 3D (le piège à ions de Paul), un piège à ions linéaire (piège 2D) et un piège électrostatique (Orbitrap), ou un piège à base de filaments magnétiques (analyseur à résonance cyclotronique ionique).

L'Orbitrap est une trappe d'ions de Kingdon améliorée qui utilise un champ purement électrostatique pour le piégeage des ions, et qui se compose de deux électrodes externes et d'une électrode centrale cylindrique en forme d'un fuseau, délimitant un espace dans lequel les ions sont piégés. Une tension continue est appliquée entre les électrodes externes et internes, produisant un potentiel logarithmique radial entre les deux électrodes. Ces électrodes sont façonnées de manière à produire la distribution quadri-logarithmique du potentiel :

$$U(r, z) = \frac{k}{2} \left(z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \cdot \ln \left[\frac{r}{R_m} \right] + C$$

r et z sont des coordonnées cylindriques ($z = 0$ étant le plan de symétrie du champ), C est une constante, k est la courbure de champ, et R_m est le rayon caractéristique [117].

Les ions accélérés sont injectés perpendiculairement à l'axe- z (Figure 32) et piégés autour de l'électrode centrale grâce au « pressage électrodynamique » (*electrodynamic squeezing*) qui contrebalance les forces centrifuges générées par l'électrode centrale afin d'éviter la perte d'ions vers les électrodes externes [118]. Les ions subissent deux types de mouvements : un mouvement circulaire autour de l'électrode centrale et un mouvement oscillatoire de va-et-vient parallèlement à l'axe z .

Comme les ions de rapports m/z différents oscillent à des fréquences différentes, ils peuvent être séparés et analysés. En mesurant les fréquences d'oscillation (mouvement oscillatoire) induites par les ions sur les électrodes externes, les spectres de masse des ions sont acquis grâce à la détection du courant image induit, qui est transformé par la suite en spectre de fréquence par la transformée de Fourier (Figure 32).

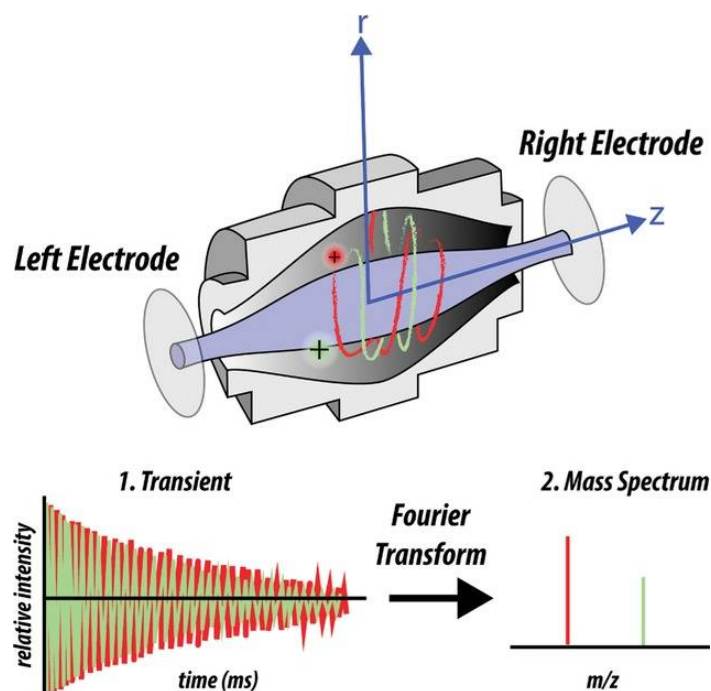


Figure 32: Représentation schématique du fonctionnement d'un analyseur de type Orbitrap [119]

L'Orbitrap est employé par des spectromètres de masse hybrides commerciaux (LTQ Orbitrap, Q Exactive, et Orbitrap Fusion) développés par Thermo Fisher Scientific®, qui ont pu atteindre de très grandes sensibilité, précisions (< 5 ppm) et résolutions ($> 150,000$) [120] qui est seulement dépassée par le FT-ICR-MS (Tableau 7).

5.2.2. L'analyseur à temps de vol :

L'un des analyseurs de masse les plus simples utilisés en spectrométrie de masse est l'analyseur de masse à temps de vol (TOF ou *time-of-flight*), souvent interfacé avec la source d'ionisation MALDI, formant une machine spectrométrique appelée MALDI-TOF. L'analyse d'un échantillon par TOF est également extrêmement rapide et offre une résolution et une sensibilité élevées, ce qui fait de cet instrument un choix populaire dans des techniques telles que la protéomique.

Après la production d'ions par le MALDI, une différence de potentiel fixe accélère tous les ions dans un tube de vol sans champ électrique (l'analyseur) où se produit une séparation par temps de vol. Comme tous les ions sont accélérés avec le même potentiel, ils ont tous la même énergie cinétique, mais traversent l'analyseur avec des vitesses différentes. L'analyseur

TOF fonctionne en mesurant le temps nécessaire (appelé temps de vol) aux ions générés dans la source pour traverser l'analyseur et atteindre le détecteur de l'autre côté. Ce temps est lié au rapport masse/charge (m/z) des ions, de façon que les ions ayant le rapport m/z le plus petit arrivent en premier au détecteur (Figure 34).

Le temps de vol d'un ion de masse m et d'énergie cinétique E pour parcourir une distance L dans une région exempte de champs électriques est donné par :

$$t = L \left(\frac{m}{2E} \right)^{1/2}$$

Un exemple de calcul pour un ion de poids moléculaire 200 Da ($3,32 \times 10^{-25}$ kg) avec une énergie cinétique de 10 keV ($1,60 \times 10^{-15}$ J), voyageant sur une distance de 1 m, donne un temps de vol de 10,18 μ s, et un ion de poids moléculaire 201 prend seulement 25 ns de plus [121].

Les analyseurs TOF fonctionnent en mode pulsé, de sorte que les ions doivent être produits ou extraits par impulsions. Pourtant, en pratique, l'impulsion n'est pas ressentie par tous les ions à la même intensité et donc tous les ions de même valeur m/z n'atteignent pas leur vitesse idéale. Pour remédier à ce problème, une réflexion est souvent appliquée à l'extrémité de la zone de vol à l'aide d'un réflectron (Figure 33) qui est constitué d'une série d'électrodes annulaires à haute tension, qui peuvent repousser et inverser la trajectoire des ions le long du tube de vol, généralement à un angle légèrement décalé.

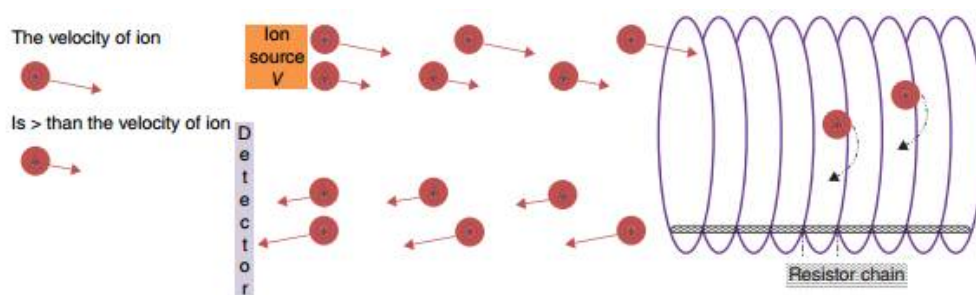


Figure 33: Principe de fonctionnement d'un réflectron dans un spectromètre de masse à temps de vol.

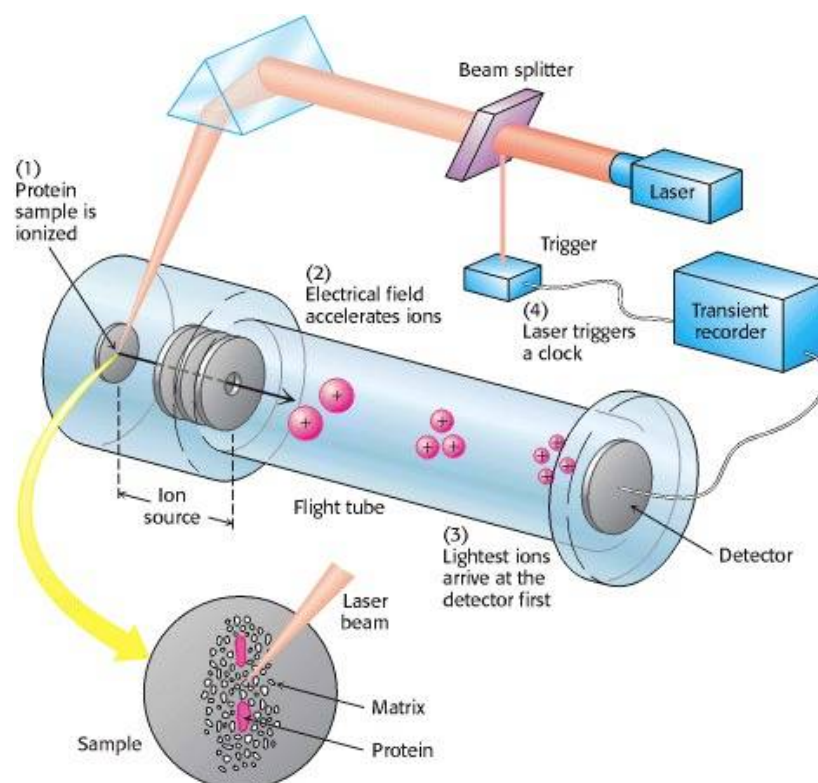


Figure 34: Schéma du spectromètre de masse MALDI-TOF [122]

5.2.3. L'analyseur quadripolaire :

L'analyseur quadripolaire (*quadrupole mass filter / quadrupole mass spectrometer*) a été développé par Wolfgang Paul et Helmut Steinwedel en 1953, et est utilisé couramment avec une source d'ionisation de type ESI.

Comme son nom l'indique, cet analyseur est composé de quatre tiges métalliques parallèles (de section circulaire ou hyperbolique), disposées autour d'un espace et un axe central. Les paires de tiges opposées sont reliées électriquement entre elles. Ce type d'analyseur de masse sépare les ions en fonction de la stabilité de leur trajectoire de vol grâce à un champ électrique oscillant dans le quadripôle (Figure 35).

Le champ électrique est généré lorsqu'une tension de radiofréquence (RF) est appliquée entre une paire de tiges opposées dans le quadripôle. Une tension continue de décalage est ensuite appliquée à l'autre paire de tiges opposées. Seuls les ions d'un certain rapport m/z auront une trajectoire stable dans le champ électrique résultant, tandis que tous les autres ions auront des trajectoires instables et vont heurter les barres métalliques et par conséquent ils n'atteindront pas le détecteur. Le comportement des ions lors de leur passage à travers l'analyseur quadripolaire est défini par l'équation de Mathieu [123].

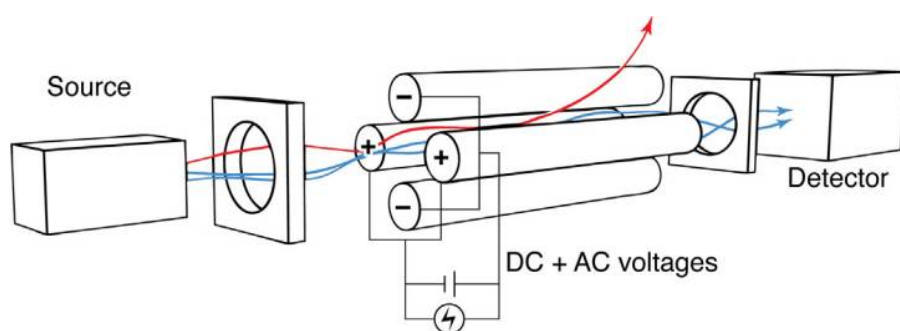


Figure 35: Représentation schématique d'un analyseur quadripolaire (Flèche en rouge : trajectoire instable ; flèche en bleu : trajectoire stable) [120]

Mass analyzer	Mass resolution ^a	Mass accuracy (ppm)	Sensitivity	m/z range	MS/MS capability	Ion source
IT	1000–1500	100–1000	Picomole	50–2000; 200–4000	MS^n	ESI
LTQ (LIT)	2000	100–500	Femtomole	50–2000; 200–4000	MS^n	ESI
Q-q-Q	1000	100–1500	Attomole	50–4000	MS/MS	ESI
TOF	10 000–20 000	5–50	Femtomole	No upper limit	n/a	MALDI
TOF-TOF	10 000–40 000	5–50	Femtomole	No upper limit	MS/MS	MALDI
Q-q-TOF	10 000–40 000	5–50	Attomole	No upper limit	MS/MS	ESI; MALDI
LTQ-FTICR	50 000–800 000	1–2	Femtomole	50–2000; 200–4000	MS^n	ESI; MALDI
LTQ-Orbitrap	50 000–500 000	<5	Femtomole	50–2000; 200–4000	MS^n	ESI; MALDI

^aMass resolution at full width half maximum.

Tableau 7: Comparaison des caractéristiques et des spécifications des analyseurs de masse couramment utilisés en protéomique [124]

5.3. Les détecteurs :

Après leur ionisation et leur filtration ou séparation par l'analyseur de masse, les molécules arrivent à leur destination finale : le détecteur. Le rôle de ce dernier est de traduire le courant ionique en un courant électrique, ce qui permet l'enregistrement du nombre d'ions pour chaque valeur m/z afin que l'ordinateur puisse digitaliser ces données et générer un spectre de masse. Les détecteurs sont également appelés « multiplicateurs » car ils multiplient (ou amplifient) le très faible courant produit par les ions arrivant à la surface du détecteur.

Dans les spectromètres de masse utilisés en protéomique, la détection des ions est souvent accomplie par un multiplicateur à galette de microcanaux (*microchannel plate* ou MCP), ou par des channeltrons (*channel electron multiplier* ou CEM) (Figure 36), à l'exception de l'Orbitrap et les spectromètres de masse utilisant la transformée de Fourier comme FT-ICR-MS (Cf. II.5.5.1).

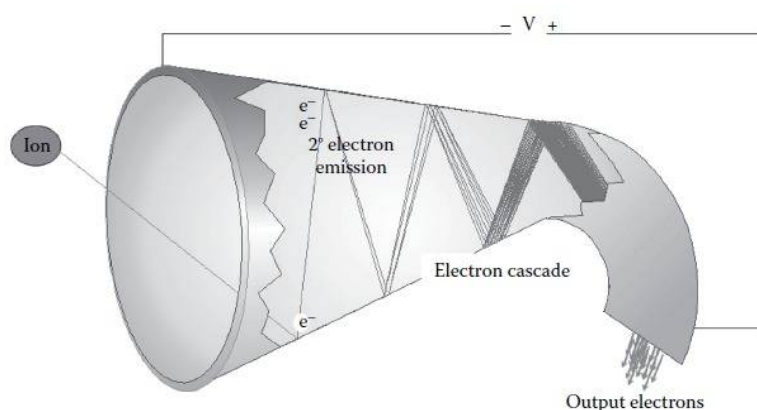


Figure 36: Représentation schématique d'un channeltron (une dynode continue) : l'amplification du courant se fait par une cascade d'émission secondaire d'électrons [125]

6. Techniques d'analyse d'interactions protéine-protéine (IPP) :

Pour remplir leur fonction dans une cellule vivante, les protéines n'agissent que rarement comme des espèces isolées. Plutôt, elles agissent en groupe via des liaisons avec d'autres protéines pour effectuer un réseau complexe de réactions importantes pour la régulation et l'exécution de la plupart des processus biologiques, à savoir la transduction de signal et la régulation des voies métaboliques. Cependant, pour révéler les mécanismes de fonctionnement dans les cellules, il est important d'identifier les IPP qui ont lieu dans l'organisme.

En utilisant une vaste collection de données sur les interactions protéiques, on a estimé que le nombre d'IPP chez l'homme est d'environ 650 000 [126]. Une majeure partie de l'interactome a déjà été cartographiée chez la levure, la plante, et l'homme, et les IPP issus d'expériences à grande et à petite échelle sont inscrits et stockés dans des bases de données spécialisées qui permettent d'avoir une vue d'ensemble de toutes les interactions possibles [127].

Grâce au concept de domaines protéiques qui a été introduit par Wetlaufer en 1973 [128], il est maintenant clair que non seulement les interactions protéine-protéine sont importantes pour la signalisation cellulaire, mais encore de nombreuses protéines contiennent des domaines modulaires reconnaissables ou des motifs qui confèrent une spécificité de liaison protéine-protéine. La cartographie des interactions physiques entre les protéines est une étape essentielle pour démêler les relations intermoléculaires complexes des réseaux protéiques et mieux comprendre la machinerie moléculaire qui sous-tend les phénotypes complexes [129].

La carte complète des interactions protéine-protéine qui peuvent se produire dans un organisme vivant font partie d'un réseau appelé « interactome ».

En outre, une meilleure compréhension des IPP et de leurs interfaces est essentielle pour le développement de nouvelles thérapies antivirales, comme la conception de petites molécules capables de se lier et de bloquer les interactions essentielles des processus viraux. Un exemple classique d'interactions virus-hôte bloquées au niveau de l'interface est l'action

du Maraviroc en tant qu'inhibiteur de l'entrée du VIH-1 dans les cellules hôtes. Ce médicament se lie au co-récepteur cellulaire CCR5, l'empêchant d'interagir avec GP120, une étape essentielle de l'infection par le VIH-1. Quant au nouveau coronavirus, une étude a identifié par AP-MS, 332 interactions à haut niveau de confiance entre les protéines de SARS-CoV-2 et les protéines humaines, qui sont liées à de multiples processus biologiques, notamment le trafic de protéines, la traduction, la transcription et la régulation de l'ubiquitination [130]. Ces interactions ont permis d'identifier 69 molécules approuvées par la FDA, en stade clinique ou pré-clinique, qui ciblent ces interactions et constituent des options thérapeutiques potentielles pour le COVID-19 [131].

D'autre part, le virus T-lymphotrope humain de type 1 (HTLV1) engendre un néoplasme de cellules T CD4+CD25+ connu sous le nom de leucémie/lymphome à cellules T de l'adulte (ATLL), dont la forme aiguë et lymphomateuses sont très agressives avec un pronostic sombre et une médiane de survie allant de 6 à 10 mois [132]. Ce néoplasme est en partie lié à la protéine Tax-1 qui est une protéine multifonctionnelle qui permet l'activation de la transcription de gènes viraux HTLV1 et cellulaires et d'agir comme médiateur de plusieurs voies cellulaires, et surtout celles régulant la prolifération et l'apoptose cellulaire. Les techniques d'analyse d'interactions protéine-protéines ont permis d'établir une carte interactomique pour mieux comprendre et avoir une vue globale sur les différentes interactions entre le Tax-1 et les protéines humaines (Figure 37).

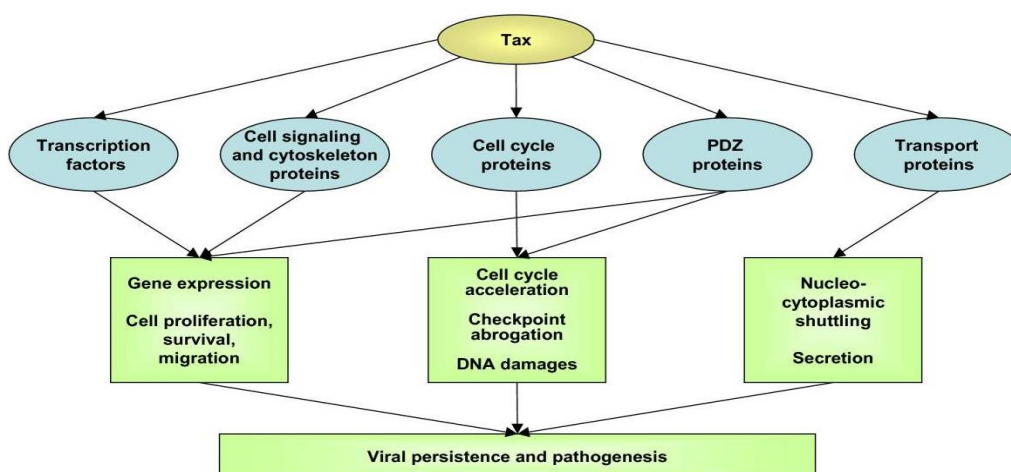


Figure 37: Aperçu des protéines de signalisation cellulaire ciblées par Tax1 [133]

6.1. Définition des interactions protéine-protéine :

Les interactions protéine-protéine sont généralement définies comme des contacts physiques par amarrage moléculaire (position et orientation des protéines spécifique) entre les protéines, qui se produisent *in vivo* dans une cellule ou dans un organisme vivant. La définition s'étend à certaines conditions supplémentaires : une interaction protéine-protéine doit être non aléatoire, excluant ainsi toutes les protéines qui peuvent se heurter les unes aux autres par hasard, et en outre, elle ne doit pas être générique, mais servir un objectif spécifique qui diffère des fonctions complètement génériques telles que la production, l'enroulement et la dégradation des protéines [134].

Les IPP sont conventionnellement classifiées en IPP permanentes et IPP transitoires, les premières dénotant une liaison étroite entre des protéines qui donnent lieu à des complexes forts et stables, tandis que les secondes se composent d'interactions relativement faibles qui peuvent se dissocier après la liaison pour réguler l'activité fonctionnelle à un moment précis.

Pour que les protéines puissent interagir entre elles, leurs sites de liaison respectifs doivent être en contact physique direct. Ces sites de liaison sont appelés « interfaces », et sont des structures tridimensionnelles formées par des ensembles de résidus d'acides aminés directement responsables de la reconnaissance des partenaires protéiques de liaison. Des mutations délétères ou bénéfiques peuvent se produire en particulier sur les interfaces, affectant ainsi l'affinité de liaison en raison d'une altération ou d'une amélioration des propriétés électrostatiques et structurelles des protéines.

Les approches de cartographie binaire interrogent des paires de protéines pour vérifier l'existence d'interactions directes entre eux (PPI). Les approches de cartographie de complexes protéiques, en revanche, visent à identifier les ensembles de protéines qui appartiennent à un complexe protéique multimérique, dans lequel des paires de protéines forment soit des contacts directs ou sont liés par des associations indirectes. Les méthodes actuelles d'identification des complexes protéiques utilisées à l'échelle du protéome ne permettent pas de distinguer les contacts directs et les associations indirectes pour toutes les paires de protéines signalées.

6.2. Méthodes conventionnelles d'analyse d'IPP :

Pour que les protéines puissent interagir entre elles, leurs sites de liaison respectifs doivent être en contact physique direct. Ces sites de liaison sont appelés « interfaces », et sont des structures tridimensionnelles formées par des ensembles de résidus d'acides aminés directement responsables de la reconnaissance des partenaires protéiques de liaison. Des mutations délétères ou bénéfiques peuvent se produire en particulier sur les interfaces, affectant ainsi l'affinité de liaison en raison d'une altération ou d'une amélioration des propriétés électrostatiques et structurelles des protéines.

6.2.1. Technique du double hybride de la levure « Y2H » :

La technique du double hybride de la levure (Yeast 2-hybrid system ou Y2H) est une technique *in vivo*, qui a été conçue en 1989 pour l'objectif d'identifier et d'analyser les interactions protéine-protéine en utilisant l'activateur transcriptionnel Gal4 de la levure *Saccharomyces cerevisiae*, une technique qui a révolutionné l'étude des IPP [135]. Ce système s'est révélé être un outil très utile et une méthode sensible pour détecter non seulement les protéines stables en interaction, mais aussi les interactions protéiques faibles et transitoires, et sa popularité s'explique par ses protocoles standardisés depuis longtemps, et aussi par sa facilité et son faible coût [136].

6.2.1.1. Principe de la technique :

Le système à double hybride de levure est basé sur la nature modulaire des activateurs transcriptionnels (facteurs de transcription) des eucaryotes qui consistent en deux domaines : un domaine de fixation à l'ADN (DNA-binding domain ou DNA-BD) et un second domaine (domaine d'activation ou AD) qui recrute l'appareil transcriptionnel assurant l'activation de la machinerie basale de transcription. Isolés, ces domaines ne peuvent en aucun cas être fonctionnels, et ne peuvent fonctionner que par le biais d'interactions protéine-protéine d'une manière non-covalente. Ainsi, une paire de protéines qui interagissent entre elles réunira un ADN-BD et un AD séparés pour reconstituer un transactivateur. Dans un essai typique de levure à deux hybrides, une protéine « appât », exprimée sous forme de fusion avec un ADN-BD, et une protéine « proie » fusionnée à un AD, sont co-exprimées dans les cellules de

levure. En cas d'interaction entre les deux protéines d'intérêt, le transactivateur reconstitué est recruté dans un ou plusieurs gènes rapporteurs contenant de l'ADN-BD, ce qui entraîne finalement l'induction et la transcription d'un gène rapporteur, généralement des marqueurs auxotrophes *HIS3*, *ADE2* et *MEL1*, qui à leur tour permettent aux colonies de levure de se développer sur des milieux sélectifs ou de changer de couleur sur des milieux colorimétriques (Figure 38). Le système souvent utilisé est celui de Gal4 [137].

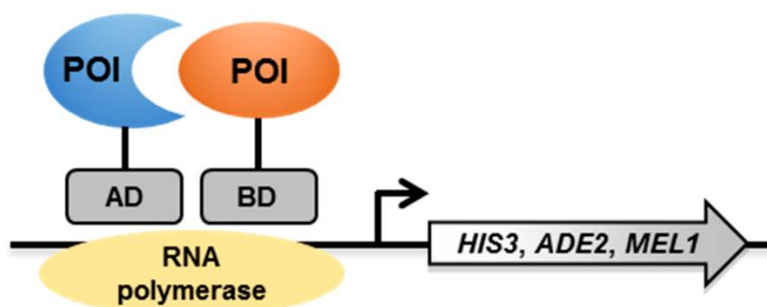


Figure 38: Représentation schématique du système Y2H. (POI : protéine d'intérêt ; AD : domaine d'activation ; BD : domaine liant l'ADN) [138]

6.2.1.2. Limites de la technique :

Bien que cette technique soit moins coûteuse et relativement rapide, on constate que les faux positifs et les faux négatifs sont élevés dans les systèmes à deux hybrides [139]. Un autre inconvénient est que certaines interactions protéiques dépendent de modifications post-traductionnelles qui peuvent ne pas se produire ou se produire de manière inappropriée dans la levure.

6.2.2. Technique AP-MS :

La technique de purification par affinité a été utilisée avec succès pour l'étude d'interactions protéine-protéine en combinaison avec la spectrométrie de masse. A titre d'exemple, cette technique combinée permet l'identification efficace et à haut débit de centaines de partenaires protéiques en interaction, ce qui a été utilisé dans plusieurs études de recherche pour cartographier le paysage global de la pluripotence [140].

Dans la purification par affinité, une protéine cible endogène ou une protéine cible à affinité exprimée ectopiquement est utilisée comme appât pour purifier son complexe protéique. L'analyse ultérieure par spectrométrie de masse du complexe protéique purifié permet d'identifier un interactome de la protéine cible.

L'une des méthodes les plus courantes utilisées pour les expériences AP-MS consiste à exprimer la protéine cible avec une étiquette de séquence peptidique unique tel que FLAG, c-Myc, HA ou V5 et à utiliser les anticorps de haute affinité bien documentés contre ces séquences peptidiques.

6.2.2.1. Purification par affinité en une seule étape :

En général, une séquence de marquage par affinité est génétiquement fusionnée à l'extrémité N ou C de la protéine cible, ce qui permet de purifier la protéine marquée ou étiquetée à partir d'extraits cellulaires ou de milieux de culture par liaison à haute affinité à une matrice de purification. Les deux petits marqueurs d'affinité peptidiques, tels que poly-His, poly-Arg, c-myc, Strep ou Flag, ainsi que les grands marqueurs peptidiques ou protéiques tels que le domaine de liaison à la cellulose ou à la chitine ou la protéine de liaison au maltose, ont été largement utilisés avec succès pour la purification des protéines en une seule étape.

6.2.2.2. Purification par affinité en tandem (TAP) :

La purification par affinité en tandem (TAP), développée par Rigaut G et al. (1999), est une approche générique pour la purification des complexes protéiques. Elle a d'abord été testée dans la levure et s'est rapidement révélée applicable à d'autres cellules ou organismes, et permet une analyse systématique des interactions protéine-protéine dans des conditions quasi-physiologiques.

En général, la purification par affinité en tandem (TAP) consiste en un enrichissement protéique en deux étapes utilisant une étiquette TAP constitué de deux marqueurs d'affinité, un peptide de liaison à la calmoduline (CBP) et une protéine A de *Staphylococcus aureus*, séparés par un site qui constitue la cible de la protéase du virus de la gravure du tabac (TEV), qui sont fusionnés à l'extrémité C ou N-terminale de la protéine appât.

Une fois exprimée dans les cellules, la protéine marquée par une étiquette TAP forme un complexe avec ses partenaires biologiques endogènes. En général, le complexe protéique contenant la cible est d'abord récupéré sur la matrice IgG (via le marqueur Protéine A). Après lavage, le matériel lié est libéré par clivage TEV pendant l'élution. La fraction éluée est ensuite incubée avec de la résine de calmoduline en présence de calcium. Enfin, la matière capturée est libérée par l'ajout d'agents chélatants. Cette deuxième étape de purification permet d'éliminer les contaminants restants après la première purification ainsi que la protéase TEV (Figures 39, 40).

Pourtant, il faut bien noter que la purification en deux étapes, qui implique deux lavages, non seulement réduit les contaminants mais peut également entraîner la perte de protéines à faible interaction et également engendrer des faux négatifs. En outre, d'autres étiquettes peuvent être utilisées à la place de CBP et la de la protéine A, à savoir le Strep-tag, HA-tag et le FLAG-tag [141].

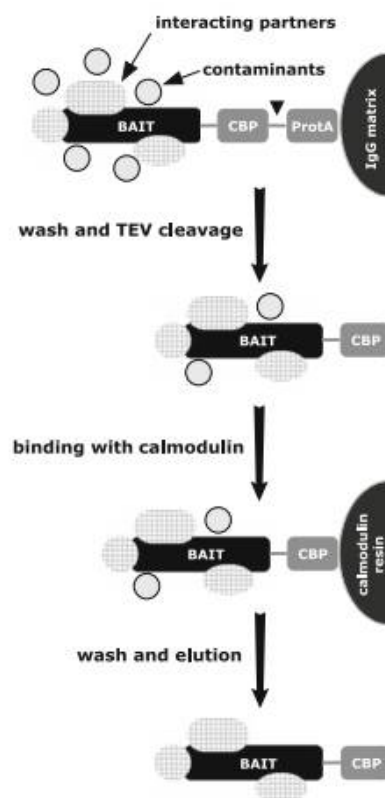


Figure 39: Aperçu schématique de la procédure TAP. (Bait : protéine appât) [142]

Bien que l'AP-MS soit une approche puissante, elle ne donne généralement pas de bons résultats sur les complexes protéiques peu solubles, les interactions très transitoires ou les IPP qui nécessitent une structure subcellulaire intacte, d'où l'utilité des nouvelles méthodes d'analyse d'IPP qui ont partiellement remédié à ces limitations.

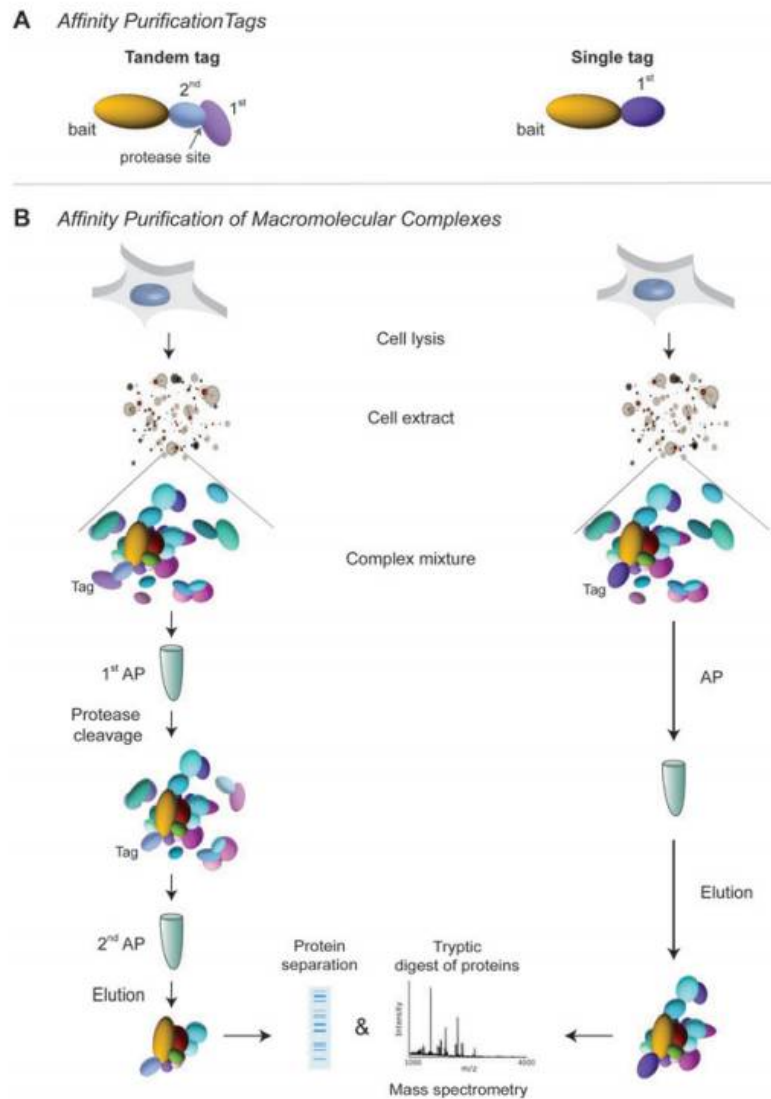


Figure 40: Comparaison entre la purification par affinité en tandem ou TAP (gauche) et en une seule étape (droite) [143].

6.3. Méthodes récentes d'analyse d'IPP :

Les méthodes récentes d'analyse des interactions protéine-protéine se basent largement sur le principe de marquage de proximité. Parmi ces méthodes, on cite la technique BioID (*proximity-dependent biotinylation identification*), une technique qui a été récemment introduite pour la première fois par Roux et ses collègues en 2012, pour le but de procurer à l'arsenal biotechnologique une nouvelle approche de criblage des interactions protéine-protéine *in vivo*, en employant une version mutante de l'enzyme BirA (dénotée BirA*), une biotine ligase extraite d'*Escherichia coli* pour l'identification de protéines voisines qui sont potentiellement en interaction. La biotine ligase est fusionnée à une protéine d'intérêt (protéine appât), puis introduite dans des cellules d'intérêt où elle biotinylera les protéines proximales après supplémentation du milieu de culture en biotine. Les protéines biotinylées peuvent ensuite être isolées de manière sélective et identifiées par des méthodes conventionnelles, notamment la spectrométrie de masse.

En présence d'ATP et de biotine, l'enzyme BirA* va générer l'espèce réactive biotinoyl-5'-AMP (biotine-AMP), la retenir dans son site actif, et faciliter son transfert covalent sur les amines primaires lysines exposées à la surface des protéines cibles de proximité (Figure 41).

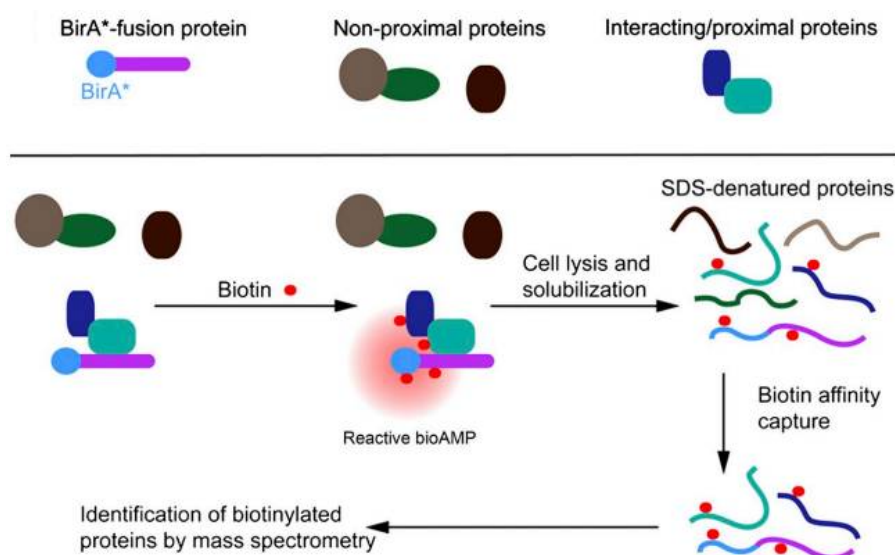


Figure 41: Modèle d'application de la technique BioID [144]

Par rapport à l'AP-MS, la BioID-MS offre plusieurs avantages clés à savoir la capture des interactions dans leur environnement cellulaire natif ; les protéines marquées plutôt que des complexes assemblés sont isolées à la suite de la lyse cellulaire. La méthode est donc plus sensible pour détecter les interactions transitoires ou faibles ou les interactions qui se produisent sur une structure subcellulaire spécifique et difficile à isoler.

D'autres versions mutantes de biotines ligases continuent d'être développées afin d'améliorer et d'accélérer la caractérisation des réseaux d'interactions protéiques. A titre d'exemple, une méthode appelée TurboID et sa variante miniTurboID, ont été récemment développées et utilisent deux biotines ligases mutantes de tailles de 35kd et 28kD respectivement. Cette nouvelle approche a achevé une biotinylation plus rapide (10 minutes) que la BioID (18 heures) [145].

6.3.1. Technique Split BioID :

Dans Split BioID, BirA* est d'abord divisé en deux fragments A (NBirA*) et B (CBirA*) non fonctionnels. Lorsque ces deux fragments A et B se mettent en proximité, ils forment une liaison entre eux pour former une enzyme BirA* fonctionnelle. Chaque fragment BirA* est fusionné à l'une des deux protéines appâts dont l'interaction est connue au préalable. Lorsque cette interaction se produit, l'activité de BirA* est restaurée et les protéines voisines sont marquées à la biotine (Figure 42).

Pour le marquage, les protéines de fusion BirA* sont exprimées dans les cellules via des plasmides ou par la création de lignées cellulaires stables et ces cellules sont ensuite cultivées dans des milieux contenant de la biotine [146]. Après leur isolement, l'identification des interacteurs pourra être réalisée par spectrométrie de masse.

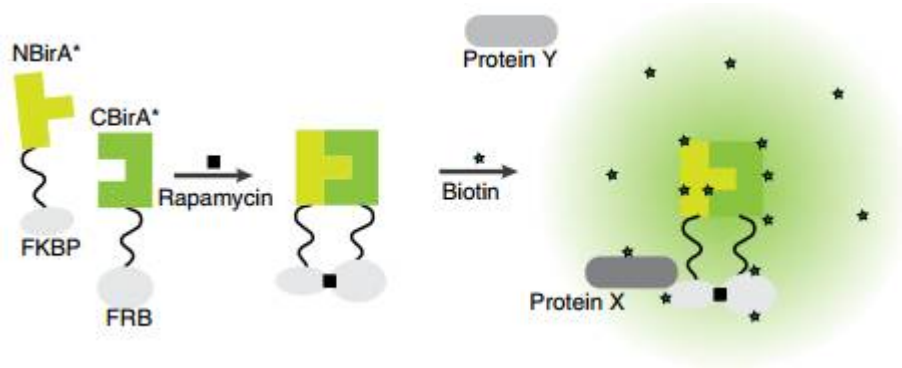


Figure 42: Représentation schématique de la technique Split BioID [147]

6.3.2. Technique BioID APEX et APEX2 :

Le marquage par l'ascorbate peroxydase ou APEX est une technique de marquage de proximité similaire à la technique BioID classique, et la différence est l'utilisation de l'ascorbate peroxydase au lieu d'une biotine ligase, et la biotine-phénol au lieu de la biotine.

Les peroxydases catalysent l'attaque de la tyrosine par un composé phénolique couplé à la biotine, lorsqu'elles sont activées par le peroxyde (H_2O_2). Une différence majeure entre la biotine ligase (BirA*) et les ascorbate peroxydases (APEX, APEX2) de BioID est le temps nécessaire à la génération d'un signal de biotinylation suffisant pour être utile aux approches d'identification basées sur la spectrométrie de masse. La BirA* nécessite généralement plusieurs heures de marquage, jusqu'à 24h dans certains protocoles. En revanche, le marquage APEX2 ne nécessite que quelques minutes de marquage, ce qui a permis des études sensibles au temps, comme la caractérisation de la signalisation des récepteurs couplés aux protéines G.

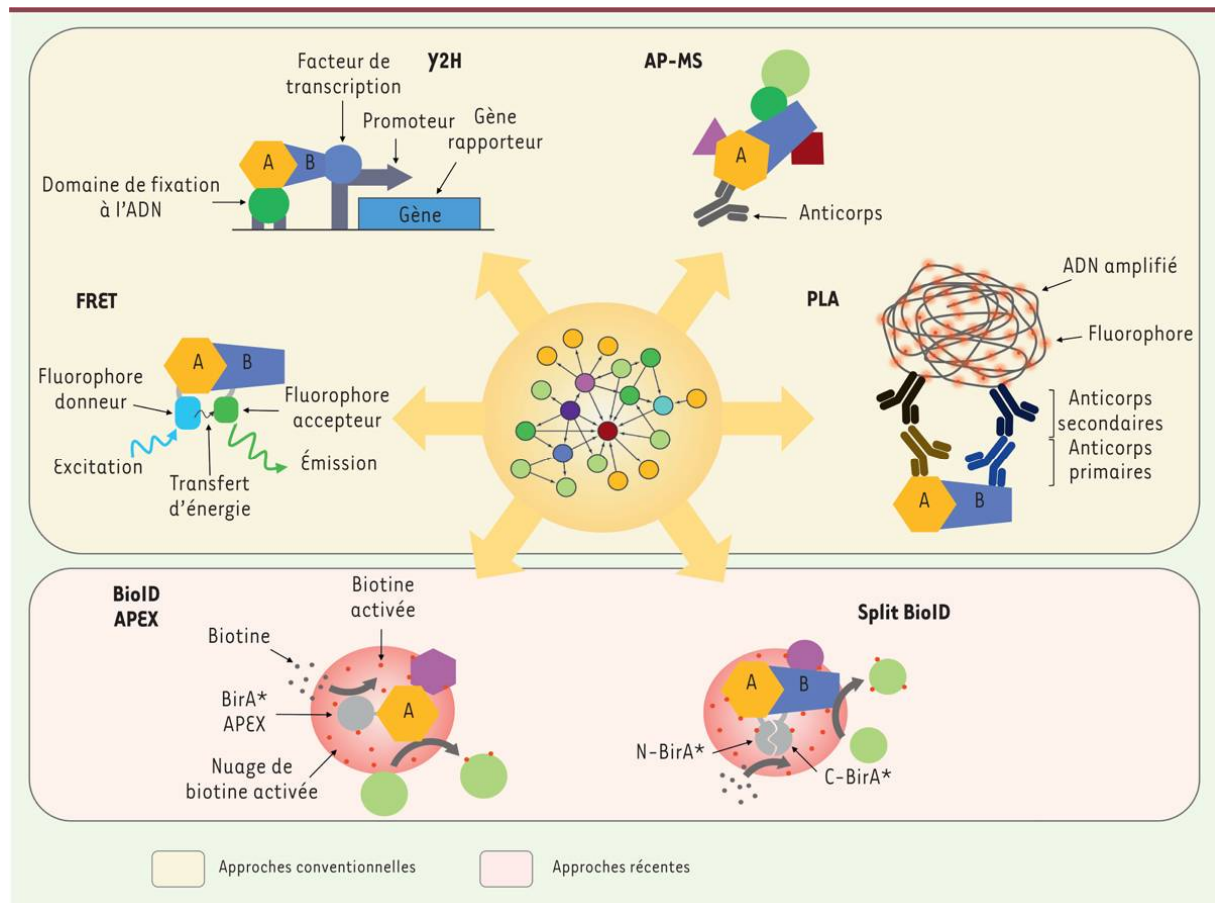


Figure 43: Vue d'ensemble des techniques d'identification des interactions protéine-protéine. [148]

7. La protéomique quantitative :

Il est passionnant de voir à quel point la protéomique a contribué à améliorer les analyses biomédicales et notre compréhension des processus cellulaires et subcellulaires. La protéomique quantitative apporte un grand avantage grâce aux nouvelles technologies qui permettent la quantification de protéines. La protéomique basée sur la spectrométrie de masse peut révéler l'état quantitatif d'un protéome et ainsi fournir des informations sur l'état biochimique de la cellule ou du tissu concerné, soit par l'emploi de stratégies de marquage (SILAC, iTRAQ, ICAT...) soit des stratégies sans marquage, directement à partir des données spectrales.

La protéomique quantitative basée sur la spectrométrie de masse est devenue un outil puissant dans la recherche biologique. Comme la spectrométrie de masse est une méthode analytique non quantitative, les peptides protéolytiques ont des réponses différentes à la spectrométrie de masse en raison de leurs diverses propriétés physico-chimiques telles que la taille, l'hydrophobicité, la charge, etc... Ainsi, il est impossible de quantifier les changements d'abondance relative entre différents échantillons expérimentaux dans la même analyse par spectrométrie de masse en raison d'un certain nombre de facteurs, notamment la réponse variable du détecteur et les rendements d'ionisation différentiels pour différentes substances. Par conséquent, quelle que soit la méthode d'ionisation utilisée, l'intensité des pics peptidiques dans les spectres de masse n'est pas suffisamment fiable pour permettre une quantification directe.

Ce problème a suscité le développement de méthodes permettant de déterminer les quantités relative et absolue des protéines dans les échantillons en utilisant la spectrométrie de masse. La protéomique basée sur les peptides peut parfois nécessiter l'introduction d'une forme de marquage avec des isotopes stables pour permettre la quantification (Figure 44).

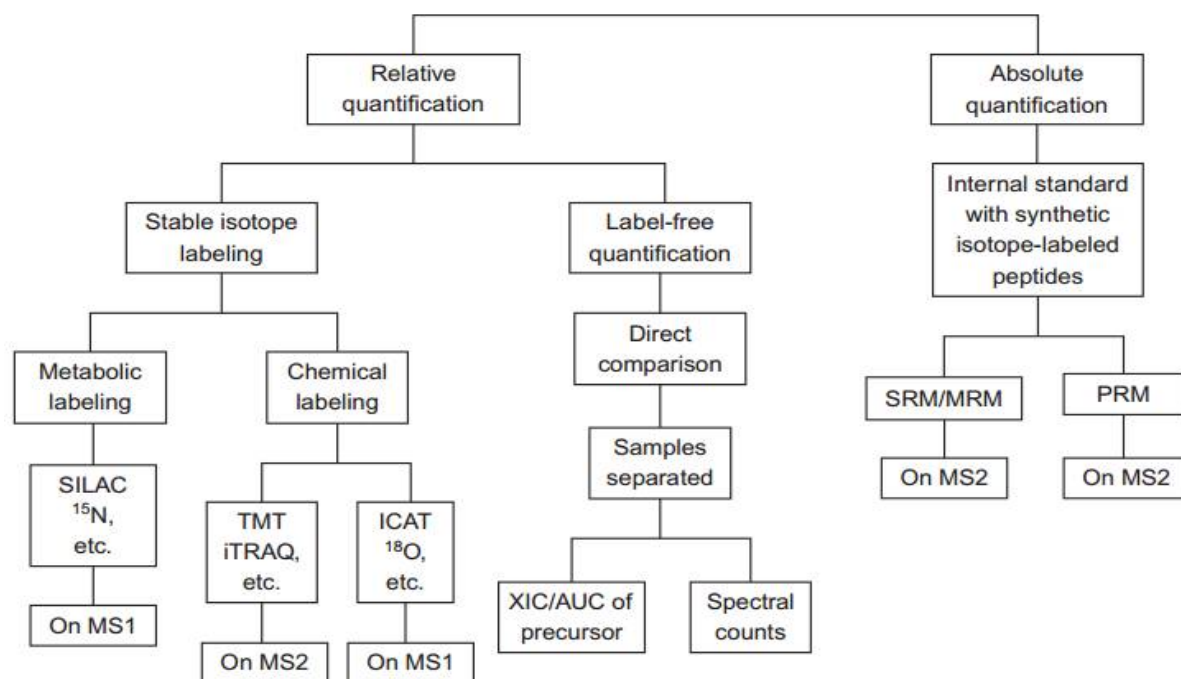


Figure 44: Résumé des stratégies de quantification basées sur la spectrométrie de masse [149]

7.1. La quantification relative des protéines :

La quantification relative des protéines consiste en une comparaison quantitative de deux ou plusieurs échantillons différents. Actuellement, deux approches sont employées pour cet objectif : l'approche sans marquage et l'approche avec marquage. Toutes ces méthodes de quantification relative ont leurs avantages et inconvénients particuliers (Tableau 8), et c'est pourquoi la méthode quantitative doit être choisie en fonction du type d'échantillon et de l'objectif des expériences à mener.

7.1.1. L'approche sans marquage :

La quantification sans marquage (ou *label-free quantification*) est une méthode spectrométrique « naturelle » qui détermine la quantité relative de protéines dans deux ou plusieurs échantillons biologiques, mais contrairement aux autres méthodes de quantification relative, elle n'utilise pas un isotope stable qui marque les peptides.

La quantification sans marquage permet de comparer, d'une façon qualitative et quantitative, les protéomes de différents états physiologiques ou individus, souvent ceux ayant une pathologie contre des sujets sains. Ceci permet ainsi, par exemple, d'identifier de nouveaux biomarqueurs dont la présence ou la teneur confirme un état physiologique donné.

Contrairement aux approches de marquage des isotopes stables dans lesquelles des échantillons de protéines marqués différemment sont combinés ensemble et analysés lors des mêmes expériences LC-MS/MS, les échantillons de protéines dans les approches sans marquage sont mesurés individuellement et séparément, ainsi les déviations causées par la préparation et prétraitement des échantillons et les instruments génèrent plus de variabilité au cours de l'acquisition des données.

La quantification sans marquage applique deux stratégies de quantification : le comptage des spectres ou la méthode de chromatogramme des ions extraits (XIC).

- **Le comptage des spectres :**

Il s'agit de la technique sans marquage la moins coûteuse et la plus simple. La quantification par comptage spectral est une méthode semi-quantitative qui repose sur le nombre de peptides identifiés à partir d'une protéine donnée. Ceci permet d'estimer les niveaux d'expression des peptides au moyen d'un comptage de spectres obtenus à partir de données MS/MS [150].

- **La méthode XIC :**

Un chromatogramme XIC est un tracé de l'intensité du signal spectrométrique à une ou plusieurs valeurs m/z sélectionnées dans une série de spectres de masse qui sont enregistrées en fonction du temps de rétention chromatographique. Cette méthode consiste à comparer les profils (obtenus via LC-MS) des échantillons et à en déterminer les ratios et l'aire-sous-la-courbe (ASC ou AUC pour *Area Under the Curve*) des pics des différents peptides. L'ASC est ensuite rapportée à celle acquise pour les mêmes peptides dans un échantillon de référence, et par conséquent, on peut quantifier **relativement** la teneur en protéine dans l'échantillon étudié.

Sur la figure ci-dessous (Figure 45), deux peptides sont représentés. Celui qui se termine par K a une concentration égale dans les deux échantillons analysés (échantillons témoin et expérimental), et celui qui se termine par R a une concentration deux fois plus élevée dans l'échantillon expérimental que dans le témoin. Le spectre MS1 enregistre le nombre d'ions pour diverses valeurs m/z du peptide intact en élution à un moment donné (B). Le temps d'élution d'un peptide est d'environ 20 secondes. Pendant ce temps, environ 10 spectres MS1 sont collectés, chacun montrant le peptide à des intensités différentes. L'intégration de cette intensité dans le temps se rapproche du nombre total d'ions ionisants dans le spectromètre de masse (C). L'ASC obtenue est une mesure du nombre total d'ions pour un peptide donné, qui peut être comparée avec celle d'un second échantillon pour avoir une quantification relative.

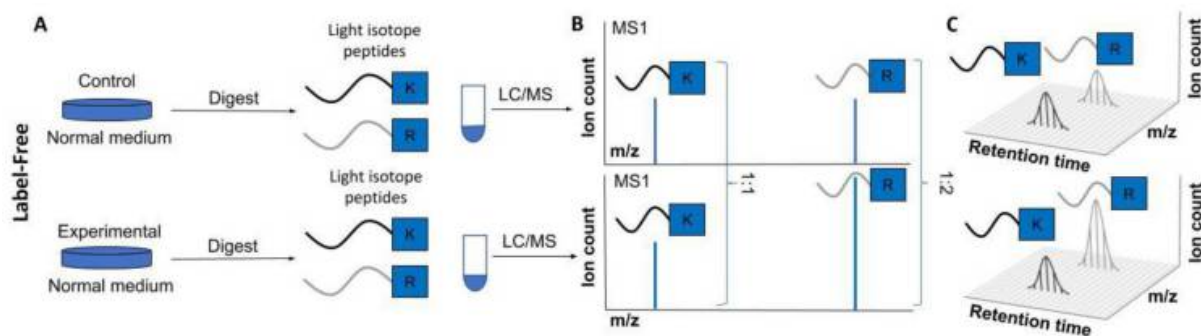


Figure 45: Principes de la quantification par l'approche sans marquage [151]

7.1.2. L'approche avec marquage :

Les méthodes protéomiques quantitatives de marquage basées sur MS se subdivisent en deux classes : le marquage métabolique et le marquage chimique/enzymatique. Les méthodes de marquage métabolique utilisent l'incorporation biologique de marqueurs isotopiques stables dans les protéines des cellules vivantes, y compris le marquage SILAC et ^{15}N , tandis que les méthodes de marquage chimiques/enzymatiques introduisent les marqueurs de masse dans les protéines ou les peptides, y compris le marquage iTRAQ, ICAT, TMT, diméthyle et ^{18}O .

• La technique SILAC :

SILAC (*stable isotope labeling by amino acids in cell culture*) est une technique basée sur la spectrométrie de masse, utilisée pour la quantification relative des protéines entre les échantillons en utilisant des marqueurs à isotopes stables (non radioactifs). Le principe du SILAC est basé sur l'incorporation par voie métabolique d'acides aminés marqués par des isotopes stables, tels que l'arginine ou la lysine marquée au ^{13}C ou au ^{15}N , dans le protéome entier pendant le métabolisme des protéines dans une culture cellulaire. En SILAC, deux populations de cellules sont cultivées dans deux milieux de culture différents, le milieu « léger » contenant un ou plusieurs acides aminés avec l'isotope naturel, et le milieu « lourd » contenant un ou plusieurs acides aminés marqués par un isotope stable.

Après un nombre suffisant de divisions cellulaires, théoriquement toutes les protéines des cellules cultivées en milieu lourd contiennent des acides aminés lourds. Une fois le marquage est complet, des quantités égales de cellules ou d'extraits protéiques marqués et non marqués sont mélangées, puis digérés. Les peptides obtenus sont ensuite analysés par LC-MS/MS. La quantification par SILAC est basée sur l'analyse du rapport entre les peptides marqués par des isotopes introduits et les peptides non marqués. Ainsi, les intensités de signal des échantillons légers et lourds permettent une comparaison quantitative de leurs abondances relatives dans le mélange (Figure 46) [152].

Les grands avantages du SILAC résident dans sa simplicité de mise en œuvre, sa grande précision et sa reproductibilité par rapport aux stratégies. De plus, le multiplexage est possible avec cette technique, c'est-à-dire, la capacité de comparer plus de 2 échantillons dans un seul essai spectrométrique, ce qui réduit le temps d'expérimentation, les variations de résultats et la quantité de l'échantillon [153].

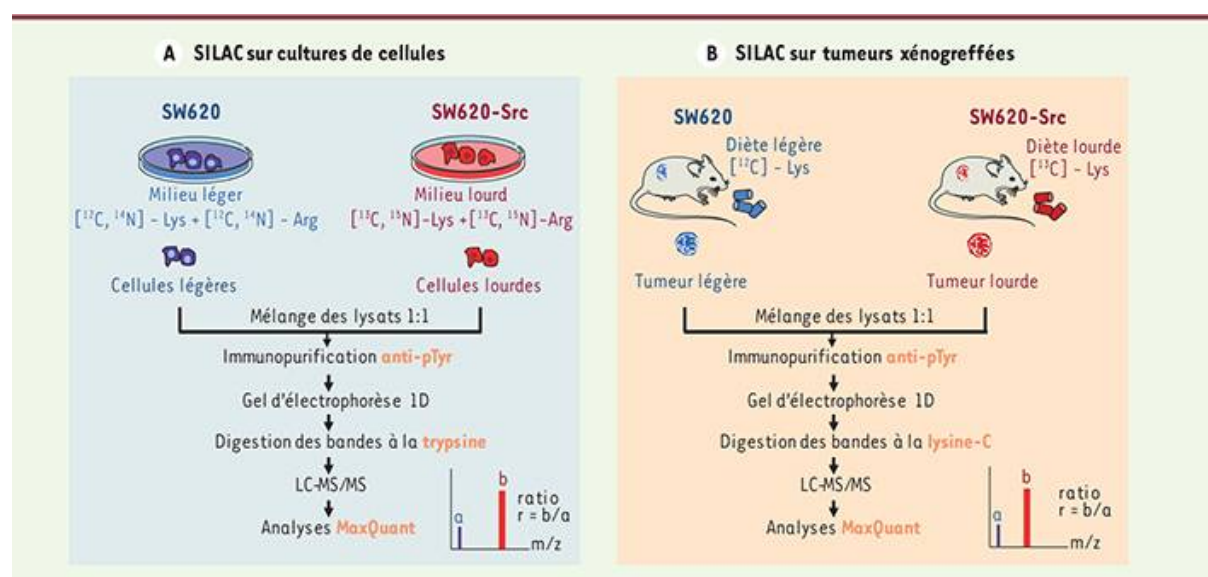


Figure 46: Comparaison de la quantification par SILAC in vitro (A) et in vivo (B) [154]

● Les techniques iTRAQ et TMT :

TMT et iTRAQ sont des méthodes de marquage chimique qui permettent de marquer les amines primaires des acides aminés avec des réactifs chimiques.

Les réactifs de marquage TMT et iTRAQ sont composés de trois segments différents. Le premier de ces segments (groupe reporteur) peut être de poids moléculaire différent et génère un fragment dans les spectres MS/MS qui est utilisé pour la quantification. Le deuxième (groupe équilibreur) fournit un équilibre pour garantir que le réactif a le même poids moléculaire pour une variété de segments de poids différents. Le troisième segment fournit un groupement réactif peptidique, généralement un ester d'hydroxysuccinimide, qui est utilisé pour fixer le réactif quantitativement aux groupements amines primaires (sur les résidus de lysine et l'extrémité N-terminale des protéines) dans tous les peptides. Les spectres MS/MS des peptides dérivés montrent les fragments de le premier segment dans la région de faible poids moléculaire des spectres (la région dite "reporter-ion region") et l'intégration de ces pics donne des quantifications relatives. Les groupes équilibreur et reporteur ont une masse totale de 145 Da dans les expériences en 4-plex et une masse totale de 304 Da dans les expériences en 8-plex. Pourtant, dans 8-plex, la masse de l'ion rapporteur n'est pas de 120, car elle donne une quantification erronée puisque l'ion ammonium de la phénylalanine est également observé à une masse de 120 Da (Figure 47).

L'avantage de ces deux techniques c'est qu'elles sont rapides par rapport aux techniques *in vivo* qui nécessitent une culture cellulaire et elles permettent de conduire des analyses par multiplexage allant jusqu'à 8-plex pour iTRAQ et jusqu'à 16-plex pour TMT [155].

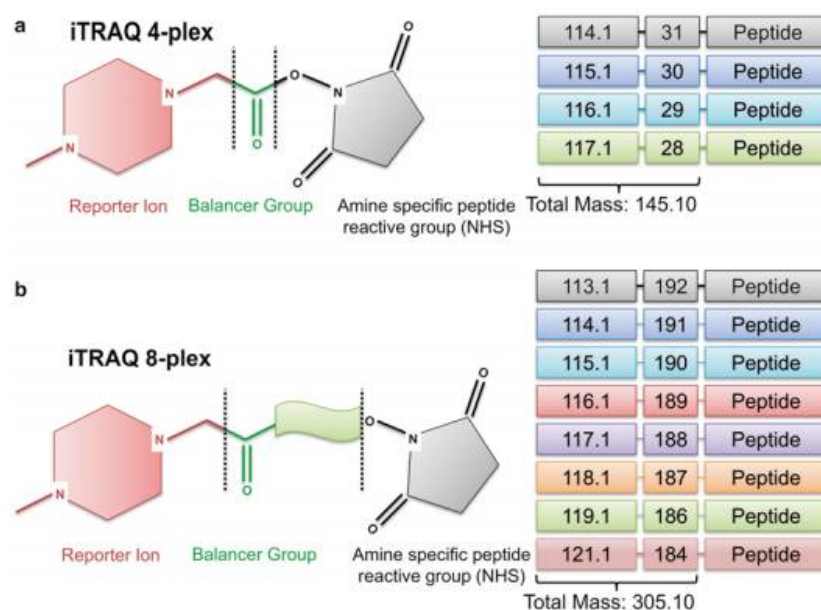


Figure 47: Structures chimiques des réactifs iTRAQ (a) 4-plex et (b) 8-plex.

• La technique ICAT :

L'étiquette d'affinité codée par des isotopes (ICAT) est une technique chimique conçue à l'origine par Gygi et ses collaborateurs [156]. Elle permet d'étiqueter les chaînes latérales de cystéine avec des étiquettes à isotopes « lourds » ou « légers », comme dans la technique SILAC. Les échantillons étiquetés distinctement sont ensuite mélangés ensemble, séparés par chromatographie en phase liquide et soumis à une analyse MS/MS.

Le réactif ICAT se compose de trois éléments (Figure 48) : un marqueur d'affinité (biotine), qui est utilisé pour isoler les peptides marqués ICAT ; un liant contenant des isotopes stables, qui peuvent être différenciés par spectrométrie de masse ; et un groupe réactif (iodoacétamide) ayant une spécificité envers les groupes thiol (c'est-à-dire envers les cystéines). Le réactif existe sous deux formes : lourde et légère. Dans le réactif ICAT original, la forme lourde contient huit deutériums et la forme légère ne contient aucun deutérium dans le liant, ce qui donne une différence de masse de 8 Da.

La différence entre cette technique et les autres approches de marquage chimique est la possibilité du marquage des protéines entières.

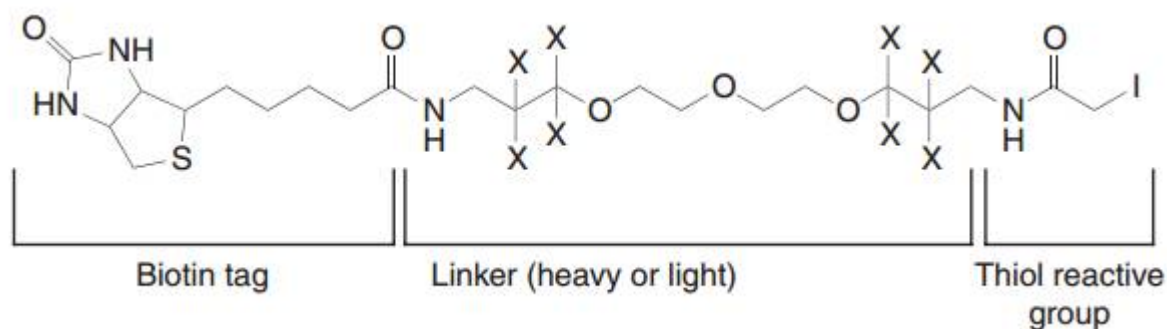


Figure 48: Structure chimique générale d'un réactif ICAT. (X = hydrogène ou deutérium)

L'ICAT est une technique précise qui peut être utile pour analyser des mélanges complexes de protéines. Elle est hautement automatisée et permet de séquencer directement les peptides à l'aide de la spectrométrie de masse en tandem. Toutefois, la principale limite de cette technique est son biais pour les protéines riches en cystéine. Le réactif ICAT de grande taille interfère souvent avec la fragmentation de la MS et réduit la qualité des spectres MS obtenus.

Quantification methods	Way to introduce label	Labeling level	Sample type	Number of conditions to be compared	Quantification accuracy	Quantitative proteome coverage	Linear dynamic range
<i>Metabolic labeling</i>							
SILAC	ex vivo, in vivo	Protein	Cells, expand to tissues or model organisms	Up to 5	+++	++	1–2 log
¹⁵ N	ex vivo, in vivo	Protein	Cells	2	+++	++	1–2 log
<i>Chemical labeling</i>							
ICAT	in vitro	Protein	Applicable to any sample (cells, animal, or human tissue samples)	2	++	++	2 log
iTRAQ	in vitro	Peptide		up to 4 or 8	++	++	2 log
TMT	in vitro	Peptide		up to 6 or 10	++	++	2 log
Dimethyl labeling	in vitro	Peptide		2–3	++	++	2 log
<i>Enzymatic labeling</i>							
¹⁸ O labeling	in vitro	Peptide	Applicable to any sample	2	++	++	1–2 log
<i>Label free</i>							
Ion intensity	n.a.	n.a.	Applicable to any sample	No limit	+	+++	2–3 log
Spectrum counting	n.a.	n.a.	(cells, tissue samples)	No limit	+	+++	2–3 log

n.a., not applicable; +, good; ++, very good; +++, excellent.

Tableau 8: Caractéristiques et applications des méthodes quantitatives basées sur MS [157]

7.2. La quantification absolue des protéines :

Les méthodes décrites dans la rubrique précédente fournissent des informations sur l'abondance relative, mais certaines applications, notamment les mesures de biomarqueurs, nécessitent une quantification absolue. La quantification absolue des protéines se fait principalement par rapport à un étalon interne, dont la concentration absolue est connue. En raison du coût élevé de ces étalons, ces expériences sont généralement limitées à la protéomique ciblée (l'étude d'un sous-ensemble plus petit du protéome). La quantification absolue, en revanche, établit la concentration exacte d'une protéine dans un mélange, et les résultats sont exprimés en unités massiques ou molaires, au contraire de la quantification relative.

En utilisant un étalon interne (ou étalon externe parfois), l'abondance absolue de toutes les protéines détectées dans un échantillon peut donc être déduite par un simple calcul.

Actuellement, trois approches principales de quantification absolue par marquage isotopique sont employées en protéomique (AQUA, QconCAT, et PSAQ), auxquelles s'ajoute une approche récemment développée : FLEXIQuant (Tableau 9).

- **La technique AQUA :**

AQUA est l'une des premières méthodes de quantification absolue développées et est toujours largement utilisée. Cette technique utilise des étalons internes comprenant des peptides synthétiques qui sont identiques aux peptides contenus dans l'échantillon d'intérêt, sauf qu'ils sont marqués avec des isotopes stables. Une concentration connue du peptide marqué est ajoutée aux lysats cellulaires pendant la digestion, et une analyse ultérieure par SRM permet d'identifier et de quantifier avec une grande précision le peptide d'intérêt en déterminant le rapport entre le peptide marqué (étalon) et le peptide non marqué (échantillon).

- **La technique QconCAT ou QCAT :**

La technique QconCAT vient de compléter la technique de quantification absolue par marquage isotopique AQUA après son introduction en 2005 (Beynon et al., 2005). Cette technique utilise des protéines artificielles constituées de peptides concaténés (appelées aussi « peptides quantotypiques »), dont chaque peptide constitue un étalon interne qui représente

une protéine spécifique. La construction d'une protéine QconCAT est exprimée à l'aide d'un gène artificiel dans *Escherichia coli* cultivé dans un milieu enrichi en isotopes stables.

Chaque protéine QconCAT est construite dans un contexte de séquence commun qui prévoit des marqueurs de purification (His Tag), des marqueurs de standardisation (peptides S1 et S2, un N-terminal et un C-terminal) et des peptides sacrificiels protecteurs (MXXK, ou en plus, MxxR) ainsi que les peptides quantotypiques (Q1-Qn) (Figure 49).

En effet, la protéine concaténée étant distincte de la protéine native d'intérêt, elle peut se comporter de manière différente lors de la digestion enzymatique.

Pour les études multiplex, l'approche QconCAT se trouve avantageuse car elle est moins coûteuse que les peptides individuels AQUA et moins laborieuse que la technique PSAQ [158].

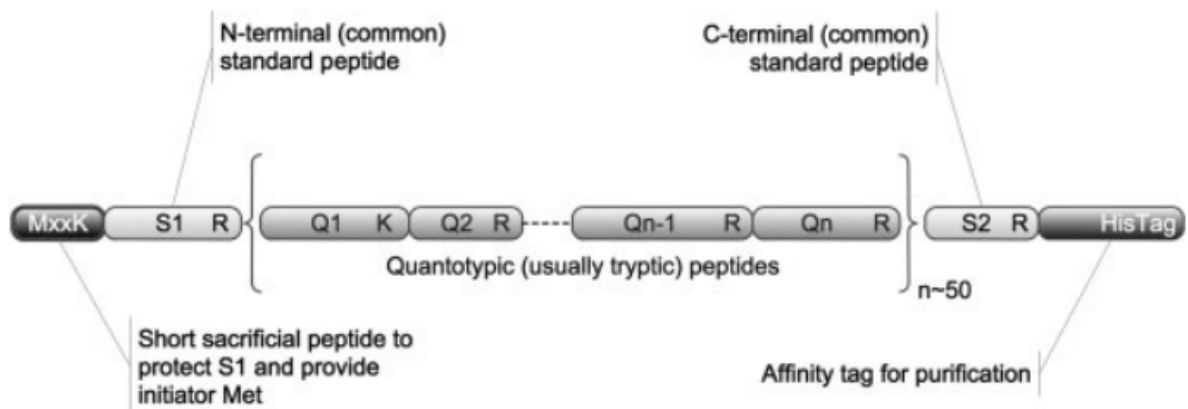


Figure 49: Structure générale d'une protéine chimérique QconCAT [159]

- **La technique PSAQ :**

Cette technique utilise un système d'expression cellulaire (*E. coli*) ou non-cellulaire (CFPS/IVT) [160] pour générer des protéines marquées par des isotopes lourds qui sont purifiées et quantifiées avant d'être ajoutées à des échantillons complexes.

Les techniques AQUA et QconCAT reposent sur l'utilisation de peptides étalons pour le processus de quantification, tandis que les protéines PSAQ sont des protéines qui conservent la séquence native dans laquelle les peptides quantifiés existent, et par conséquent, toute différence entre l'analyte et l'étalon dans le clivage protéolytique et les pertes de procédure sont minimisées. Cette technique permet donc d'obtenir une quantification très précise.

• **La technique FLEXIQuant :**

FLEXIQuant (*Full-Length Expressed Stable Isotope-labeled Proteins for Quantification*) est une approche récente qui utilise des protéines marquées sur toute leur longueur comme des étalons internes, bien que FLEXIQuant génère les étalons in vitro en utilisant un extrait de germe de blé. Cette technique est principalement employée pour l'identification et la quantification absolue des modifications post-traductionnelles, et c'est pourquoi la protéine synthétisée doit être dépourvue de toute modification pour servir d'étalon approprié [161].

	Standard Types	Measured Values	Spiking Time Point	Normalization	Accuracy	Coverage	Applicability to Post-Translational Modification	Noise Origins
Isotope labeling	Synthetic peptide	Ratio to standard	Before or after digestion	—	Medium	Low	Applicable	S/N of ion peak, Missed cleavage
	Intact protein	Ratio to standard	Just after protein extraction	—	Very High	Low	Not	S/N of ion peak
	Peptide concatemer (QconCAT, PCS)	Ratio to standard	Prior to digestion	—	High	Medium	Not	S/N of ion peak, Missed cleavage (QconCAT)
Label-free	—	Peak intensity	—	Average of most intense three peaks	Low	Medium	Not	Variation of ionization efficiency
	—	Spectral count	—	Observable peptides	Low	High	Not	Stochastically calculated index

Tableau 9: Résumé des approches pour la quantification absolue de l'abondance des protéines par spectrométrie de masse [162]

8. La bioinformatique :

Dans la protéomiques, les données collectées par la spectrométrie de masse n'ont aucun sens sans un traitement et interprétation par le système bioinformatique. La protéomique dépend essentiellement de la bioinformatique pour traiter les données brutes du spectre de masse en données sur les protéines. Bien qu'ils soient couramment utilisés par tous les laboratoires, les logiciels les plus critiques sont ceux qui prennent la cartographie des peptides et/ou les résultats de la spectrométrie de masse en tandem et déterminent la protéine ou la séquence peptidique qui correspond le plus étroitement aux données expérimentales.

Pour aider la communauté de la recherche scientifique, divers outils bioinformatiques, bases de données et algorithmes de calcul ont été développés pour le stockage, la diffusion et l'analyse ultérieure de ces données protéomiques.

La bioinformatique est l'utilisation des technologies de l'information pour analyser les données protéomiques. Pour l'identification des protéines, trois techniques de recherche dans les bases de données sont utilisées, dont l'empreinte de masse peptidique (PMF), la MS/MS non interprétée et les étiquettes de séquence (sequence tags). Dans la PMF, les protéines sont séparées par SDS-PAGE 1D ou 2D, soumises à une digestion en gel et à un profilage de masse peptidique. Les résultats sont recherchés dans une base de données de protéines traduites théoriquement à partir de gènes connus. La PMF est rapide et pratique, mais ses inconvénients sont évidents : l'analyte doit avoir une pureté relativement élevée et la confiance dans l'identification est faible. La technique de MS/MS non interprétée a amélioré l'identification des protéines et est entrée dans une nouvelle ère où les spectres de MS/MS sont impliqués. En se basant sur la masse de l'ion précurseur, plusieurs peptides candidats de la base de données sont déterminés. Le système de notation utilisant différents algorithmes est ensuite utilisé pour classer la signification de l'identification. Bien que la recherche soit lente, la méthode MS/MS non interprétée permet une identification de grande confiance dans un mélange de protéines. La troisième méthode, le marquage de séquence, utilise la masse de l'ion précurseur et des ions produits contenant trois acides aminés ou plus pour l'identification. Elle permet la caractérisation des peptides par PTM mais augmente également l'ambiguïté. Quelle que soit la méthode bioinformatique utilisée, un fait essentiel est que la plupart des séquences de peptides d'environ six acides aminés ou plus sont largement uniques dans le protéome d'un organisme. Par conséquent, il est important d'avoir un ion précurseur de longueur appropriée pour l'identification de la protéine au cours de la recherche dans la base de données.

Troisième partie (III) : APPLICATIONS EN PROTEOMIQUE

1. La protéomique fonctionnelle :

La protéomique fonctionnelle se concentre sur l'élucidation des fonctions biologiques de protéines inconnues et la définition des mécanismes cellulaires et subcellulaires au niveau moléculaire. Dans les cellules, de nombreuses protéines exposent leurs fonctions biologiques par l'association rapide et transitoire au sein de grands complexes protéiques. La compréhension des fonctions des protéines ainsi que l'élucidation des mécanismes moléculaires au sein de la cellule dépendent alors de l'identification des partenaires protéiques en interaction. L'association d'une protéine inconnue avec des partenaires appartenant à un complexe protéique spécifique impliqué dans un mécanisme particulier serait, en fait, fortement suggestive de sa fonction biologique. En outre, une description détaillée des voies de signalisation cellulaire pourrait largement bénéficier de l'élucidation des interactions protéine-protéine in vivo [163].

2. La protéomique différentielle :

Généralement, les études de protéomique différentielle visent à étudier les modèles d'expression de protéines dans les cellules anormales (malignes, pathologiques, ou affectées par un traitement médicamenteux, etc.), par rapport aux cellules normales. Dans les applications en biologie médicale, cette approche comparative est généralement utilisée pour identifier les protéines qui sont régulées positivement ou négativement de manière spécifique à la maladie pour être utilisées comme marqueurs de diagnostic, pronostic ou cibles thérapeutiques. Dans ces études, une analyse fiable des changements quantitatifs de l'expression des protéines est essentielle. Ces changements sont souvent initialement observés à partir de l'intensité de la coloration des taches protéiques sur les gels.

3. Applications dans la recherche médicale et biomédicale :

3.1. Relation protéome-pathologie :

Les protéines et leurs fonctions sont déterminées par leur localisation spécifique dans les cellules, les événements d'épissage, les différentes modifications post-traductionnelles et les interactions avec d'autres protéines. Tout cela ajoute à la complexité du protéome, mais fournit également des informations précieuses pour les chercheurs sur l'état physiologique des cellules.

Les protéines représentent les éléments clés de chaque organisme. Si les informations fonctionnelles et structurales exactes et les capacités de mise en réseau de toutes les protéines d'un organisme étaient disponibles, alors il devrait être possible de définir exactement toutes les fonctions de cet organisme, son état physiologique et les maladies au niveau moléculaire. Ces considérations constituent le fondement de la protéomique, qui vise à évaluer le protéome de manière holistique, idéalement toutes les protéines contenues dans l'échantillon étudié. Les maladies sont généralement le résultat de changements protéomiques, la pathologie (moléculaire) étant la conséquence de changements spécifiques dans la structure, l'abondance et/ou la fonction des protéines. Ceci n'est pas surprenant tant qu'en général, tous les médicaments agissent au niveau du protéome, en ciblant plus ou moins spécifiquement les protéines.

3.2. Découverte de nouveaux biomarqueurs :

La découverte de biomarqueurs a augmenté de façon spectaculaire au cours de la dernière décennie, sous l'impulsion d'une demande croissante de nouvelles modalités thérapeutiques telles que la médecine régénérative et la disponibilité de nouvelles technologies « omiques ».

En 2001, le terme biomarqueur a été défini par un groupe de travail faisant partie de National Institutes of Health comme « une caractéristique qui est objectivement mesurée et évaluée comme un indicateur de processus biologiques normaux, de processus pathogènes ou de réponses pharmacologiques à une intervention thérapeutique » [164].

Les biomarqueurs comprennent les gènes, les protéines, les variations génétiques et les différences d'expression métabolique provenant de différentes sources telles que les fluides ou les tissus corporels (Tableau 10).

La découverte de biomarqueurs par la protéomique suit une série d'étapes distinctes (Figure 50). Cette recherche comprend d'abord une étape préliminaire pour la détermination du plan d'expériences (question clinique, type échantillon et flux de travail...). L'étape suivante dite de « découverte » ou de criblage est achevée par l'utilisation d'une technique de protéomique explorative (approche bottom-up). Cette étape peut être réalisée sur de petites cohortes, ainsi qu'à partir d'échantillons de patients (sang, biopsie de tumeur fraîche ou fixée, selles, etc.) dont le contenu protéique sera extrait et analysé par spectrométrie de masse.

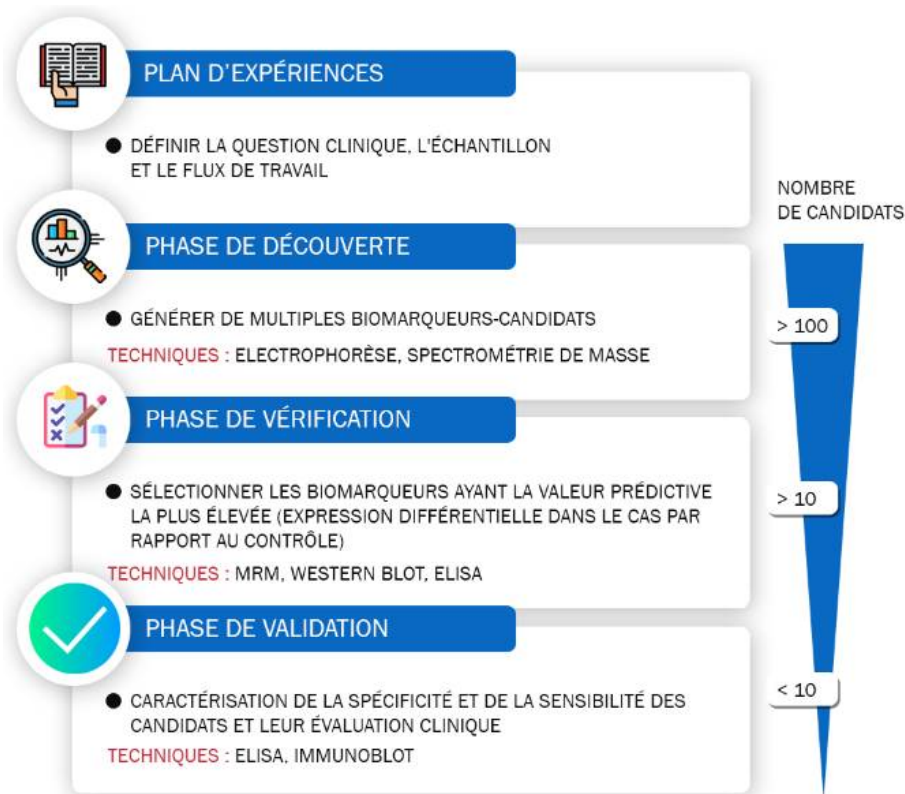


Figure 50: Schéma des différentes étapes du développement des biomarqueurs [165, 166]

Après la phase de découverte, les protéines présentant les plus grandes variations entre les différents groupes sont sélectionnées pour la phase de vérification. Cette étape implique une protéomique ciblée utilisant différentes techniques, telles que la SRM-MS (surveillance des réactions sélectionnées - spectrométrie de masse), la MRM-MS (surveillance des réactions multiples - spectrométrie de masse), ou le SWATH. Cette étape permettra une quantification précise des protéines cibles. Elle implique généralement une cohorte plus importante que lors de l'étape de découverte, bien que l'étape de vérification soit souvent effectuée sur des modèles cellulaires ou animaux avant d'être appliquée à une cohorte de patients. Enfin, l'étape de validation est la phase d'évaluation clinique des biomarqueurs.

	Biomarqueur génomique	Biomarqueur transcriptomique	Biomarqueur protéomique	Biomarqueur métabolomique
Technique à bas débit	Technique de séquençage	PCR Northern blot	-----	RMN et chromatographie en phase gazeuse
Technique à moyen débit	Spectrométrie de masse	SAGE	Spectrométrie de masse et ELISA	Chromatographie liquide/spectrométrie de masse
Technique à haut débit	PCR	Puce à ADN	Puce à protéines Puce à anticorps	-----

Tableau 10: Comparaison des techniques de découverte des biomarqueurs usuelles [167]

• Détection de pathologies neurodégénératives :

Notre compréhension actuelle des changements moléculaires qui déterminent la pathogenèse des maladies neurodégénératives est incomplète. Des études protéomiques impartiales, basées sur la spectrométrie de masse, fournissent un moyen efficace et complet d'examiner quantitativement des milliers de protéines à la fois en utilisant des quantités microscopiques de tissu cérébral humain.

Des preuves empiriques suggèrent que le génotype APOE renseigne sur le risque pré-symptomatique d'une grande variété de maladies, et est précieux pour le diagnostic de la dysbetalipoprotéinémie de type III, augmente le risque de démence dans les maladies neurodégénératives et est associé à un mauvais pronostic après une lésion cérébrale aiguë. Le génotype APOE4 est le facteur le plus important du risque génétique de la maladie d'Alzheimer à apparition tardive, dont les porteurs ont un risque plus élevé de développer la maladie que les non-porteurs [168].

Dans une étude protéomique publiée cette année par Konijnenberg et ses collègues, 142 protéines parmi un total de 313 ont été analysées avec la surveillance des réactions multiples (MRM) par spectroscopie de masse ciblée, et de nombreuses différences dans le protéome du LCR ont été observées entre les porteurs et les non-porteurs du génotype ApoE4. Les altérations des concentrations de protéines du complément ont été accompagnées d'une série de marqueurs inflammatoires chez les porteurs d'APOE4, certains d'entre eux montrant des niveaux altérés chez les porteurs qui avaient encore des niveaux amyloïdes normaux β , ce qui suggère que les processus d'inflammation peuvent jouer un rôle dans le développement de la MA avant que l'agrégation amyloïde ne devienne manifeste dans le LCR [169].

De plus, dans une autre étude publiée par Abdi et ses collègues, plus de 1500 protéines ont été identifiées en utilisant l'approche quantitative iTRAQ. Ces chercheurs ont constaté que 136 protéines étaient modifiées de manière unique dans le LCR de la maladie d'Alzheimer par rapport au LCR de la maladie de Parkinson, de la démence avec corps de Lewy, et des témoins [170]. Ces observations peuvent induire les chercheurs à explorer et à étudier davantage ces différences pour mieux comprendre les mécanismes moléculaires soulignant cette triade neurodégénérative. Sur la base de ces résultats, Zhang et ses collègues ont effectué une validation à grande échelle de leurs meilleurs biomarqueurs-candidats potentiels par une analyse multiplex et ont constaté qu'un panel de huit protéines (Protéine tau, β -amyloïde 42, β -2-microglobuline, interleukine-8, protéine de liaison à la vitamine D, apolipoprotéines A-II et E, et BDNF) était très efficace pour détecter la maladie de Parkinson [171].

La neurogenèse est le processus biologique par lequel de nouveaux neurones sont formés. Les cellules souches neurales prolifèrent, migrent et se différencient au cours de l'embryogenèse en neurones matures qui formeront finalement le système nerveux central. Lugert et al. ont utilisé des cellules souches neurales humaines pour rechercher des biomarqueurs potentiels du LCR de la différenciation neuronale, et ont identifié le protéoglycan Glypican-2 (Gpc2) comme un marqueur putatif sécrété par des neurones immatures en utilisant le profilage de l'expression protéique basé sur MS [172].

Finalement, les biomarqueurs jouent un rôle essentiel dans l'amélioration du processus de développement des médicaments, ainsi que dans la recherche biomédicale. La compréhension de la relation entre les processus biologiques mesurables et les résultats cliniques est essentielle pour élargir notre arsenal de traitements pour toutes les maladies, et pour approfondir notre compréhension de la physiologie normale et saine.

3.3. Surveillance de la progression du cancer :

Un autre avantage potentiel de la recherche en protéomique relève de la capacité prédictive de biomarqueurs. L'approche protéomique dans l'étude du cancer du sein a également progressé et a donné des résultats prometteurs avec des applications diagnostiques et thérapeutiques. Un exemple d'approches combinées in vitro et in vivo impliquant une analyse approfondie de lignées cellulaires de cancer du sein cultivées a été obtenu à partir de tumeurs à des stades définis du cancer du sein et validé à l'aide de tissus de cancer du sein humain. Cette approche a démontré que les signatures protéomiques spécifiques des stades de la tumeur extraites de l'étude in vitro ont été validées sur des micropuces tissulaires. Les cellules transformées ont montré des signatures protéomiques caractérisant la perte d'architecture tissulaire et les changements métaboliques cellulaires. En outre, l'analyse protéomique a également identifié des niveaux élevés d>IDH2 et de CRABP2 et de faibles niveaux de SEC14L2 comme étant des marqueurs de pronostic pour la survie globale du cancer du sein [173].

3.4. Diagnostic biologique de pathologies :

Un certain nombre de biomarqueurs protéomiques urinaires candidats ont été identifiés, qui peuvent prédire l'évolution de la maladie rénale dans le diabète. Dans une première étude sur la protéomique urinaire chez les Indiens Pima, les profils protéomiques ont permis de prédire la macroalbuminurie 10 ans avant le développement de la néphropathie [174]. Merchant et al ont identifié trois peptides qui diminuent (fragments de collagène de type IV et de type V $\alpha 1$ et de tenascin-X) et trois peptides qui augmentent (fragments d'inositol-pentakisphosphate 2-kinase, zona occludens 3 et FAT tumor suppressor 2) dans l'urine de patients atteints de diabète de type 1 et de déclin fonctionnel rénal précoce [175]. Dans une étude sur le diabète de type 1 (n = 465), un panel de quatre biomarqueurs protéiques (glycoprotéine de Tamm-Horsfall, $\alpha 1$ -acide-glycoprotéine, clusterine et progranuline) a prédit des lésions rénales précoces [176]. Associés aux données des biopsies rénales, les résultats ont indiqué que les fragments de peptides urinaires reflètent des changements dans l'expression des protéines intactes dans le rein.

Des preuves supplémentaires de la capacité de la protéomique urinaire à prédire le déclin fonctionnel rénal ont récemment été fournies, lorsque l'haptoglobine urinaire s'est avérée capable de prédire le déclin fonctionnel rénal précoce chez les patients atteints de diabète de type 2 [177].

Dans une étude transversale comparant des patients atteints d'IRC ayant des étiologies sous-jacentes variables de la maladie avec des témoins sains, 273 peptides (connus plus tard sous le nom de score CKD273) ont été identifiés comme étant associés à l'IRC [178].

3.5. Suivi thérapeutique :

L'utilisation de la protéomique dans la recherche des marqueurs biologiques témoignant l'évolution d'un processus pathogénique suivant le traitement par un agent médicamenteux peut améliorer la prise en charge thérapeutique des patients en détectant l'inefficacité possible du traitement précocement.

3.6. Elucidation de mécanismes physiopathologiques :

La protéomique constitue également un outil efficace pour l'élucidation des hypothèses sur les mécanismes moléculaires et physiopathologiques de diverses pathologies.

En considérant le psoriasis comme exemple, plusieurs études ont été conduites dans l'objectif de déconstruire les éléments protéiques pouvant être liés au mécanisme physiopathologique de son processus inflammatoire.

Le psoriasis est une maladie de la peau inflammatoire, chronique et à médiation immunitaire, qui touche l'homme dans le monde entier et dont la prévalence et l'incidence augmentent en fonction des zones géographiques. La forme la plus courante est le psoriasis en plaques (ou *psoriasis vulgaris*), qui se caractérise par des plaques érythémateuses bien délimitées, avec des écailles argentées, représentant environ 90 % de tous les patients psoriasiques [179]. On pense que le psoriasis est principalement dû à un système immunitaire inné et adaptatif qui peut être modifié par des facteurs génétiques et environnementaux (par exemple, l'alcool, les drogues, et le stress). Plusieurs sources de données ont montré que les cellules T, en particulier les cellules Th1 et Th17, jouent un rôle crucial dans ces réponses inflammatoires.

Une autre étude publiée par Geyer et ses collègues (2016) a exploré les effets d'une perte de poids soutenue sur le protéome du plasma humain. Dans une autre étude publiée par Lifang Xie et ses collègues, l'utilisation de l'approche ascendante de la protéomique basée sur l'ESI-LC-MS/MS a conduit à la découverte que les fibroblastes diploïdes humains traités par H₂O₂ augmentent la production de la protéine IGFBP-6. Les résultats obtenus par cette analyse protéomique ont été validés par des analyses Western blot et RT-PCR semi-quantitatives [180]. Ce résultat suggère que l'IGFBP-6 pourrait servir de biomarqueur de la dégénérescence ou des lésions cellulaires in vitro et in vivo. Cette étude est un autre exemple qui confirme l'importance des techniques protéomiques dans la découverte de biomarqueurs et de nouvelles voies associées à une condition cellulaire particulière.

4. Applications en microbiologie :

Le typage des organismes par spectrométrie de masse en tandem (MS/MS), communément appelé « protéotypage », est basé sur l'identification de séquences peptidiques spécifiques qui permettent l'identification non ambiguë des microorganismes [181]. L'unicité des rapports masse/charge et des schémas de fragmentation mesurés en MS/MS permet l'identification des peptides qui différencient les organismes au niveau des sous-espèces. Bien que l'identification classique des agents pathogènes par MS en milieu clinique repose sur la technologie MALDI-TOF sur cellules entières, ce domaine a prospéré grâce aux progrès de la vitesse, de la sensibilité et de la précision des nouveaux instruments de MS au cours de la dernière décennie. Le couplage des instruments de nouvelle génération avec le pouvoir de séparation de la chromatographie liquide fait de la LC-MS/MS une technologie précieuse à mettre en œuvre dans la routine des laboratoires cliniques. Parmi les quelques exemples disponibles dans la littérature, la LC-MS/MS s'est avérée capable de détecter le virus de la grippe purifié et le métapneumovirus humain dans des échantillons cliniques [182].

La capacité d'évaluer l'ensemble du réseau protéique d'une cellule à un moment donné dans l'espace ou dans le temps aura un effet profond sur la science microbienne, car la fonction des protéines est inextricablement liée au phénotype. Bien qu'une telle situation soit encore hors de portée des technologies actuelles, les progrès rapides de la spectrométrie de masse, de la bioinformatique et des technologies de séparation des protéines ont entraîné un changement radical de nos capacités protéomiques actuelles. Par conséquent, un nombre restreint, mais en constante augmentation, de groupes tirent parti de cette technologie de pointe pour en savoir plus sur la physiologie et le métabolisme des micro-organismes [183].

4.1. Applications en bactériologie :

Le typage et la caractérisation rapides et précis des bactéries infectieuses sont d'une grande importance dans les pratiques médicales modernes et sont essentiels pour un diagnostic correct et un traitement efficace des patients. En particulier, avec la circulation rapide de souches pathogènes de bactéries exprimant une multirésistance aux antibiotiques, une détection et une identification rapides et efficaces sont de plus en plus essentielles pour

répondre aux épidémies des maladies infectieuses. Une grande variété de méthodes permettant d'identifier le contenu bactérien dans les échantillons cliniques a été mise au point. Ces méthodes vont des méthodes traditionnelles basées sur la culture, le profilage des isolats résultants en phénotypes, à des techniques moléculaires plus récentes, y compris les tests de PCR et le séquençage de l'ADN pour la détection de biomarqueurs génétiques et pour la classification des génotypes et des lignées génétiques. Bien qu'elles soient bien établies, un bon nombre des méthodes existantes de caractérisation microbienne présentent des inconvénients importants : les méthodes basées sur la culture sont exigeantes en main-d'œuvre et intrinsèquement lentes, et les méthodes basées sur l'ADN, c'est-à-dire le profilage par PCR et le séquençage des gènes, sont généralement limitées aux applications ciblant des caractéristiques connues des bactéries. La dernière décennie a vu des développements rapides dans les technologies de séquençage dites « next generation » (NGS), qui ont permis le dépistage systématique des bactéries par détermination de la séquence du génome entier (WGS) à des coûts décroissants [184]. Le séquençage du génome microbien offre des analyses complètes des agents pathogènes avec un débit élevé et devient un outil important pour le diagnostic en temps réel des infections bactériennes. Toutefois, le WGS ne fournit aucune information sur l'expression des gènes et se limite donc aux informations présentes dans le génotype. En revanche, les récents développements et améliorations des performances des instruments de SM en tandem permettent la caractérisation de parties importantes des protéomes exprimés des bactéries. En ce qui concerne MALDI-TOF, qui a récemment été établi comme un outil de diagnostic moléculaire polyvalent pour les maladies infectieuses, les techniques basées sur la SEP/SM fournissent des informations sur l'expression des gènes au niveau de la résolution des acides aminés. Cela permet une détection rapide et sensible des caractéristiques génotypiques exprimées, telles que les protéines associées à la résistance aux antibiotiques et à la virulence ; ainsi, la spectrométrie de masse en tandem a le potentiel d'améliorer encore l'analyse diagnostique des infections bactériennes (Figure 51).

La spectrométrie de masse MALDI-TOF a été largement adoptée par de nombreux laboratoires de microbiologie clinique au cours de la dernière décennie [185]. Le MALDI-TOF MS est l'outil le plus émergent utilisé pour identifier les microorganismes au niveau de l'espèce, voire la différenciation des souches sur la base de profils moléculaires [186].

À l'heure actuelle, un système commercial MALDI-TOF MS (VITEK® MS) a été approuvé par la Food and Drug Administration américaine après des essais cliniques approfondis et réussis [187].

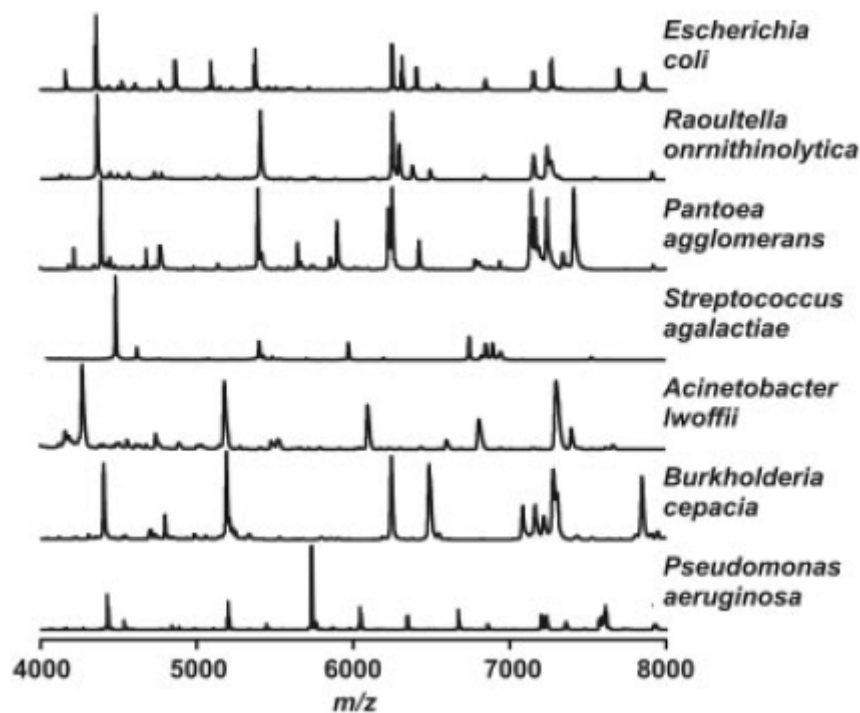


Figure 51: Spectres de masse exemplaires de cellules intactes d'une variété d'espèces bactériennes cliniques courantes

Dans une étude publiée par Sogawa et ses collègues, les colonies bactériennes obtenues à partir d'un total de 468 souches de 92 espèces bactériennes isolées ont été directement placées sur des plaques MALDI cibles. Les plaques ont ensuite été soumises à une mesure MALDI-TOF MS et les micro-organismes ont été identifiés par correspondance avec les bibliothèques du logiciel BioTyper 2.0. Le taux de réussite de l'identification au niveau de l'espèce et du genre était de 91,7 % (429/468) et de 97,0 % (454/468), respectivement. Il est intéressant de noter que les différences entre les colonies et les effets de la durée de culture sur les résultats se sont avérés minimes, et il n'a fallu que 5 minutes pour obtenir les résultats de cette étude, ce qui est beaucoup plus court que les cinq à 48 heures nécessaires aux méthodes conventionnelles pour obtenir les mêmes résultats [188].

4.2. Applications en virologie :

Un virus est un petit agent infectieux qui ne peut se multiplier à l'intérieur d'organismes vivants qu'en dirigeant la machinerie de la cellule hôte pour générer davantage de particules virales. Le matériel génétique d'un virus est soit l'ARN, soit l'ADN.

De nouvelles espèces de virus humains sont toujours identifiées, au rythme de trois ou quatre par an, et les virus représentent plus des deux tiers de tous les nouveaux agents pathogènes humains, d'où la nécessité de l'établissement de méthodes d'étude d'interaction virus-hôte sensibles et rapides [189].

Pour que les virus provoquent une maladie, ils doivent d'abord infecter leur hôte, se propager dans l'hôte et endommager les tissus cibles. Pour assurer leur propagation, les virus doivent ensuite être transmis à d'autres individus sensibles, c'est-à-dire qu'ils doivent être excrétés dans l'environnement à l'aide de sécrétions ou d'excrétions, être absorbés par un autre hôte ou un vecteur, ou être transmis congénitalement de la mère à la progéniture. Les virus ont développé une variété remarquable de stratégies pour assurer leur propre survie. De même, les virus individuels provoquent des maladies par le biais d'une variété considérable de mécanismes pathogènes distincts. La compréhension de ces mécanismes peut garantir une meilleure prise en charge thérapeutique et clinique, ainsi que le développement de nouvelles thérapies médicamenteuses.

En comparant les protéomes des lipoprotéines circulantes de patients infectés par le VHC et de donneurs sains, il a été observé que le VHC induit une altération de l'association de l'ApoA-I avec les lipoprotéines de basse densité [190]. Il est intéressant de noter que la régulation négative de l'ApoA-I induit une déficience significative de la production du VHC, ce qui renforce l'idée que le VHC a besoin des mécanismes lipoprotéiques de l'hôte [191].

L'identification de biomarqueurs peut également être obtenue avec des tissus ou des cultures cellulaires comme source de protéines. Par exemple, en analysant la lignée cellulaire HepG2 d'hépatome répliquant le virus de l'hépatite B avec une approche iTRAQ, Feng et ses collègues ont identifié l'apolipoprotéine (Apo)A-I et la glycoprotéine a2-HS comme candidats biomarqueurs pronostiques pour le carcinome hépatocellulaire (CHC) associé au virus de l'hépatite B [192].

Les études globales sur le protéome ont fourni une multitude d'informations quantitatives concernant l'identification et l'abondance des protéines intracellulaires pendant l'infection virale. Associées à la détection par spectrométrie de masse, des analyses génomiques et bioinformatiques parallèles ont permis la catégorisation fonctionnelle des voies de réponse de l'hôte et l'identification de produits génétiques viraux non caractérisés auparavant.

5. Applications dans la recherche pharmaceutique :

Les chercheurs des domaines médical et pharmaceutique sont constamment à la recherche de nouveaux ou de meilleurs moyens pour traiter les différentes pathologies, afin d'élargir les possibilités de traitements, notamment médicamenteux, ce qui potentiellement permet de contourner les différentes limitations de thérapies existantes et d'améliorer la qualité de vie des patients.

Tout au long du processus de découverte et de développement des médicaments, la protéomique peut aider les chercheurs dans toute une série de processus. Ceux-ci comprennent la recherche de nouvelles cibles thérapeutiques, la compréhension des mécanismes de la maladie, la conception de médicaments basée sur la structure et la recherche du mécanisme d'action des composés principaux.

5.1. Détermination de mécanismes d'action des médicaments :

La compréhension et l'étude de mécanismes d'actions des substances actives permettent non seulement de développer d'autres molécules ciblant le même mécanisme, mais aussi de déceler un processus toxique et les interactions médicamenteuses potentielles. L'utilisation de techniques protéomiques basées sur la spectrométrie de masse peut démasquer des données protéomiques pouvant aider les chercheurs à détecter ces mécanismes d'action.

Une première étude protéomique établissant le profil de l'effet des composants de la combinaison antipaludique CoArtem a été publiée par Makanga et ses collègues [193]. Cette combinaison antipaludique (artéméther et luméfantrine) est largement préconisée comme le médicament antipaludique de choix. Non seulement CoArtem est cliniquement la plus importante combinaison médicamenteuse contenant de l'artémisinine disponible, mais le(s)

mécanisme(s) d'action de chaque composant n'est pas entièrement compris et chaque composant est considéré comme ayant des mécanismes indépendants de destruction du parasite *Plasmodium falciparum*. Le premier point à noter est l'observation d'altérations significatives des schémas d'expression des protéines suite à l'exposition du parasite à ces deux molécules. Deuxièmement, les altérations induites étaient clairement spécifiques au médicament. Par exemple, alors que le traitement par l'artéméther entraîne une réduction plus de trois fois supérieure de l'expression d'un certain nombre d'enzymes glycolytiques (énolase, phosphoglycérate kinase, fructose-biphosphate aldolase et glyceraldéhyde-3-phosphate déshydrogénase), l'expression de ces mêmes enzymes est augmentée de plus de trois fois après le traitement par la luméfantrine.

Plusieurs études protéomiques visant à comprendre les mécanismes d'action des produits extraits des plantes médicinales et utilisés en médecine traditionnelle ont été aussi publiées. Par exemple, La *Scutellaria baicalensis*, ou scutellaire chinoise, est largement utilisée comme plante médicinale en Chine depuis des milliers d'années, où la préparation à partir de ses racines est appelée Huang-Qin. Elle a été appliquée dans le traitement de plusieurs pathologies et symptômes à savoir l'hypertension, la dysenterie, les infections respiratoires et l'inflammation grâce aux composés bioactifs qu'elle contient (wogonine, baicaléine et baicaline) [194, 195] . Les flavones naturelles telles que l'apigénine et la baicaléine ont des activités anticancéreuses observées in vitro ou dans des modèles animaux [196, 197]. Une étude protéomique a démontré, en utilisant l'électrophorèse SDS-PAGE et MALDI-TOF/TOF-MS, que la baicaléine provoquait une élévation de l'abondance de la peroxiredoxine 6 (PRDX6), qui atténue la génération des espèces réactives de l'oxygène et inhibe la croissance des cellules cancéreuses colorectales [198].

5.2. Découverte de nouvelles cibles thérapeutiques :

Les protéines de surface cellulaire sont les principaux bio-effecteurs et les points d'origine des signaux qui sont transmis par la cellule à la suite de changements dans son environnement externe et interne. Elles jouent un rôle essentiel dans les interactions entre cellules, l'immunité cellulaire, les nutriments et le transport des métabolites. Les protéines qui n'adhèrent que temporairement à la membrane plasmique sont définies comme des protéines

de surface cellulaire extrinsèques/périphériques. Les protéines de surface cellulaire incorporées dans la membrane par des liaisons peptidiques covalentes (par exemple, l'hélice alpha) ou des ancrages fixés de manière covalente (par exemple, les lipides) sont définies comme des protéines membranaires intégrales intrinsèques [199].

Pourtant, les protéines de surface cellulaire sont parmi les plus difficiles à préparer et à analyser dans le contexte de la protéomique LC-MS ascendante. Les principaux obstacles sont les suivants : (i) faible solubilité de la plupart des protéines de surface cellulaire dans les tampons aqueux, (ii) degré élevé de modifications post-traductionnelles hétérogènes, et (iii) niveau d'expression relativement faible. Leur isolation/enrichissement avant l'analyse par MS est une procédure difficile et complexe.

En utilisant la spectrométrie de masse, Ducret et ses collègues ont identifié 32 candidats de protéines de surface cellulaire spécifiques aux hépatocytes et ont sélectionné six marqueurs qui peuvent fournir une spécificité adéquate pour cibler les thérapies sur le foie. En outre, ils ont démontré la faisabilité et l'importance de mener de telles études directement dans un tissu cliniquement pertinent, car le protéome de la surface cellulaire des hépatocytes fraîchement isolés a différé sensiblement des lignées cellulaires cultivées [200].

6. Applications en toxicologie :

L'un des domaines émergents est l'application des technologies protéomiques à la recherche toxicologique, ce qui a conduit au développement d'un nouveau domaine, appelé toxicoprotéomique. La toxicoprotéomique, bien qu'elle n'en soit qu'à ses débuts, a attiré une grande attention en toxicologie et présente un grand potentiel d'application. Cette discipline est axée sur l'étude protéomique de la toxicité des substances toxiques, y compris, mais sans s'y limiter, les médicaments, les toxines, les facteurs de stress environnementaux, les produits chimiques et tout autre matériau susceptible de provoquer des réactions pathologiques importantes [201]. Le domaine doit combiner les connaissances en toxicologie, pathologie et autres disciplines pour définir, au niveau du protéome, les voies de toxicité, les modes d'action et les événements moléculaires clés.

Les applications de la protéomique en toxicologie peuvent être divisées en deux grands domaines qui se chevauchent : (1) les études d'investigation ou « mécanistiques » et (2) la toxicologie de dépistage ou prédictive. Étant donné la nouveauté de cette technologie, il est important de reconnaître que les expériences de soutien doivent être conçues à l'aide de protocoles basés sur de bons principes toxicologiques. Il est également nécessaire de procéder à une analyse complète et minutieuse des résultats et il est important que les nouvelles approches, telles que la protéomique, se comparent favorablement aux techniques de toxicologie conventionnelles.

Dans le but d'une évaluation globale du protéome sérique fœtal et de l'identification de biomarqueurs protéiques reflétant des expositions toxiques in utero liées au tabagisme maternel, Colquhoun et ses collègues ont quantifié les protéines dans les sérums de sang de cordon ombilical par la spectrométrie de masse en tandem en conjonction avec des marqueurs isobariques (iTRAQ) [202]. Dans une autre tentative d'identification des biomarqueurs protéiques de l'ovomucoïde (principal allergène alimentaire présent dans le blanc d'œuf de poulet) dans des échantillons de plasma et de sérum de souris BALB/c, la 2D-DIGE suivie d'une détection par LC-MS a été réalisée pour caractériser les biomarqueurs candidats à l'évaluation de l'allergie causée par l'ovomucoïde [203].

Une étude de Bantscheff et ses collègues a démontré que la mesure quantitative des kinases cibles était réalisable en utilisant iTRAQ. Ils ont utilisé des inhibiteurs de kinases non sélectifs immobilisés (kinobeads) pour capturer des centaines de protéines kinases exprimées de manière endogène et des protéines de liaison aux purines qui interagissent avec les inhibiteurs de kinases étudiés. Le profilage quantitatif des cellules K562 traitées par l'imatinib, le dasatinib et le bosutinib a confirmé les cibles connues, y compris les kinases des familles Abl et Src, et a identifié le récepteur tyrosine kinase DDR1 et l'oxydoréductase NQO2 comme nouvelles cibles de l'imatinib [204]. Cette étude de cas démontre que les technologies protéomiques sont des outils précieux pour la confirmation des cibles médicamenteuses connues et l'identification de nouvelles cibles. Les nouvelles cibles découvertes pourraient être associées à des toxicités organiques, ce qui offrirait une grande opportunité d'amélioration du médicament original pour prévenir son interaction avec la ou les cibles induisant la toxicité.

7. Applications extra-médicales :

7.1. Taxonomie des organismes microscopiques :

Tandis que les séquences du génome entier constituent la référence ultime pour définir la phylogénie et la taxonomie microbiennes, les stratégies basées sur des biomarqueurs sélectionnés continuent à fournir les moyens pour la majeure partie des études microbiennes systématiques. La spectrométrie de masse MALDI-TOF, qui génère des modèles de masse protéomiques pour les caractérisations basées sur les « empreintes digitales », a permis une percée importante dans l'identification microbienne. Cependant, la spectrométrie de masse MALDI-TOF est limitée dans le nombre de cibles qui peuvent être détectées pour la caractérisation des souches. Les méthodes avancées de spectrométrie de masse en tandem, dans lesquelles les protéines et les peptides générés à partir de protéines sont caractérisés et identifiés, en utilisant la LC-MS/MS, permettent de détecter des milliers de marqueurs de souches microbiennes exprimées pour des caractérisations et des identifications à haute résolution. Les études de modèles démontrent l'application des analyses protéomiques pour la détection et l'identification au niveau des espèces et des souches bactériennes et pour la caractérisation de bactéries importantes pour l'environnement et diversifiées sur le plan métabolique.

Les approches basées sur la protéomique représentent un complément émergent aux méthodes traditionnelles de caractérisation des microorganismes, permettant l'élucidation des biomarqueurs exprimés de l'information sur la séquence du génome, qui peut être appliquée aux applications de « protéotypage » des microorganismes à tous les niveaux taxonomiques.

L'échantillon bactérien peut être analysé par une approche sur gel ou par spectrométrie de masse (MS). Dans l'approche sur gel, le lysat bactérien est préparé et soumis à une électrophorèse sur gel de polyacrylamide et de dodécylsulfate de sodium (SDS-PAGE) unidimensionnelle ou bidimensionnelle. Le SDS-PAGE peut être analysé en le comparant directement aux images de gel disponibles dans la base de données, ou en excisant les taches de protéines et en utilisant la digestion à la trypsine et l'analyse de masse pour l'identification.

Sur la base de l'analyse du profil protéique ou des protéines identifiées, ou des deux, la bactérie à partir de laquelle le lysat a été préparé sera identifiée à l'aide d'une analyse bioinformatique. Dans l'approche basée sur la spectrométrie de masse, les lysats bactériens ou la cellule entière sont analysés à l'aide de différentes techniques de spectrométrie de masse, telles que MALDI-TOF et ESI-MS. La bactérie inconnue sera identifiée soit en comparant les spectres de masse résultants avec une base de données protéomique collective contenant des spectres de masse de bactéries connues, soit en recherchant et en faisant correspondre la séquence d'un panel de protéines avec des protéines de bactéries connues dans la base de données de protéines (Figure 52).

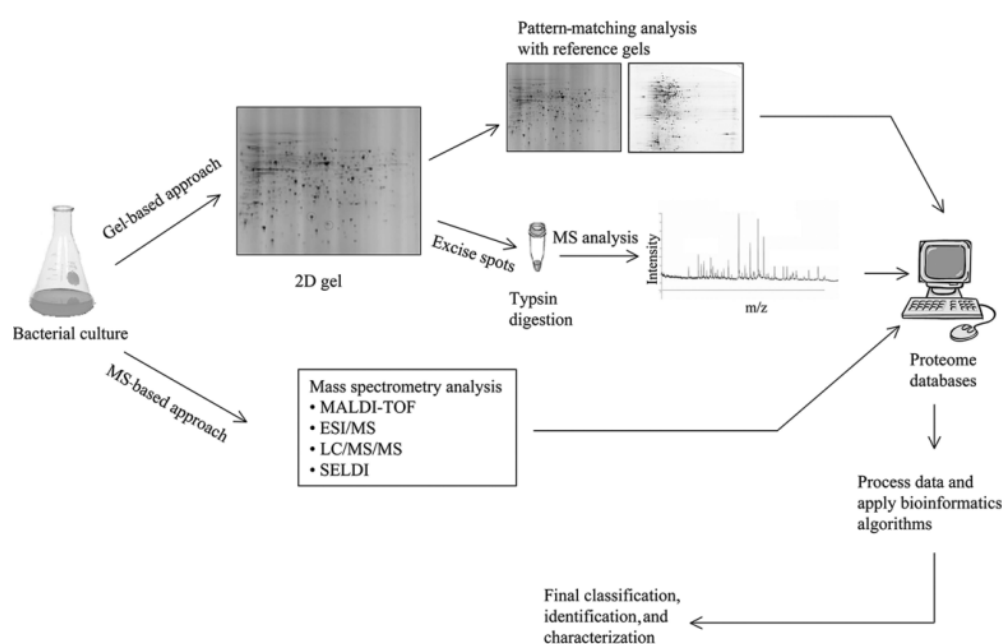


Figure 52: Aperçu des approches protéomiques dans l'identification et la caractérisation des bactéries [205]

Les techniques de protéomique actuellement disponibles présentent des limites auxquelles il convient de remédier. Un nombre des techniques utilisées pour définir les protéomes bactériens analysent les bactéries après leur croissance sur des milieux de laboratoire définis. Il est bien connu que le schéma de synthèse des protéines dans ces conditions ne reflète pas entièrement la synthèse des protéines bactériennes *in vivo* [206].

7.2. Applications en agroalimentaire :

La protéomique est en train de devenir un outil puissant pour la recherche sur les technologies alimentaires, car elle contribue à relever les principaux défis auxquels sont confrontés les analystes et les chercheurs dans le domaine de l'alimentation :

- le développement de méthodologies simples et rapides pour une utilisation en routine ;
- l'analyse de matrices alimentaires complexes ou hautement transformées ; et,
- la quantification de traces d'analytes avec un haut degré de sélectivité.

L'identification des espèces alimentaires a traditionnellement reposé sur l'analyse morphologique. Cependant, les caractéristiques morphologiques sont particulièrement difficiles à différencier entre les certaines espèces, telles que les espèces de fruits de mer, en raison de leurs similitudes phénotypiques. Dans le cas des produits alimentaires transformés, la différenciation des composants peut être encore plus difficile, car les caractéristiques extérieures sont souvent supprimées. Par conséquent, il est urgent la nécessité de disposer de méthodes d'identification moléculaire rapides et fiables qui permettent aux autorités et les industries alimentaires disposant des outils nécessaires pour se conformer aux exigences en matière d'étiquetage et de traçabilité tant au niveau des espèces que de l'origine, ce qui permet de garantir la qualité des produits et de protéger valeur pour le consommateur.

Parmi les exemples de ces applications figurent l'identification et l'analyse taxonomique des espèces de crevettes commercialement pertinentes [207], ainsi que l'identification des espèces commerciales de merlu et de grenadier par analyse protéomique de la fraction parvalbumine [208].

De plus, grâce à LC-ESI-MS et MALDI-TOF-MS, il a été possible de détecter rapidement et avec sensibilité les agents pathogènes ou les micro-organismes d'altération affectant la qualité et la sécurité des aliments pendant la transformation et le stockage [209, 210]. Une description plus précise des micro-organismes contaminants a été obtenue par l'intégration de la protéomique aux méthodes peptidiques et métabolomiques capables de fournir une identification structurelle ou quantitative des métabolites spécifiques produits par les différents micro-organismes d'altération.

8. Applications dans le contexte du COVID-19 :

La pandémie COVID-19 a pris son origine à Wuhan, dans la province de Hubei en Chine, en décembre 2019. Depuis lors, la maladie s'est répandue dans plus de 210 pays et territoires. Il s'agit d'une maladie virale due au virus du syndrome respiratoire aigu sévère 2 (SARS-CoV-2). Les patients présentent des symptômes de type grippal avec une toux sèche, un mal de gorge, une forte fièvre et des problèmes respiratoires. La maladie due au SARS-CoV-2 a été baptisée COVID-19. Plus de 10 millions de personnes ont été infectées, avec plus d'un million de fatalités dans le monde. Les États-Unis d'Amérique sont le pays le plus touché, avec le plus grand nombre de patients, soit environ 0,7 million. Malgré de grands efforts, il n'existe aucun traitement confirmé pour cette maladie, mais heureusement, les vaccins seraient bientôt mis sur le marché. Cependant, la prévention et la gestion sont les meilleures options. La détermination de thérapies efficaces est entravée par une connaissance limitée des détails moléculaires de l'infection, mais certaines études en protéomique pourraient donner des résultats pertinents.

8.1. Découverte de cibles thérapeutiques :

David E. Gordon et ses 124 autres collègues ont publié un article de recherche dans une tentative de déterminer le profil interactomique des protéines de SARS-CoV-2 dont certaines interactions pourraient être exploitées pour le développement de nouveaux principes actifs et pour le repositionnement de médicaments existants [211].

Comme résultat, ces chercheurs ont pu identifier les protéines humaines qui sont physiquement associées à chacune des 26 protéines du SARS-CoV-2 en utilisant la spectrométrie de masse avec purification par affinité, identifiant 332 interactions protéine-protéine entre le SRAS-CoV-2 et les protéines humaines, liées à de multiples processus biologiques, notamment le transport de protéines, la traduction, la transcription et la régulation de l'ubiquitination. Parmi celles-ci, 66 protéines humaines ciblées par 69 composés médicamenteux, dont 29 composés sont déjà approuvés par le corps réglementaire américain FDA, 12 sont en essais cliniques et 28 sont des composés précliniques.

8.2. Recherche de biomarqueurs :

Dans une étude publiée dans Cell Systems, par des chercheurs de l'institut Francis Crick et Charité – Universitätsmedizin Berlin (Messner et al. 2020), 27 biomarqueurs potentiels qui sont présents à différents niveaux chez les patients atteints de COVID-19 – selon la gravité de leurs symptômes – ont été déterminés, par l'analyse de prélèvements de plasma grâce à une technique raffinée de spectrométrie de masse de très haut débit (Figure 53) [212].

Les biomarqueurs découverts étaient validés par le recrutement d'un groupe témoin, composé de 15 volontaires sains et de 17 autres patients souffrant de COVID-19. Ces marqueurs pourraient aider les médecins à prédire l'évolution de la maladie d'un patient pour aider à prendre des décisions sur la meilleure façon de gérer la maladie pour chaque patient, et fournir aux scientifiques du domaine pharmaceutique de nouvelles cibles potentielles pour le développement de nouveaux médicaments.

Une autre étude récemment publiée dans la même revue a réalisé un profilage des échantillons de plasma de patients atteints du COVID-19 et a révélé que les altérations des protéines plasmatiques de l'hôte sont liées au développement clinique de COVID-19. Cette étude a permis d'identifier 11 protéines hôtes et un ensemble de combinaisons de biomarqueurs qui peuvent servir de biomarqueurs en utilisant le pipeline « POC-19 » basé sur l'apprentissage automatique (*machine learning*). En plus de ça, ce travail a également révélé de nombreuses altérations de l'abondance de protéines plasmatiques de l'hôte qui pourraient contribuer à la pathogénèse de COVID-19. Par exemple, les niveaux plasmatiques d'ORM1, ORM2, S100A9, CRP, AZGP1, CFI, SERPINA3/ACT et LCP1/LPL (dont la plupart sont des protéines de phase aiguë), étaient significativement élevés dans des cas plus sévères de COVID-19. La validation sérologique de ces biomarqueurs mis en évidence était faite par ELISA, et les résultats se sont avérés consistants [213].

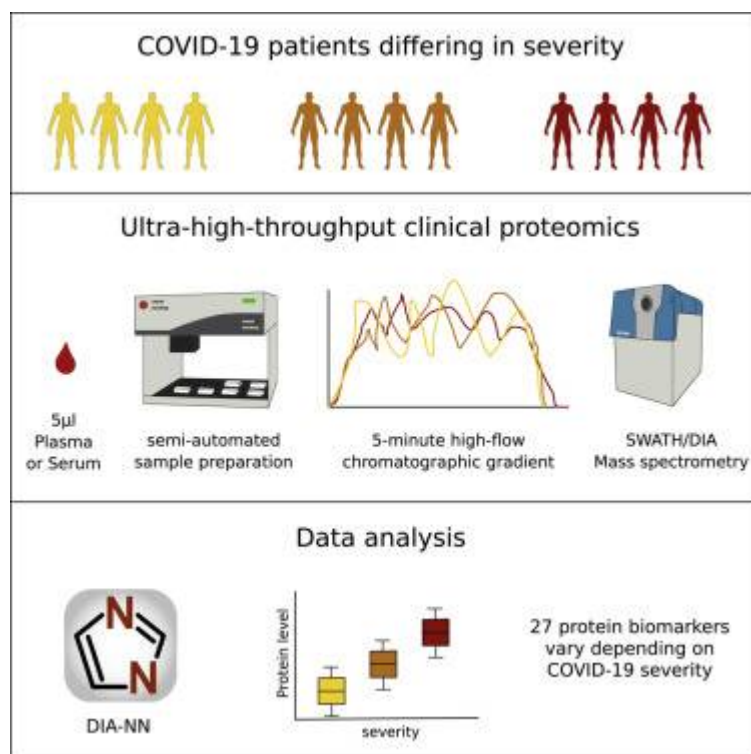


Figure 53: Vue d'ensemble du processus d'analyse protéomique du plasma des patients atteint de COVID-19 [214]

8.3. Identification du SARS-CoV-2 par MS :

Actuellement, le test préféré pour le SARS-CoV-2 est celui de l'amplification de l'ADN par RT-PCR, la méthode la plus efficace et la plus sensible pour le diagnostic de l'infection par le SARS-CoV-2. Néanmoins, quelques études publiées confirment l'utilité de la spectrométrie de masse dans l'identification du SARS-CoV-2 par l'analyse de divers échantillons biologiques. Par exemple, l'étude publiée par Nachtigall et ses collègues a utilisé des échantillons recueillis par écouvillonnage nasopharyngé et préalablement confirmés par PCR, chez 362 personnes, dont 211 personnes sont SARS-CoV-2-positives et 151 sont négatives. L'analyse par MALDI-TOF-MS couplée à six types de méthodes d'apprentissage automatique (DT, NB, KNN, RF, SVM-L et SVM-R) a été menée à bien. En comparant les résultats de la RT-PCR et de la MALDI/ML, ces chercheurs ont trouvé un taux de concordance acceptable comme approche de diagnostic clinique (>80 %) [215].

Une deuxième étude publiée par P. Singh et ses collègues a démontré une méthode LC-MS/MS simple et rapide qui pourrait détecter deux peptides, QIAPGQTGK et AIVSTIQRKYK, à partir de la glycoprotéine S et de la polyprotéine pp1ab (réplicase 1ab) respectivement avec une sensibilité de 90% et une spécificité de 100%. En utilisant des échantillons de suivi de patients qui se sont rétablis symptomatiquement et qui ont également été testés négatifs pour les analyses RT-PCR, ces chercheurs ont montré que ces peptides sont présents dans ces échantillons, ce qui indique le potentiel de diagnostic asymptomatique de COVID-19 chez les patients.

Comme résultat, l'analyse par spectrométrie de masse a révélé que dans les 20 échantillons confirmés négatifs par PCR, les peptides n'ont pas été détectés. Sur les 63 échantillons positifs, 57 ont été trouvés positifs grâce à la méthode LC-MS/MS développée par cette équipe. Les données indiquent donc une grande sensibilité de la méthode MS employée et peuvent être utilisées pour le diagnostic initial, y compris le diagnostic des cas asymptomatiques.

Bien que la RT-PCR soit sans aucun doute utile pour la détection du SARS-CoV-2, la technologie de spectrométrie de masse est une alternative prometteuse étant donné sa rapidité, sa simplicité et son faible coût, ainsi que la disponibilité d'équipements et d'expertise dans de nombreux laboratoires des pays en développement.

CONCLUSION

La protéomique est devenue l'une des technologies de pointe à la disposition des chercheurs de l'ère post-génomique, en raison du rôle crucial et complexe des protéines et des interactions protéine-protéine dans la fonction cellulaire. Les protéines étant plus proches de la fonction biologique que l'ADN, l'accent a été mis sur le développement de nouveaux outils de protéomique.

L'objectif de ce travail était de fournir une compréhension et une connaissance de base de la protéomique, y compris les méthodes et les applications qui relèvent de cette discipline. En comprenant les structures et les fonctions de toutes les protéines de l'organisme, la protéomique permettra de mieux comprendre les mécanismes des maladies, ce qui se traduira par une amélioration des soins aux patients.

Des progrès très importants et extrêmement utiles ont été réalisés dans les méthodes de protéomique ascendante et descendante. Ces méthodes, qui offrent une plus grande sensibilité, une plus grande rapidité et une plus grande couverture du protéome, sont souvent présentées avec l'affirmation que ces méthodes sont appelées à remplacer les méthodes plus traditionnelles, notamment en utilisant des gels bidimensionnels. En outre, la combinaison de ces technologies a permis de purifier, d'analyser, de caractériser, de quantifier, d'analyser la séquence et la structure et de procéder à des analyses bio-informatiques d'un grand nombre de protéines dans tous les types d'organismes. Tous les domaines liés aux sciences biologiques ont bénéficié de l'utilisation croissante des techniques protéomiques. Les techniques de validation telles que la MS ciblée et les techniques quantitatives utilisant des approches isotopiques (SILAC...) ont fait énormément progresser les applications de recherche en protéomique.

Grâce à la protéomique, les recherches permettront d'identifier de nouvelles protéines, des voies de signalisation cellulaire, des fonctions et des structures dans les organismes. Grâce à ces découvertes, les changements protéiques entre les états sains et les états pathologiques peuvent être clarifiés et utilisés pour développer des plateformes de diagnostic pratiques et applicables dans le cadre clinique.

Ainsi, les progrès de la protéomique auront un impact non seulement sur les activités universitaires, mais aussi sur la recherche et le développement dans le domaine pharmaceutique, biotechnologique et diagnostique.

Pour conclure, d'autres études et recherches seront nécessaires afin de fournir une base de données protéiques complète et d'élargir davantage notre compréhension totale sur la dynamique protéique du corps humain et aussi des microorganismes pathogènes.

RESUMES

RESUME

Titre : La protéomique : généralités, méthodes et applications

Auteur : BAGACHOUL Younesse

Rapporteur : Pr. IBRAHIMI Azeddine

Mots clés : Protéine, protéomique, protéome, spectrométrie de masse, électrophorèse.

Les gènes sont l'unité de l'hérédité, mais ils ne prennent vie que lorsqu'ils sont traduits en protéines - les principaux acteurs fonctionnels de la biologie. Le protéome désigne le nombre total de protéines présentes dans une cellule, un tissu ou un organisme à un moment donné. Cette population de protéines peut varier dans le temps, entre les conditions de croissance et entre les types de cellules, en raison de différences dans l'expression des gènes. La protéomique est la science des protéines et peptides qui sont des molécules biologiques d'immense importance car elles permettent l'exécution de plusieurs rôles structuraux et fonctionnels, de la respiration jusqu'à la prolifération cellulaire et la réparation du génome. L'étude extensive des protéines n'a pas connu une avancée significative jusqu'au début des années 80, à la suite de l'introduction de la spectrométrie de masse MALDI-TOF et ESI. Le développement de flux de travail de la protéomique à haut débit et de la protéomique quantitative au cours des deux dernières décennies a élargi le champ de nos connaissances sur la structure, la fonction, la modification et la dynamique globale des protéines de différents organismes eucaryotes et procaryotes. Les technologies développées sont sensibles, rapides, et permettent une plus grande couverture du protéome, et la combinaison de ces technologies a permis de purifier, d'analyser, de caractériser, de quantifier, d'analyser la séquence et la structure et de procéder à des analyses bioinformatiques d'un grand nombre de protéines.

Ce travail donne une vue globale sur les bases, les outils, et les méthodes conventionnelles et récentes utilisées pour l'étude protéomique des échantillons et préparations biologiques et leurs limitations, et ainsi que ses champs d'application.

ABSTRACT

Title : Proteomics : generalities, methods and applications

Author : BAGACHOUL Younesse

Thesis director: Pr. IBRAHIMI Azeddine

Keywords : protein, proteomics, proteome, mass spectrometry, electrophoresis.

Genes are the unit of heredity, but they only come to life when they are translated into proteins - the main functional players in biology. The proteome refers to the total number of proteins present in a cell, tissue or organism at a given time. This population of proteins can vary over time, between growth conditions and between cell types, due to differences in gene expression. Proteomics is the science of proteins and peptides, which are biological molecules of immense importance because they enable the performance of many structural and functional roles, from respiration to cell proliferation and genome repair. The extensive study of proteins did not advance significantly until the early 1980s, following the introduction of MALDI-TOF and ESI mass spectrometry. The development of high-throughput proteomics and quantitative proteomics workflows over the past two decades has broadened the scope of our knowledge of the structure, function, modification, and overall dynamics of proteins from various eukaryotic and prokaryotic organisms. The technologies developed are sensitive, rapid, and provide greater coverage of the proteome, and the combination of these technologies has enabled purification, analysis, characterization, quantification, sequence and structure analysis, and bioinformatics analysis of a large number of proteins.

This work provides an overview of the basics, tools, and conventional and recent methods used for the proteomic study of biological samples and their limitations, as well as their fields of application.

ملخص

العنوان : البروتيوميّات : عموميّات، طرق و تطبيقات

المؤلف : يونس بكشول

المشرف : البروفيسور إبراهيمي عز الدين

الكلمات الأساسية : بروتين، بروتينيّات، بروتيوم، مطيافية الكتلة، الترحيل الكهربائي

الجينات تعتبر وحدة الوراثة، لكنها لا ترى الحياة سوى عندما تترجم إلى بروتينات - اللاعبين الوظيفيين الرئيسيين في علم الأحياء. يشير البروتيوم إلى العدد الإجمالي للبروتينات الموجودة في خلية أو نسيج أو كائن حي في وقت معين. يمكن أن تختلف مجموعة البروتينات هذه بمرور الوقت، بين ظروف النمو وبين أنواع الخلايا، بسبب الاختلافات في التعبير الجيني. البروتيوميّات هي علم البروتينات والببتيدات، وهي جزيئات بيولوجية ذات أهمية هائلة لأنها تمكن من أداء العديد من الأدوار الهيكلية والوظيفية، من التنفس إلى تكاثر الخلايا وإصلاح الجينوم. لم تتقدم الدراسة المكثفة للبروتينات بشكل كبير حتى أوائل الثمانينات، بعد إدخال تقنية MALDI-TOF و ESI وقد وسع تطوير البروتيوميّات عالية الإنتاجية وسير عمل البروتيوميّات الكمية على مدى العقدين الماضيين نطاق معرفتنا بالبنية والوظيفة والتعديل والديناميكيات العامة للبروتينات من مختلف الكائنات حقيقية النواة وبدائيات النواة. التقنيات المتقدمة حساسة وسريعة وتوفر تغطية أكبر للبروتيوم، وقد مكن الجمع بين هذه التقنيات من التنقية والتحليل والتوصيف والتقدير الكمي وتحليل التسلسل والبنية وتحليل البيانات المعلوماتية الحيوية لعدد كبير من البروتينات.

يقدم هذا العمل لمحة عامة عن الأساسيات والأدوات والأساليب التقليدية والحديثة المستخدمة في الدراسة البروتينية للعينات البيولوجية وقبورها، وكذلك مجالات تطبيقها.

REFERENCES

- [1] Adkins, J.N., et al., *Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry*. Mol Cell Proteomics, 2002. **1**(12): p. 947-55.
- [2] Baudin, B., *Protéomique et spectrométrie de masse*. Revue Francophone des Laboratoires, 2011. **2011**(437): p. 31-40.
- [3] Yu, L., K. Li, and X. Zhang, *Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review*. Oncotarget, 2017. **8**(70): p. 115774-115786.
- [4] de Oliveira Veras, A.A., et al., *Chapter 12 - Computational Techniques in Data Integration and Big Data Handling in Omics*, in *Omics Technologies and Bio-Engineering*, D. Barh and V. Azevedo, Editors. 2018, Academic Press. p. 209-222.
- [5] Heo, L. and M. Feig, *High-accuracy protein structures by combining machine-learning with physics-based refinement*. Proteins, 2020. **88**(5): p. 637-642.
- [6] Vitzthum, F., et al., *Proteomics: from basic research to diagnostic application. A review of requirements & needs*. J Proteome Res, 2005. **4**(4): p. 1086-97.
- [7] Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. **14**(3): p. 206-214.
- [8] Villoutreix, B.O., et al., *A leap into the chemical space of protein-protein interaction inhibitors*. Curr Pharm Des, 2012. **18**(30): p. 4648-67.
- [9] Al Feteisi, H., et al., *Choice of LC-MS methods for the absolute quantification of drug-metabolizing enzymes and transporters in human tissue: a comparative cost analysis*. AAPS J, 2015. **17**(2): p. 438-46.
- [10] D.C., L., *Overview of Analytical Proteomics*. In: *Introduction to Proteomics*. Humana Press, Totowa, NJ. 2002.
- [11] Will, C.L. and R. Luhrmann, *Spliceosome structure and function*. Cold Spring Harb Perspect Biol, 2011. **3**(7).

- [12] Wang, Y., et al., *Mechanism of alternative splicing and its regulation*. Biomed Rep, 2015. **3**(2): p. 152-158.
- [13] Ben-Shem, A., et al., *The structure of the eukaryotic ribosome at 3.0 Å resolution*. Science, 2011. **334**(6062): p. 1524-9.
- [14] Litwack, G., *Chapter 11 - Protein Biosynthesis*, in *Human Biochemistry*, G. Litwack, Editor. 2018, Academic Press: Boston. p. 319-336.
- [15] Veuthey, A.L., et al., *Application of text-mining for updating protein post-translational modification annotation in UniProtKB*. BMC Bioinformatics, 2013. **14**: p. 104.
- [16] Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity*. Nat Methods, 2013. **10**(3): p. 186-7.
- [17] Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks*. Cell, 2006. **127**(3): p. 635-48.
- [18] Hanke, S. and M. Mann, *The phosphotyrosine interactome of the insulin receptor family and its substrates IRS-1 and IRS-2*. Mol Cell Proteomics, 2009. **8**(3): p. 519-34.
- [19] Huang, S. and M.P. Czech, *The GLUT4 glucose transporter*. Cell Metab, 2007. **5**(4): p. 237-52.
- [20] Maqbool, M. and N. Hoda, *GSK3 Inhibitors in the Therapeutic Development of Diabetes, Cancer and Neurodegeneration: Past, Present and Future*. Curr Pharm Des, 2017. **23**(29): p. 4332-4350.
- [21] Rassy, E., R. Flippot, and L. Albiges, *Tyrosine kinase inhibitors and immunotherapy combinations in renal cell carcinoma*. Ther Adv Med Oncol, 2020. **12**: p. 1758835920907504.
- [22] Houshmand, M., et al., *Chronic myeloid leukemia stem cells*. Leukemia, 2019. **33**(7): p. 1543-1556.

- [23] Greuber, E.K., et al., *Role of ABL family kinases in cancer: from leukaemia to solid tumours*. Nat Rev Cancer, 2013. **13**(8): p. 559-71.
- [24] Zhu, L., et al., *Ebola virus replication is regulated by the phosphorylation of viral protein VP35*. Biochem Biophys Res Commun, 2020. **521**(3): p. 687-692.
- [25] Yoo, B.H., et al., *CKAP2 phosphorylation by CDK1/cyclinB1 is crucial for maintaining centrosome integrity*. Exp Mol Med, 2017. **49**(7): p. e354.
- [26] Icardi, L., K. De Bosscher, and J. Tavernier, *The HAT/HDAC interplay: multilevel control of STAT signaling*. Cytokine Growth Factor Rev, 2012. **23**(6): p. 283-91.
- [27] Reed, S.M. and D.E. Quelle, *p53 Acetylation: Regulation and Consequences*. Cancers (Basel), 2014. **7**(1): p. 30-69.
- [28] Akhtar, A. and P.B. Becker, *Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila*. Mol Cell, 2000. **5**(2): p. 367-75.
- [29] Aksnes, H., R. Ree, and T. Arnesen, *Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases*. Mol Cell, 2019. **73**(6): p. 1097-1114.
- [30] Ree, R., S. Varland, and T. Arnesen, *Spotlight on protein N-terminal acetylation*. Exp Mol Med, 2018. **50**(7): p. 1-13.
- [31] Nakayasu, E.S., et al., *Ancient Regulatory Role of Lysine Acetylation in Central Metabolism*. mBio, 2017. **8**(6).
- [32] Ouidir, T., et al., *Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14*. J Proteomics, 2015. **114**: p. 214-25.
- [33] Tanaka, S., et al., *Structural basis for the broad substrate specificity of the human tyrosylprotein sulfotransferase-1*. Sci Rep, 2017. **7**(1): p. 8776.

- [34] de Graffenried, C.L. and C.R. Bertozzi, *Golgi localization of carbohydrate sulfotransferases is a determinant of L-selectin ligand biosynthesis*. J Biol Chem, 2003. **278**(41): p. 40282-95.
- [35] Moore, K.L., *The biology and enzymology of protein tyrosine O-sulfation*. J Biol Chem, 2003. **278**(27): p. 24243-6.
- [36] Mueller, J.W. and N. Shafqat, *Adenosine-5'-phosphosulfate--a multifaceted modulator of bifunctional 3'-phospho-adenosine-5'-phosphosulfate synthases and related enzymes*. FEBS J, 2013. **280**(13): p. 3050-7.
- [37] Westmuckett, A.D., et al., *Early postnatal pulmonary failure and primary hypothyroidism in mice with combined TPST-1 and TPST-2 deficiency*. Gen Comp Endocrinol, 2008. **156**(1): p. 145-53.
- [38] Seibert, C. and T.P. Sakmar, *Toward a framework for sulfoproteomics: Synthesis and characterization of sulfotyrosine-containing peptides*. Biopolymers, 2008. **90**(3): p. 459-77.
- [39] Dorfman, T., et al., *A tyrosine-sulfated peptide derived from the heavy-chain CDR3 region of an HIV-1-neutralizing antibody binds gp120 and inhibits HIV-1 infection*. J Biol Chem, 2006. **281**(39): p. 28529-35.
- [40] Kim, J.H., et al., *The Role of Protein Arginine Methyltransferases in Inflammatory Responses*. Mediators Inflamm, 2016. **2016**: p. 4028353.
- [41] Li, L., I. Vorobyov, and T.W. Allen, *The different interactions of lysine and arginine side chains with lipid membranes*. J Phys Chem B, 2013. **117**(40): p. 11906-20.
- [42] Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function*. Mol Cell, 2005. **18**(3): p. 263-72.
- [43] Yang, Y. and M.T. Bedford, *Protein arginine methyltransferases and cancer*. Nat Rev Cancer, 2013. **13**(1): p. 37-50.

- [44] Tewary, S.K., Y.G. Zheng, and M.C. Ho, *Protein arginine methyltransferases: insights [into the enzyme structure and mechanism at the atomic level]*. Cell Mol Life Sci, 2019. **76**(15): p. 2917-2932.
- [45] Tang, J., et al., *PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells*. J Biol Chem, 2000. **275**(11): p. 7723-30.
- [46] Castillo, J., G. Lopez-Rodas, and L. Franco, *Histone Post-Translational Modifications and Nucleosome Organisation in Transcriptional Regulation: Some Open Questions*. Adv Exp Med Biol, 2017. **966**: p. 65-92.
- [47] Davidson, T., Q. Ke, and M. Costa, *Chapter 9 - Selected Molecular Mechanisms of Metal Toxicity and Carcinogenicity*, in *Handbook on the Toxicology of Metals (Fourth Edition)*, G.F. Nordberg, B.A. Fowler, and M. Nordberg, Editors. 2015, Academic Press: San Diego. p. 173-196.
- [48] Perche, P.-Y., et al., *Différenciation du nucléosome: le rôle des variants de l'histone H2A*. M/S : médecine sciences, 2003. **19**(11): p. 1137-1145.
- [49] Guengerich, F.P., *Introduction: Metals in Biology: alpha-Ketoglutarate/Iron-Dependent Dioxygenases*. J Biol Chem, 2015. **290**(34): p. 20700-20701.
- [50] Lee, C.C., C.Y. Wu, and H.Y. Yang, *Discoveries of how cells sense oxygen win the 2019 Nobel Prize in Physiology or medicine*. Biomed J, 2020. **43**(5): p. 434-437.
- [51] Schofield, C.J. and P.J. Ratcliffe, *Oxygen sensing by HIF hydroxylases*. Nat Rev Mol Cell Biol, 2004. **5**(5): p. 343-54.
- [52] Fan, L., et al., *The hypoxia-inducible factor pathway, prolyl hydroxylase domain protein inhibitors, and their roles in bone repair and regeneration*. Biomed Res Int, 2014. **2014**: p. 239356.
- [53] Berchner-Pfannschmidt, U., et al., *Imaging of the hypoxia-inducible factor pathway: insights into oxygen sensing*. Eur Respir J, 2008. **32**(1): p. 210-7.

- [54] Perrin, B.J. and J.M. Ervasti, *The actin gene family: function follows isoform*. Cytoskeleton (Hoboken), 2010. **67**(10): p. 630-4.
- [55] Luo, W., et al., *PHD3-mediated prolyl hydroxylation of nonmuscle actin impairs polymerization and cell motility*. Mol Biol Cell, 2014. **25**(18): p. 2788-96.
- [56] Zurlo, G., et al., *New Insights into Protein Hydroxylation and Its Important Role in Human Diseases*. Biochimica et biophysica acta, 2016. **1866**(2): p. 208-220.
- [57] Welsh, K.J., M.S. Kirkman, and D.B. Sacks, *Role of Glycated Proteins in the Diagnosis and Management of Diabetes: Research Gaps and Future Directions*. Diabetes Care, 2016. **39**(8): p. 1299-306.
- [58] Shental-Bechor, D. and Y. Levy, *Folding of glycoproteins: toward understanding the biophysics of the glycosylation code*. Curr Opin Struct Biol, 2009. **19**(5): p. 524-33.
- [59] Shental-Bechor, D. and Y. Levy, *Effect of glycosylation on protein folding: a close look at thermodynamic stabilization*. Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8256-61.
- [60] Martinez, M.R., et al., *Stress-induced O-GlcNAcylation: an adaptive process of injured cells*. Biochemical Society transactions, 2017. **45**(1): p. 237-249.
- [61] de Virgilio, M., et al., *Degradation of a short-lived glycoprotein from the lumen of the endoplasmic reticulum: the role of N-linked glycans and the unfolded protein response*. Mol Biol Cell, 1999. **10**(12): p. 4059-73.
- [62] Corfield, A., *Eukaryotic protein glycosylation: a primer for histochemists and cell biologists*. Histochem Cell Biol, 2017. **147**(2): p. 119-147.
- [63] Mu, Y., et al., *O-linked beta-N-acetylglucosamine transferase plays an essential role in heart development through regulating angiopoietin-1*. PLoS Genet, 2020. **16**(4): p. e1008730.

- [64] Tong, L., *Structure and function of biotin-dependent carboxylases*. Cell Mol Life Sci, 2013. **70**(5): p. 863-91.
- [65] Lombard, J. and D. Moreira, *Early evolution of the biotin-dependent carboxylase family*. BMC Evol Biol, 2011. **11**: p. 232.
- [66] Donti, T.R., P.R. Blackburn, and P.S. Atwal, *Holocarboxylase synthetase deficiency pre and post newborn screening*. Mol Genet Metab Rep, 2016. **7**: p. 40-4.
- [67] Resh, M.D., *Chapter 13 - Lipid Modification of Proteins*, in *Biochemistry of Lipids, Lipoproteins and Membranes (Sixth Edition)*, N.D. Ridgway and R.S. McLeod, Editors. 2016, Elsevier: Boston. p. 391-414.
- [68] Hooff, G.P., et al., *Isoprenoids, small GTPases and Alzheimer's disease*. Biochim Biophys Acta, 2010. **1801**(8): p. 896-905.
- [69] Kuchay, S., et al., *GGTase3 is a newly identified geranylgeranyltransferase targeting a ubiquitin ligase*. Nat Struct Mol Biol, 2019. **26**(7): p. 628-636.
- [70] Perez-Sala, D., *Protein isoprenylation in biology and disease: general overview and perspectives from studies with genetically engineered animals*. Front Biosci, 2007. **12**: p. 4456-72.
- [71] Gomes, A.Q., et al., *Membrane Targeting of Rab GTPases Is Influenced by the Prenylation Motif*. Molecular Biology of the Cell, 2003. **14**(5): p. 1882-1899.
- [72] Pylypenko, O., et al., *Structure of Rab escort protein-1 in complex with Rab geranylgeranyltransferase*. Mol Cell, 2003. **11**(2): p. 483-94.
- [73] Wilkins, J.A., et al., *Characterization of Prenylated C-terminal Peptides Using a Thiopropyl-based Capture Technique and LC-MS/MS* ^{*}. Molecular & Cellular Proteomics, 2020. **19**(6): p. 1005-1016.

- [74] Gilardi, M., et al., *Tipifarnib as a Precision Therapy for HRAS-Mutant Head and Neck Squamous Cell Carcinomas*. Molecular Cancer Therapeutics, 2020. **19**(9): p. 1784.
- [75] Maurer-Stroh, S., S. Washietl, and F. Eisenhaber, *Protein prenyltransferases*. Genome Biol, 2003. **4**(4): p. 212.
- [76] Schrader, E.K., K.G. Harstad, and A. Matouschek, *Targeting proteins for degradation*. Nat Chem Biol, 2009. **5**(11): p. 815-22.
- [77] Iconomou, M. and D.N. Saunders, *Systematic approaches to identify E3 ligase substrates*. Biochem J, 2016. **473**(22): p. 4083-4101.
- [78] Bhogaraju, S. and I. Dikic, *Cell biology: Ubiquitination without E1 and E2 enzymes*. Nature, 2016. **533**(7601): p. 43-4.
- [79] Balaji, V. and T. Hoppe, *Regulation of E3 ubiquitin ligases by homotypic and heterotypic assembly*. F1000Res, 2020. **9**.
- [80] Deol, K.K., S. Lorenz, and E.R. Strieter, *Enzymatic Logic of Ubiquitin Chain Assembly*. Front Physiol, 2019. **10**: p. 835.
- [81] Nakagawa, T. and K. Nakayama, *Protein monoubiquitylation: targets and diverse functions*. Genes Cells, 2015. **20**(7): p. 543-62.
- [82] Swatek, K.N. and D. Komander, *Ubiquitin modifications*. Cell Res, 2016. **26**(4): p. 399-422.
- [83] Dikic, I., S. Wakatsuki, and K.J. Walters, *Ubiquitin-binding domains - from structures to functions*. Nat Rev Mol Cell Biol, 2009. **10**(10): p. 659-71.
- [84] Haglund, K. and I. Dikic, *Ubiquitylation and cell signaling*. EMBO J, 2005. **24**(19): p. 3353-9.
- [85] Haakonsen, D.L. and M. Rape, *Branching Out: Improved Signaling by Heterotypic Ubiquitin Chains*. Trends Cell Biol, 2019. **29**(9): p. 704-716.

- [86] Ohtake, F., et al., *Ubiquitin acetylation inhibits polyubiquitin chain elongation*. EMBO Rep, 2015. **16**(2): p. 192-201.
- [87] Stolz, A. and I. Dikic, *Heterotypic Ubiquitin Chains: Seeing is Believing*. Trends Cell Biol, 2018. **28**(1): p. 1-3.
- [88] Randles, L. and K.J. Walters, *Ubiquitin and its binding domains*. Front Biosci (Landmark Ed), 2012. **17**: p. 2140-57.
- [89] Hospenhal, M.K., S.M. Freund, and D. Komander, *Assembly, analysis and architecture of atypical ubiquitin chains*. Nat Struct Mol Biol, 2013. **20**(5): p. 555-65.
- [90] Cajee, U.F., R. Hull, and M. Ntwasa, *Modification by ubiquitin-like proteins: significance in apoptosis and autophagy pathways*. Int J Mol Sci, 2012. **13**(9): p. 11804-31.
- [91] Lorincz, P. and G. Juhasz, *Autophagosome-Lysosome Fusion*. J Mol Biol, 2020. **432**(8): p. 2462-2482.
- [92] Iwai, K., *Diverse roles of the ubiquitin system in NF- κ B activation*. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2014. **1843**(1): p. 129-136.
- [93] Poltronieri, P. and N. Čerekovic, *Roles of Nicotinamide Adenine Dinucleotide (NAD⁺) in Biological Systems*. Challenges, 2018. **9**(1).
- [94] Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer*. Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.
- [95] Mayne, J., et al., *Bottom-Up Proteomics (2013-2015): Keeping up in the Era of Systems Biology*. Anal Chem, 2016. **88**(1): p. 95-121.
- [96] Paulo, J.A., *Practical and Efficient Searching in Proteomics: A Cross Engine Comparison*. Webmedcentral, 2013. **4**(10).

- [97] Schrader, C.U., et al., *Simultaneous Proteoform Analysis of Histones H3 and H4 with a Simplified Middle-Down Proteomics Method*. Anal Chem, 2018. **90**(5): p. 3083-3090.
- [98] Virág, D., et al., *Current Trends in the Analysis of Post-translational Modifications*. Chromatographia, 2020. **83**(1): p. 1-10.
- [99] Diaz Marin, R., et al., *RELi protocol: Optimization for protein extraction from white, brown and beige adipose tissues*. MethodsX, 2019. **6**: p. 918-928.
- [100] Verollet, R., *A major step towards efficient sample preparation with bead-beating*. Biotechniques, 2008. **44**(6): p. 832-3.
- [101] Behjati, S. and P.S. Tarpey, *What is next generation sequencing?* Archives of disease in childhood. Education and practice edition, 2013. **98**(6): p. 236-238.
- [102] Gagaoua, M. and K. Hafid, *Three Phase Partitioning System, an Emerging Non-Chromatographic Tool for Proteolytic Enzymes Recovery and Purification*. 2016. **5**(1): p. 100134.
- [103] Friedman, D.B., S. Hoving, and R. Westermeier, *Isoelectric focusing and two-dimensional gel electrophoresis*. Methods Enzymol, 2009. **463**: p. 515-40.
- [104] Magdeldin, S., et al., *Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis*. Clin Proteomics, 2014. **11**(1): p. 16.
- [105] Bunai, K. and K. Yamane, *Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives*. J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **815**(1-2): p. 227-36.
- [106] Zong, C., et al., *Two-dimensional electrophoresis-based characterization of post-translational modifications of mammalian 20S proteasome complexes*. Proteomics, 2008. **8**(23-24): p. 5025-37.

- [107] Zheng, S., et al., *Monolithic column HPLC separation of intact proteins analyzed by LC-MALDI using on-plate digestion: An approach to integrate protein separation and identification*. Anal Chem, 2006. **78**(14): p. 5198-204.
- [108] Swartz, M.E., *UPLC™: An Introduction and Review*. Journal of Liquid Chromatography & Related Technologies, 2005. **28**(7-8): p. 1253-1263.
- [109] Emadali, A. and M. Gallagher-Gambarelli, *[Quantitative proteomics by SILAC: practicalities and perspectives for an evolving approach]*. Med Sci (Paris), 2009. **25**(10): p. 835-42.
- [110] Tang, K., et al., *Electrospray Ionization in Mass Spectrometry*, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, J.C. Lindon, G.E. Tranter, and D.W. Koppenaal, Editors. 2017, Academic Press: Oxford. p. 476-481.
- [111] Pomastowski, P. and B. Buszewski, *Complementarity of Matrix- and Nanostructure-Assisted Laser Desorption/Ionization Approaches*. Nanomaterials (Basel), 2019. **9**(2).
- [112] Pomastowski, P. and B. Buszewski, *Two-dimensional gel electrophoresis in the light of new developments*. TrAC Trends in Analytical Chemistry, 2014. **53**: p. 167-177.
- [113] Reddy, G. and E.A. Dalmasso, *SELDI ProteinChip(R) Array Technology: Protein-Based Predictive Medicine and Drug Discovery Applications*. J Biomed Biotechnol, 2003. **2003**(4): p. 237-241.
- [114] Canas, B., et al., *Mass spectrometry technologies for proteomics*. Brief Funct Genomic Proteomic, 2006. **4**(4): p. 295-320.
- [115] El-Aneed, A., A. Cohen, and J. Banoub, *Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers*. Applied Spectroscopy Reviews, 2009. **44**(3): p. 210-230.

- [116] Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal Chem, 2000. **72**(6): p. 1156-62.
- [117] Hecht, E.S., et al., *Fundamentals and Advances of Orbitrap Mass Spectrometry*. Encyclopedia of Analytical Chemistry, 2019: p. 1-40.
- [118] Brion, C., et al., *Orbitrap mass analyser for in situ characterisation of planetary environments: Performance evaluation of a laboratory prototype*. Planetary and Space Science, 2016. **131**: p. 33-45.
- [119] Savaryn, J.P., T.K. Toby, and N.L. Kelleher, *A researcher's guide to mass spectrometry-based proteomics*. Proteomics, 2016. **16**(18): p. 2435-43.
- [120] Kyle, P.B., *Chapter 7 - Toxicology: GCMS*, in *Mass Spectrometry for the Clinical Laboratory*, H. Nair and W. Clarke, Editors. 2017, Academic Press: San Diego. p. 131-163.
- [121] Rockwood, A.L., M.M. Kushnir, and N.J. Clarke, *2 - Mass Spectrometry*, in *Principles and Applications of Clinical Mass Spectrometry*, N. Rifai, A.R. Horvath, and C.T. Wittwer, Editors. 2018, Elsevier. p. 33-65.
- [122] Berg JM, T.J., Stryer L. , *Biochemistry*. 5th edition. New York: W H Freeman; 2002. Section 4.1, *The Purification of Proteins Is an Essential First Step in Understanding Their Function*.
- [123] Syed, S.U., et al., *Quadrupole mass filter: design and performance for operation in stability zone 3*. J Am Soc Mass Spectrom, 2013. **24**(10): p. 1493-500.
- [124] Cunsolo, V., et al., *Mass spectrometry in food proteomics: a tutorial*. J Mass Spectrom, 2014. **49**(9): p. 768-84.
- [125] Parker CE, Warren MR, Mocanu V. *Mass Spectrometry for Proteomics*. In: Alzate O, editor. *Neuroproteomics*. Boca Raton (FL): CRC Press/Taylor & Francis; 2010. Chapter 5. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK56011/>.

- [126] Stumpf, M.P., et al., *Estimating the size of the human interactome*. Proc Natl Acad Sci U S A, 2008. **105**(19): p. 6959-64.
- [127] Rolland, T., et al., *A proteome-scale map of the human interactome network*. Cell, 2014. **159**(5): p. 1212-1226.
- [128] Wetlaufer, D.B., *Nucleation, rapid folding, and globular intrachain regions in proteins*. Proc Natl Acad Sci U S A, 1973. **70**(3): p. 697-701.
- [129] De Las Rivas, J. and C. Fontanillo, *Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell*. Brief Funct Genomics, 2012. **11**(6): p. 489-96.
- [130] Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing*. Nature, 2020. **583**(7816): p. 459-468.
- [131] Tiwari, V., et al., *Discovering small-molecule therapeutics against SARS-CoV-2*. Drug Discov Today, 2020. **25**(8): p. 1535-1544.
- [132] Hermine, O., J.C. Ramos, and K. Tobinai, *A Review of New Findings in Adult T-cell Leukemia-Lymphoma: A Focus on Current and Emerging Treatment Strategies*. Adv Ther, 2018. **35**(2): p. 135-152.
- [133] Boxus, M., et al., *The HTLV-1 Tax interactome*. Retrovirology, 2008. **5**: p. 76.
- [134] De Las Rivas, J. and C. Fontanillo, *Protein-protein interactions essentials: key concepts to building and analyzing interactome networks*. PLoS Comput Biol, 2010. **6**(6): p. e1000807.
- [135] Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions*. Nature, 1989. **340**(6230): p. 245-6.
- [136] Peterson, T.A., M.A. Stamnes, and R.C. Piper, *A Yeast 2-Hybrid Screen in Batch to Compare Protein Interactions*. J Vis Exp, 2018(136).

- [137] Lin, J.S. and E.M. Lai, *Protein-Protein Interactions: Yeast Two-Hybrid System*. Methods Mol Biol, 2017. **1615**: p. 177-187.
- [138] Veyron-Churlet, R. and C. Loch, *In Vivo Methods to Study Protein-Protein Interactions as Key Players in Mycobacterium Tuberculosis Virulence*. Pathogens, 2019. **8**(4).
- [139] Xing, S., et al., *Techniques for the Analysis of Protein-Protein Interactions in Vivo*. Plant Physiol, 2016. **171**(2): p. 727-58.
- [140] Huang, X. and J. Wang, *The extended pluripotency protein interactome and its links to reprogramming*. Curr Opin Genet Dev, 2014. **28**: p. 16-24.
- [141] Gloeckner, C.J., et al., *Tandem affinity purification of protein complexes from mammalian cells by the Strep/FLAG (SF)-TAP tag*. Methods Mol Biol, 2009. **564**: p. 359-72.
- [142] Li, Y., *The tandem affinity purification technology: an overview*. Biotechnol Lett, 2011. **33**(8): p. 1487-99.
- [143] Oeffinger, M., *Two steps forward--one step back: advances in affinity purification mass spectrometry of macromolecular complexes*. Proteomics, 2012. **12**(10): p. 1591-608.
- [144] Roux, K.J., et al., *A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells*. J Cell Biol, 2012. **196**(6): p. 801-10.
- [145] Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID*. Nat Biotechnol, 2018. **36**(9): p. 880-887.
- [146] De Munter, S., et al., *Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions*. FEBS Lett, 2017. **591**(2): p. 415-424.
- [147] Schopp, I.M., et al., *Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes*. Nat Commun, 2017. **8**: p. 15690.

- [148] Beganton, B., et al., *[New approaches for protein-protein interaction study]*. Med Sci (Paris), 2019. **35**(3): p. 223-231.
- [149] Zhang, C. and Y. Liu, *Retrieving Quantitative Information of Histone PTMs by Mass Spectrometry*. Methods Enzymol, 2017. **586**: p. 165-191.
- [150] Arike, L. and L. Peil, *Spectral counting label-free proteomics*. Methods Mol Biol, 2014. **1156**: p. 213-22.
- [151] Pappireddi, N., L. Martin, and M. Wuhr, *A Review on Quantitative Multiplexed Proteomics*. Chembiochem, 2019. **20**(10): p. 1210-1224.
- [152] Hoedt, E., G. Zhang, and T.A. Neubert, *Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) for Quantitative Proteomics*. Adv Exp Med Biol, 2019. **1140**: p. 531-539.
- [153] McMillan, L.J., et al., *Multiplex quantitative SILAC for analysis of archaeal proteomes: a case study of oxidative stress responses*. Environ Microbiol, 2018. **20**(1): p. 385-401.
- [154] Sirvent, A., S. Urbach, and S. Roche, *[Analysis of oncogenic signaling induced by tyrosine kinases in tumors by SILAC-based quantitative proteomic approach]*. Med Sci (Paris), 2014. **30**(5): p. 558-66.
- [155] Wang, Z., et al., *High-throughput and Deep-proteome Profiling by 16-plex Tandem Mass Tag Labeling Coupled with Two-dimensional Chromatography and Mass Spectrometry*. J Vis Exp, 2020(162).
- [156] Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags*. Nat Biotechnol, 1999. **17**(10): p. 994-9.
- [157] Chen, X., et al., *Quantitative proteomics using SILAC: Principles, applications, and developments*. Proteomics, 2015. **15**(18): p. 3175-92.
- [158] Voges, R., et al., *Absolute quantification of Corynebacterium glutamicum glycolytic and anaplerotic enzymes by QconCAT*. J Proteomics, 2015. **113**: p. 366-77.

- [159] Brownridge, P., et al., *Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy*. Proteomics, 2011. **11**(15): p. 2957-70.
- [160] Narumi, R., et al., *Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics*. Synth Syst Biotechnol, 2018. **3**(2): p. 97-104.
- [161] Singh, S., et al., *A practical guide to the FLEXIQuant method*. Methods Mol Biol, 2012. **893**: p. 295-319.
- [162] Kito, K. and T. Ito, *Mass spectrometry-based approaches toward absolute quantitative proteomics*. Curr Genomics, 2008. **9**(4): p. 263-74.
- [163] Monti, M., et al., *Functional proteomics*. Clinica Chimica Acta, 2005. **357**(2): p. 140-150.
- [164] Strimbu, K. and J.A. Tavel, *What are biomarkers?* Curr Opin HIV AIDS, 2010. **5**(6): p. 463-6.
- [165] Chan, P.P., V.C. Wasinger, and R.W. Leong, *Current application of proteomics in biomarker discovery for inflammatory bowel disease*. World J Gastrointest Pathophysiol, 2016. **7**(1): p. 27-37.
- [166] Amiri-Dashatan, N., et al., *Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry*. Iranian journal of pharmaceutical research : IJPR, 2018. **17**(4): p. 1523-1536.
- [167] Romanetto, J., *Intérêts des marqueurs biologiques dans les essais cliniques. Sciences pharmaceutiques*. 2011.
- [168] Kanekiyo, T., H. Xu, and G. Bu, *ApoE and Abeta in Alzheimer's disease: accidental encounters or partners?* Neuron, 2014. **81**(4): p. 740-54.
- [169] Konijnenberg, E., et al., *APOE epsilon4 genotype-dependent cerebrospinal fluid proteomic signatures in Alzheimer's disease*. Alzheimers Res Ther, 2020. **12**(1): p. 65.

- [170] Abdi, F., et al., *Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders*. J Alzheimers Dis, 2006. **9**(3): p. 293-348.
- [171] Zhang, J., et al., *CSF multianalyte profile distinguishes Alzheimer and Parkinson diseases*. Am J Clin Pathol, 2008. **129**(4): p. 526-9.
- [172] Lugert, S., et al., *Glypican-2 levels in cerebrospinal fluid predict the status of adult hippocampal neurogenesis*. Sci Rep, 2017. **7**: p. 46543.
- [173] Geiger, T., et al., *Proteomic portrait of human breast cancer progression identifies novel prognostic markers*. Cancer Res, 2012. **72**(9): p. 2428-39.
- [174] Otu, H.H., et al., *Prediction of Diabetic Nephropathy Using Urine Proteomic Profiling 10 Years Prior to Development of Nephropathy*. Diabetes Care, 2007. **30**(3): p. 638.
- [175] Merchant, M.L., *Mass spectrometry in chronic kidney disease research*. Advances in chronic kidney disease, 2010. **17**(6): p. 455-468.
- [176] Schlatzer, D., et al., *Novel urinary protein biomarkers predicting the development of microalbuminuria and renal function decline in type 1 diabetes*. Diabetes care, 2012. **35**(3): p. 549-555.
- [177] Bhensdadia, N.M., et al., *Urine haptoglobin levels predict early renal functional decline in patients with type 2 diabetes*. Kidney international, 2013. **83**(6): p. 1136-1143.
- [178] Good, D.M., et al., *Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease*. Molecular & cellular proteomics : MCP, 2010. **9**(11): p. 2424-2437.
- [179] Rendon, A. and K. Schakel, *Psoriasis Pathogenesis and Treatment*. Int J Mol Sci, 2019. **20**(6).

- [180] Xie, L., G. Tsapralis, and Q.M. Chen, *Proteomic identification of insulin-like growth factor-binding protein-6 induced by sublethal H₂O₂ stress from human diploid fibroblasts*. Mol Cell Proteomics, 2005. **4**(9): p. 1273-83.
- [181] Karlsson, R., et al., *Proteotyping: Proteomic characterization, classification and identification of microorganisms – A prospectus*. Systematic and Applied Microbiology, 2015. **38**(4): p. 246-257.
- [182] Foster, M.W., et al., *Targeted Proteomics of Human Metapneumovirus in Clinical Samples and Viral Cultures*. Analytical Chemistry, 2015. **87**(20): p. 10247-10254.
- [183] Hecker, M. and U. Völker, *Towards a comprehensive understanding of Bacillus subtilis cell physiology by physiological proteomics*. PROTEOMICS, 2004. **4**(12): p. 3727-3750.
- [184] Peker, N., et al., *A Comparison of Three Different Bioinformatics Analyses of the 16S–23S rRNA Encoding Region for Bacterial Identification*. 2019. **10**(620).
- [185] Chen, B., et al., *Proteomics progresses in microbial physiology and clinical antimicrobial therapy*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2017. **36**(3): p. 403-413.
- [186] Wieser, A., et al., *MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond (mini review)*. Applied Microbiology and Biotechnology, 2012. **93**(3): p. 965-974.
- [187] Luo, L., et al., *Evaluation of the VITEK MS knowledge base version 3.0 for the identification of clinically relevant Mycobacterium species*. Emerging microbes & infections, 2018. **7**(1): p. 114-114.
- [188] Sogawa, K., et al., *Use of the MALDI BioTyper system with MALDI-TOF mass spectrometry for rapid identification of microorganisms*. Analytical and Bioanalytical Chemistry, 2011. **400**(7): p. 1905.

- [189] Woolhouse, M., et al., *Human viruses: discovery and emergence*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1604): p. 2864-2871.
- [190] Mancone, C., et al., *Hepatitis C virus production requires apolipoprotein A-I and affects its association with nascent low-density lipoproteins*. Gut, 2011. **60**(3): p. 378.
- [191] Huang, H., et al., *Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins*. Proceedings of the National Academy of Sciences, 2007. **104**(14): p. 5848.
- [192] Feng, H., M. Wang, and W.N. Chen, *iTRAQ-Coupled 2D LC-MS/MS Analysis of Secreted Proteome of HBV-Replicating HepG2 Cells: Potential in Biomarkers for Prognosis of HCC*. Current Microbiology, 2010. **61**(4): p. 280-284.
- [193] Makanga, M., et al., *Towards a proteomic definition of CoArtem action in Plasmodium falciparum malaria*. Proteomics, 2005. **5**(7): p. 1849-58.
- [194] Zhu, D., et al., *Dose Dependent Dual Effect of Baicalin and Herb Huang Qin Extract on Angiogenesis*. PLoS One, 2016. **11**(11): p. e0167125.
- [195] Li-Weber, M., *New therapeutic aspects of flavones: the anticancer properties of Scutellaria and its main active constituents Wogonin, Baicalein and Baicalin*. Cancer Treat Rev, 2009. **35**(1): p. 57-68.
- [196] Gao, Y., et al., *Anticancer properties of baicalein: a review*. Medicinal Chemistry Research, 2016. **25**(8): p. 1515-1523.
- [197] Yan, X., et al., *Apigenin in cancer therapy: anti-cancer effects and mechanisms of action*. Cell & Bioscience, 2017. **7**(1): p. 50.
- [198] Huang, W.S., et al., *Proteomic analysis of the effects of baicalein on colorectal cancer cells*. Proteomics, 2012. **12**(6): p. 810-9.

- [199] Ye, X., et al., *Chapter 24 - Cell surface protein enrichment for biomarker and drug target discovery using mass spectrometry-based proteomics*, in *Proteomic and Metabolomic Approaches to Biomarker Discovery (Second Edition)*, H.J. Issaq and T.D. Veenstra, Editors. 2020, Academic Press: Boston. p. 409-420.
- [200] Ducret, A., et al., *Identification of six cell surface proteins for specific liver targeting*. *Proteomics Clin Appl*, 2015. **9**(7-8): p. 651-61.
- [201] Suman, S., S. Mishra, and Y. Shukla, *Toxicoproteomics in human health and disease: an update*. *Expert Review of Proteomics*, 2016. **13**(12): p. 1073-1089.
- [202] Colquhoun, D.R., et al., *Global screening of human cord blood proteomes for biomarkers of toxic exposure and effect*. *Environ Health Perspect*, 2009. **117**(5): p. 832-8.
- [203] Hobson, D.J., et al., *Proteomic analysis of ovomucoid hypersensitivity in mice by two-dimensional difference gel electrophoresis (2D-DIGE)*. *Food Chem Toxicol*, 2007. **45**(12): p. 2372-80.
- [204] Bantscheff, M., et al., *Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors*. *Nat Biotechnol*, 2007. **25**(9): p. 1035-44.
- [205] Emerson, D., et al., *Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics*. *BioScience*, 2008. **58**(10): p. 925-936.
- [206] Pavlov, M.Y. and M. Ehrenberg, *Optimal control of gene expression for fast proteome adaptation to environmental change*. *Proceedings of the National Academy of Sciences*, 2013. **110**(51): p. 20527.
- [207] Ortea, I., et al., *Arginine kinase peptide mass fingerprinting as a proteomic approach for species identification and taxonomic analysis of commercially relevant shrimp species*. *J Agric Food Chem*, 2009. **57**(13): p. 5665-72.
- [208] Carrera, M., et al., *Identification of commercial hake and grenadier species by proteomic analysis of the parvalbumin fraction*. *Proteomics*, 2006. **6**(19): p. 5278-87.

- [209] Bohme, K., et al., *Species differentiation of seafood spoilage and pathogenic gram-negative bacteria by MALDI-TOF mass fingerprinting*. J Proteome Res, 2010. **9**(6): p. 3169-83.
- [210] Böhme, K., et al., *Chapter 25 - Proteomics of Food Spoilage Pathogens*, in *Proteomics in Food Science*, M.L. Colgrave, Editor. 2017, Academic Press. p. 417-431.
- [211] Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing*. Nature, 2020. **583**(7816): p. 459-468.
- [212] Messner, C.B., et al., *Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection*. Cell Syst, 2020. **11**(1): p. 11-24 e4.
- [213] Shu, T., et al., *Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19*. Immunity, 2020. **53**(5): p. 1108-1122 e5.
- [214] Messner, C.B., et al., *Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection*. Cell Systems, 2020. **11**(1): p. 11-24.e4.
- [215] Nachtigall, F.M., et al., *Detection of SARS-CoV-2 in nasal swabs using MALDI-MS*. Nat Biotechnol, 2020. **38**(10): p. 1168-1173.



Serment de Galien

Je jure en présence des maîtres de cette faculté :

- *D'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.*
- *D'exercer ma profession avec conscience, dans l'intérêt de la santé publique, sans jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine.*
- *D'être fidèle dans l'exercice de la pharmacie à la législation en vigueur, aux règles de l'honneur, de la probité et du désintéressement.*
- *De ne dévoiler à personne les secrets qui m'auraient été confiés ou dont j'aurais eu connaissance dans l'exercice de ma profession, de ne jamais consentir à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser les actes criminels.*
- *Que les hommes m'accordent leur estime si je suis fidèle à mes promesses, que je sois méprisée de mes confrères si je manquais à mes engagements.*

جامعة محمد الخامس
كلية الطب والصيدلة
- الرباط -

قسم الصيدلي

بسم الله الرحمن الرحيم

الحمد لله العظم



- ◀ أن أراقب الله في مهنتي
 - ◀ أن أبجل أساتذتي الذين تعلمت على أيديهم مبادئ مهنتي وأعترف لهم بالجميل وأبقى دوما وفيا لتعاليمهم.
 - ◀ أن أزاول مهنتي بوازع من ضميري لما فيه صالح الصحة العمومية، وأن لا أقصر أبدا في مسؤوليتي وواجباتي تجاه المريض وكرامته الإنسانية.
 - ◀ أن ألتزم أثناء ممارستي للصيدلة بالقوانين المعمول بها وبأدب السلوك والشرف، وكذا بالاستقامة والترفع.
 - ◀ أن لا أفشي الأسرار التي قد تعهد إلى أو التي قد أطلع عليها أثناء القيام بمهامي، وأن لا أوافق على استعمال معلوماتي لإفساد الأخلاق أو تشجيع الأعمال الإجرامية.
 - ◀ لأحظى بتقدير الناس إن أنا تقيدت بعهودي، أو أحتقر من طرف زملائي إن أنا لم أف بالتزاماتي.
- "والله على ما أقول شهيد"



المملكة المغربية
جامعة محمد الخامس بالرباط
كلية الطب والصيدلة
الرباط



أطروحة رقم: 78

سنة : 2021

البروتيوميات: عموميات، طرق وتطبيقات

أطروحة

قدمت ونوقشت علانية يوم : / / 2021

من طرف

السيد يونس بكشول

المزاد في 01 يوليوز 1995 بماسة (اشتوكة آيت باها)

لنيل شهادة

دكتور في الصيدلة

الكلمات الأساسية : بروتين؛ بروتيوميات؛ بروتيوم؛ مطيافية الكتلة؛ الترحيل الكهربائي

أعضاء لجنة التحكيم:

رئيس

مشرف

عضو

عضو

السيد امحمد أنصار

أستاذ في الكيمياء العضوية

السيد عز الدين إبراهيمي

أستاذ في التكنولوجيا الحيوية

السيد جواد الحارتي

أستاذ في الكيمياء العلاجية

السيد يوسف الرملي

أستاذ في الكيمياء