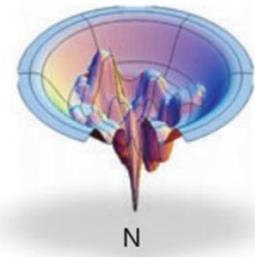
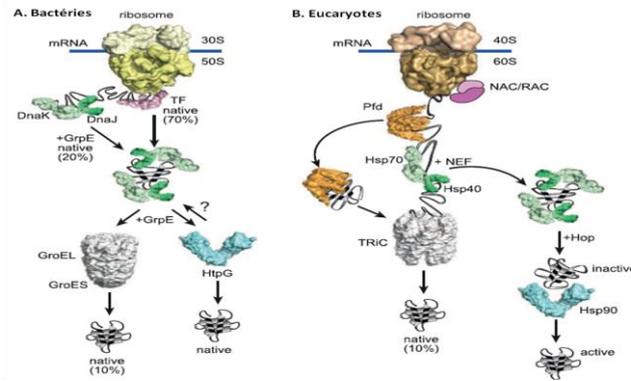
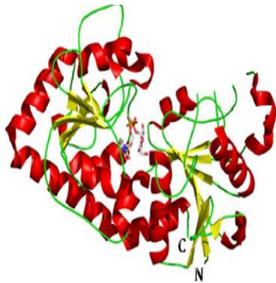
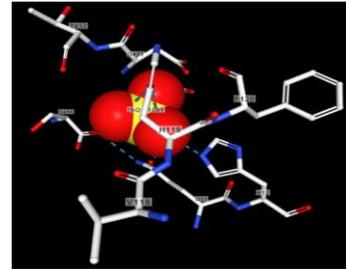
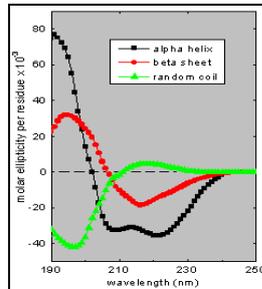
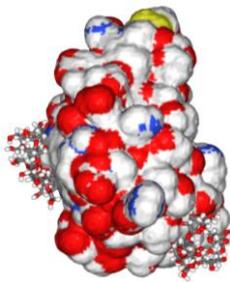


Polycopié de cours:

# Ingénierie des Protéines

Filière des Sciences Biologiques  
 Master 1 Biochimie



**BAKLI Mahfoud**, PhD  
 Maître de Conférences

## SYNOPSIS

Le syllabus du cours de la matière, Ingénierie des protéines tel que présenté dans le canevas de Master est présenté ci-dessous:

**Intitulé du Master : Biochimie      Semestre : 1**

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14-16 sem	C	TD	TP	Autres			Continu	Examen
<b>UE fondamentales</b>									
<b>UEF1 (O/P) Ingénierie des protéines et Enzymologie</b>	<b>135</b>	<b>4h30</b>	<b>3h</b>	<b>1h30</b>	<b>165</b>	<b>06</b>	<b>12</b>		
Matière 1 : Ingénierie des protéines	67h30	3h	1h30	/	82h30	03	06	✓	✓
Matière 2 : Enzymologie	67h30	1h30	1h30	1h30	82h30	03	06	✓	✓

**Objectifs de l'enseignement**

L'objectif de la formation est d'approfondir la compréhension des propriétés des protéines naturelles et d'introduire l'étudiant au domaine de l'ingénierie des protéines ainsi qu'aux notions théoriques et pratiques permettant d'appréhender les problématiques de la stabilité et du repliement des protéines.

**Connaissances préalables recommandées**

Biochimie structurale et fonctionnelle

**Contenu de la matière**

- 1- Stabilité, repliement et dynamique des protéines
- 2- Thermodynamique de la stabilité et du repliement des protéines (théorie et méthodes d'investigation)
- 3- Dénaturation réversible et irréversible
- 4- Repliement des protéines *in vivo* (mécanisme de repliement, formation des ponts disulfure, isomérisation des prolines, protéines chaperons, maladies conformationnelles)
- 5- Ingénierie de nouvelles interactions protéine-ligand
- 6- Ingénierie des enzymes
- 7- Modification chimique des protéines *in vitro* et *in vivo*
- 8- Ingénierie des protéines *in silico*

## Sommaire

Introduction .....	5
Chapitre 1. Stabilité, repliement et dynamique des protéines .....	7
1.1. Stabilité d'une protéine.....	7
1.2. Repliement des protéines.....	10
1.3. Dynamique des protéines.....	11
1.3.1. Fluctuations d'ordre atomique.....	12
1.3.2. Mouvements collectifs .....	12
1.3.3. Changements conformationnels provoqués .....	12
1.4. Approcher la dynamique des protéines .....	12
Chapitre 2. Thermodynamique de la stabilité et du repliement des protéines .....	14
2.1. Le code du repliement et la thermodynamique.....	14
2.2. Cinétique du repliement.....	15
2.3. Modèles et hypothèses de repliement des protéines .....	16
2.3.1. Modèle de diffusion-collision .....	16
2.3.2. Modèle de nucléation-condensation.....	17
2.3.3. Modèle de fermeture "zipping" et assemblage (ZA) .....	18
2.3.4. Modèle de puzzle de Jigsaw .....	19
2.3.5. Modèle d'effondrement hydrophobe et hypothèse d'entonnoir de repliement .....	20
2.3.6. Hypothèse de repliement des protéines pilotée par la stœchiométrie .....	23
2.4. Méthodes d'investigation de stabilité de protéine.....	23
2.4.1. Calorimétrie différentielle à balayage (DSC) .....	25
2.4.2. Dichroïsme circulaire (CD).....	26
Chapitre 3. Dénaturation des protéines .....	28
Chapitre 4. Repliement des protéines <i>in vivo</i> .....	32
4.1. Mécanisme de repliement.....	32
4.2. Comparaison entre le repliement des protéines <i>in vitro</i> et <i>in vivo</i> .....	34
4.3. Protéines chaperons .....	36
4.4. Formation des ponts disulfure .....	40
4.5. Isomérisation des prolines .....	41
4.6. Maladies conformationnelles.....	43
Chapitre 5. Ingénierie de nouvelles interactions protéine-ligand.....	45

5.1. Interactions protéine-ligand et mécanismes de liaison .....	45
5.2. Mécanismes de liaison protéine-ligand .....	46
5.3. Mécanismes physicochimiques de l'interaction protéine-ligand .....	48
5.3.1. Cinétique de liaison protéine-ligand .....	48
5.3.2. Thermodynamique de l'interaction protéine-ligand .....	49
5.4. Prédiction de l'interaction protéine-ligand.....	50
5.4.1. Docking (Amarrage moléculaire) .....	50
5.4.2. Criblage virtuel .....	51
Chapitre 6. Ingénierie des enzymes.....	53
6.1. Rappels sur les mécanismes de la catalyse enzymatique: .....	53
6.2. Histoire de l'ingénierie enzymatique .....	55
6.3. Technologie et ingénierie des enzymes .....	58
6.4. Classification des enzymes .....	60
6.5. Application industrielle des enzymes .....	60
6.6. Production des enzymes .....	62
6.6.1. Sources des enzymes.....	62
6.6.2. Criblage des enzymes.....	64
6.6.3. Processus de fermentation.....	65
6.6.4. Séparation et purification des enzymes.....	67
6.7. Enzyme immobilisée .....	69
Chapitre 7. Modification chimique des protéines <i>in vitro</i> et <i>in vivo</i> .....	73
Chapitre 8. Ingénierie des protéines <i>in silico</i> .....	79
8.1. Prédiction de la structure des protéines .....	79
8.1.1. Méthodes basées sur la bioinformatique .....	79
8.1.2. Méthodes basées sur la physique .....	80
8.2. Conception des protéines .....	81
8.3. La biologie synthétique .....	82
Références .....	87

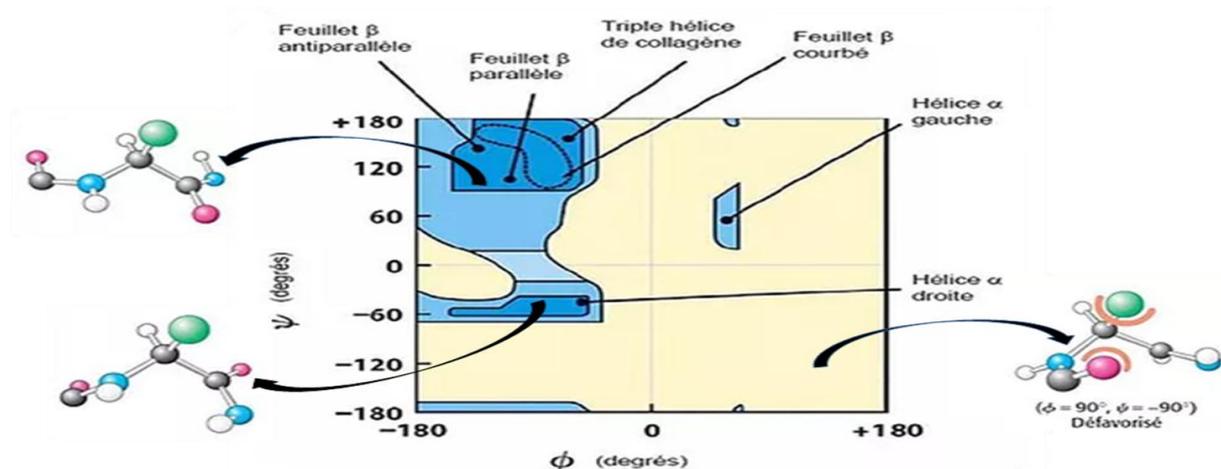
## Introduction

Les protéines sont les biomolécules clés dans presque tous les processus physiologiques et pathologiques qui se produisent dans la cellule. La fonctionnalité des protéines est liée à leur conformation (relation structure-fonction), qui est finalement dictée par leur séquence d'acides aminés unique qui est codé aussi par la séquence génique. De grandes variétés de protéines sont disponibles, des plus petites aux plus grandes, monomères à oligomères, globulaires ou fibrillaires, hydrophobes ou hydrophiles, existantes ou associées à d'autres protéines / biomolécules telles que les glucides, les acides nucléiques,.. etc., avec conformation fixe ou avec une gamme de facteurs dynamiques. Une telle nature diversifiée de protéines peut être attribuée à l'arrangement différentiel d'acides aminés dans une chaîne polypeptidique.

En outre, l'organisation des chaînes de protéines par les événements moléculaires tels que les modifications de co-traduction et les modifications post-traductionnelles ajoute une couche supplémentaire de diversification à ces entités biologiques. Toutes ces variations dans les protéines conduisent à des structures tridimensionnelles différentielles qui les rendent finalement polyvalentes quant à leurs caractéristiques fonctionnelles. Les protéines remplissent une grande variété de fonctions dans les cellules: organisation structurelle, maintien de la forme des cellules, synthèse de biomolécules, capteurs moléculaires, aides au mouvement cellulaire, interaction avec d'autres cellules et homologues biomoléculaires, défense contre les agents infectieux, ...etc.

Par ailleurs, la polymérisation des acides aminés se produit par la formation de liaisons CO–NH ou de liaisons peptidiques entre les acides aminés par élimination de molécules d'eau. Différents types de polypeptides ont leurs différentes séquences d'acides aminés et deviennent des protéines biologiquement actives en se repliant dans leurs structures 3D spécifiques. L'organisation structurelle des protéines a été divisée en quatre classes; à savoir structure primaire, secondaire, tertiaire et quaternaire.

Les groupes peptidiques existent généralement dans une conformation *trans* stable. Deux types de liaisons qui permettent la rotation autour du squelette polypeptidique sont N-C $\alpha$  et C-C $\alpha$ . Les angles autour de ces liaisons sont appelés angles de torsion (rotation ou dièdre)  $\Phi$  et  $\Psi$ .  $\Psi$  est l'angle entre la liaison C-C $\alpha$  et la liaison peptidique adjacente, alors que  $\Phi$  est l'angle entre C $\alpha$ -N et la liaison peptidique adjacente. La conformation du polypeptide peut être dictée par ces angles. Les valeurs pour les deux angles de torsion s'avèrent être de  $180^\circ$  lorsque la chaîne polypeptidique est dans sa conformation complètement étendue. Beaucoup de contraintes stériques sont imposées sur ces angles de torsion qui limitent la gamme de conformation des polypeptides. Les valeurs de  $\Psi$  et de  $\Phi$  autorisées stériquement ou énergétiquement favorables ont été représentées dans un graphe 2D connu sous le nom de diagramme de Ramachandran qui a été développé en 1963 par G.N. Ramachandran (biophysicien indien), basé sur la distance de Van der Waals (Fig. 1). Le digramme de Ramachandran entre  $\Psi$  et  $\Phi$  montre 75% de la région inaccessible du point de vue de la conformation de la chaîne principale du polypeptide. Les valeurs de  $\Psi$  et de  $\Phi$  tombent dans les régions autorisées pour toutes les structures secondaires observées dans les protéines (hélices  $\alpha$  et feuillets  $\beta$ ). Certains des résidus où le repliement de la protéine compense les contraintes stériques peuvent avoir les valeurs  $\Psi$  et  $\Phi$  situées en dehors de la région autorisée.



**Fig. 1. Diagramme de Ramachandran.** Montre les états les plus stables (bleu foncé), les états limites (bleu clair) et les états impossibles "défavorisés" du fait du risque de collision des atomes (en rose). Il est possible de localiser aussi les différentes structures protéiques secondaires.

## Chapitre 1. Stabilité, repliement et dynamique des protéines

### 1.1. Stabilité d'une protéine

Pour remplir leur fonction biologique, la plupart des protéines doivent se replier en une seule structure bien définie, l'état natif, et rester repliées de manière stable pendant une période prolongée. Pour y parvenir, l'énergie libre des états natifs doit être inférieure à celle des états non repliés ou mal repliés. La stabilité des protéines présente un large éventail de manifestations pratiques dans la nature et en laboratoire, notamment la résistance aux températures élevées, au dénaturant, aux protéases et au pH non physiologique; de plus, les rendements *in vivo* de protéine fonctionnelle, de solubilité et de localisation cellulaire ou extracellulaire appropriée sont souvent corrélés à la stabilité de la protéine. Les protéines naturelles ne sont souvent que faiblement stables, avec des différences d'énergie libre entre état natif et les états dépliés ou mal repliés aussi bas que 5 kcal / mol. Étant donné que la stabilité marginale entraîne souvent une sensibilité aux mutations ou aux modifications de l'environnement, elle a de vastes implications pour la compréhension des contraintes liées à l'évolution des protéines, des causes sous-jacentes de la maladie, et pour la conception de variants améliorés pour la recherche, la biotechnologie et la médecine. Par exemple, les mutations qui causent des maladies humaines ne modifient souvent pas la fonction moléculaire mais diminuent la stabilité de la protéine au point que trop peu de la protéine est disponible pour remplir sa fonction; Dans un exemple éclairant, les mutations associées au cancer dans p53 déstabilisent souvent la protéine, de sorte que moins de la moitié de la protéine exprimée est correctement repliée. Dans d'autres cas, des mutations provoquant une maladie peuvent provoquer un mal repliement ou induire une agrégation potentiellement cytotoxique et à la formation de fibrilles. Tous les organismes ont développé des mécanismes complexes qui traitent des problèmes potentiels liés à la stabilité marginale des protéines; collectivement, ces mécanismes sont connus sous le nom de réseau de protéostasie. Ce dernier

est essentiel pour que de nombreuses protéines puissent se replier et fonctionner efficacement dans des conditions normales et de stress. Il veille au contrôle qualité de la protéine et assure un repliement optimal et l'élimination des protéines mal repliées et / ou endommagées. Ce réseau impliqué dans la protéostasie comporte, entre autres, essentiellement des chaperons moléculaires (voir chapitre 4, protéines chaperons) et des enzymes détoxifiantes. De plus, le vieillissement cellulaire a été associé à une diminution des fonctions de ce réseau, qui conduit à la formation progressive d'amyloïdes cytotoxiques (voir chapitre 4, maladies conformationnelles). Ces énormes implications pour chaque aspect de la recherche biomédicale ont conduit à un effort continu pour décoder les principes sous-jacents qui régissent la stabilité des protéines et ses nombreuses manifestations. En termes généraux, la stabilité des protéines est étudiée dans deux directions complémentaires: premièrement, la stabilité cinétique ou thermodynamique des protéines, mesurée *in vitro* face au stress, telle que dénaturant, protéase ou température élevée; et deuxièmement, la stabilité *in vivo* ou l'efficacité du repliement des protéines à l'état natif et du maintien de cet état dans les cellules vivantes.

La structure des protéines à différents niveaux d'organisation, y compris secondaire, tertiaire et / ou quaternaire, est maintenue ou stabilisée par l'effet additif de nombreuses forces faibles. Ces forces sont la résultante de faibles interactions non-covalentes qui se produisent dans la protéine elle-même, mais certaines de ces forces sont dues à l'interaction des protéines avec l'environnement externe. La résultante nette de ces forces entraîne la conformation stable de repliement des protéines. Les plus importantes d'entre elles sont les interactions hydrophobes qui amènent la plus part des chaînes latérales des acides aminés hydrophobes vers l'intérieur de la protéine pour les protéger de l'eau (effet hydrophobe). En effet, on suppose que l'eau construit une "cage" autour de ces molécules car elle ne peut pas les intégrer dans son réseau de liaisons d'hydrogènes. La conséquence en est l'augmentation de l'ordre de l'eau, et une

diminution d'entropie. L'agrégation de groupements hydrophobes "apolaires" minimise cette perte d'entropie de l'eau. D'autres contributions significatives sont les liaisons hydrogènes et les ponts salins entre les carboxylates des acide aspartique (Asp) et glutamique (Glu) et les chaînes latérales protonées des résidus lysine (Lys), arginine (Arg) et histidine (His), porteuses d'une charge opposée. Bien que la force individuelle soit faible par rapport à celle d'une liaison covalente typique qui est d'environ 90 kcal/mol, l'ensemble de ces nombreuses interactions confèrent un haut degré de stabilité à la conformation biologiquement fonctionnelle d'une protéine (tableau 1.1). Certaines protéines contiennent des liaisons covalentes disulfures (S-S) qui relient les groupements sulfhydryles de résidus cystéines (Cys). La formation des liaisons disulfures implique l'oxydation des groupements sulfhydryles des cystéines et nécessite de l'oxygène. Les liaisons disulfures intrachânes augmentent encore la stabilité de la conformation repliée d'un polypeptide, tandis que les liaisons disulfures interchânes stabilise la structure quaternaire de certaines protéines oligomériques.

En résumé; les forces qui stabilisent la conformation repliée de la protéine sont: (1) liaisons hydrogènes intramoléculaires, (2) entropie de déshydratation, (3) interactions hydrophobes, (4) ponts salins et (5) interactions dipolaires. Les forces qui déstabilisent la conformation repliée de la protéine sont: (1) liaisons hydrogènes avec l'eau et (2) perte d'entropie configurationnelle (voir chapitre 2).

**Table1.1 Interactions chimiques qui stabilisent les polypeptides**

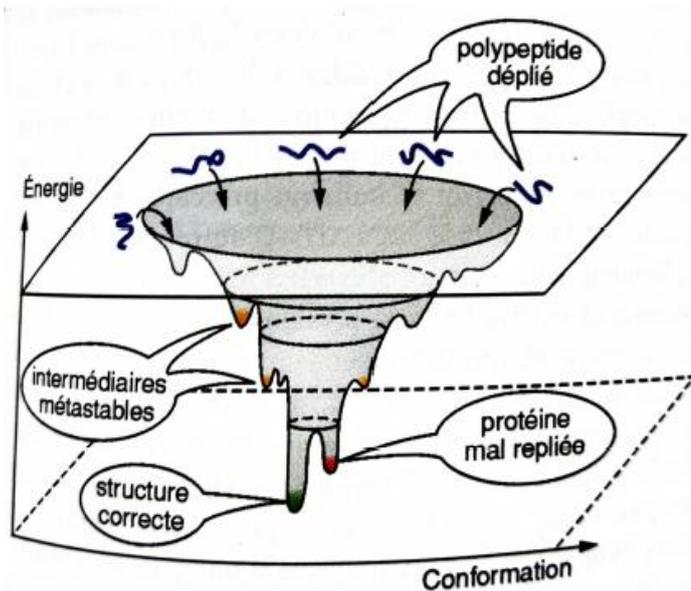
Interaction	Dépendance à distance	Distance typique	Energie libre (enthalpie de dissociation des liaisons pour les liaisons covalentes)
Covalente	-	1.5 Å	356 kJ/mol (610 kJ/mol pour la liaison C=C)
Disulfure	-	2.2 Å	167 kJ/mol
Pont salin	Donneur et accepteur < 3.5 Å	2.8 Å	12.5-17 kJ/mol, peut atteindre 30 kJ / mol pour les ponts salin entièrement ou partiellement enterrés, moins si le pont salin est externe
Liaison hydrogène	Donneur et accepteur < 3.5 Å	3.0 Å	2-6 kJ/mol dans l'eau, 12.5-21 kJ/mol si un donneur ou un accepteur est chargé
Electrostatique	Dépend de la constante diélectrique du milieu. Criblé par l'eau. Dépendance 1/r	Variable	Dépend de la distance et l'environnement. Peut être très forte dans la région nonpolaire, mais très faible dans l'eau
Van der Waals	Courte portée. Chutes de plus de 4 Å de séparation. Dépendance 1/r <sup>6</sup>	3.0 Å	4 kJ/mol (4-17 dans l'intérieur de la protéine) dépendant de la taille du groupe

## 1.2. Repliement des protéines

Le repliement protéique "folding" est le processus de transformation d'une information linéaire unidimensionnelle "contenue dans l'ADN" codée dans la séquence d'acides aminés en une structure tridimensionnelle et fonctionnelle. Il s'agit essentiellement d'un processus de chimie physique pour lequel le mécanisme reste insaisissable à cause des trois problèmes incomplètement résolus suivants: i) la question thermodynamique de comment une structure native résulte des forces inter-atomiques agissant sur une séquence d'acides aminés le code de repliement; ii) le problème cinétique de la façon dont une structure native peut se replier si vite taux de Repliement; iii) le problème informatique de la prédiction de la structure native d'une protéine à partir de sa séquence d'acides aminés, la prédiction de la structure protéique.

En effet, les protéines ont besoin pour se replier de moins d'une seconde en générale. L'exemple numérique simple d'un polypeptide de 100 acides aminés montre que cette vitesse n'est rien moins qu'évidente : même si les angles de torsion  $\Psi$  et  $\Phi$  le long de la chaîne principale étaient restreints à deux combinaisons énergétiquement favorables (deux points dans l'espace des conformations au diagramme de Ramachandran), il existerait encore  $2^{100}$ , soit environ  $10^{30}$  conformations possibles de la chaîne principale. La rotation d'un angle à un autre ne prend certes qu'un très court, de l'ordre de  $10^{-11}$  secondes. Néanmoins, essayer toutes les  $10^{30}$  combinaisons prendrait encore dans cette hypothèse 100 milliards d'années!

Le repliement des protéines à partir du polypeptide néo synthétisé ne peut donc pas être un processus qui fonctionne par essai et erreur. Mais il semble par ailleurs qu'il n'existe pas de chemin unique prédéterminé de repliement, comme cela a été souvent supposé dans le passé, pour résoudre ce paradoxe, dit de Levinthal, du repliement rapide des protéines. En fait, les polypeptides non repliés se meuvent dans un entonnoir énergétique vers la structure de la protéine native (Fig. 1.1).



**Fig. 1.1. Représentation schématique d'un « entonnoir de repliement »** (Müller-Esterl *et al.*, 2007). Les polypeptides dépliés peuvent se mouvoir par une quantité innombrable de chemins vers la structure tertiaire correcte. Sur ces chemins se trouvent souvent des intermédiaires stables de repliements (des minima locaux d'énergie), mais aussi parfois des « pièges » dans lesquels, la protéine reste dans un état incorrectement replié.

Dans cet entonnoir, le nombre de conformations énergétiquement possibles est rapidement limité et plusieurs chemins peuvent conduire au but. Ces chemins passent aussi souvent par des intermédiaires métastables, c'est-à-dire des « dépressions » locales de l'entonnoir. Parfois, ces dépressions peuvent constituer des pièges, ou la protéine reste emprisonnée dans un état incorrectement replié. La force motrice, comme dans tout processus naturel, est la diminution d'énergie libre.

### 1.3. Dynamique des protéines

Les protéines, comme toutes les autres macromolécules biologiques, sont des molécules non rigides "souples" qui oscillent en permanence et dont les mouvements structuraux ont une signification fonctionnelle. Leur structure fluctue autour de son état d'équilibre et adopte de nombreuses conformations. Ces fluctuations confèrent aux protéines la possibilité de s'adapter à un changement de milieu environnant ou de fixer des ligands d'une manière optimale, pour la catalyse par exemple. De plus, des changements conformationnels importants permettant la transition d'un état à un autre sont nécessaires pour réguler l'activité, par exemple le passage d'un état inactif à un état actif ou d'un état de faible affinité à un état de forte affinité. Ainsi, dans le cas des protéines allostériques "enzymes allostériques", l'activité biologique est contrôlée par des mouvements globaux faisant suite aux interactions des substrats eux-mêmes ou de leurs produits, ou encore à la fixation d'effecteurs. En raison de la compacité atomique

importante qui règne au sein des protéines, les changements conformationnels de grande amplitude s'accompagnent de déplacements concertés d'un grand nombre d'atomes. À ce propos, il convient de remarquer qu'une densité élevée des atomes à l'intérieur de la protéine et une relative rigidité de certains de ses segments peptidiques sont nécessaires à la transmission de mouvements à travers la molécule. Certaines protéines satisfont à ces deux conditions et assument la transmission vectorielle de mouvements. Les mouvements intramoléculaires des protéines sont classés en trois grandes catégories selon leur cohérence :

**1.3.1. Fluctuations d'ordre atomique**, telles que les vibrations de liaison individuelles, dont les périodes vont de  $10^{-15}$  à  $10^{-11}$  s et de déplacement de 0,01 à 1 Å.

**1.3.2. Mouvements collectifs**, où les groupes d'atomes liés par covalence, dont la taille est comprise entre celles de chaînes latérales d'acides aminés et celles de domaines entiers, se déplacent comme des entités avec des périodes qui vont de  $10^{-12}$  à  $10^{-3}$  s et des déplacements compris entre 0,01 et  $> 5$  Å. De tels mouvements se font plus ou moins fréquemment comparés à leurs périodes caractéristiques.

**1.3.3. Changements conformationnels provoqués**, où des groupes d'atomes, de tailles qui vont de la chaîne latérale à la sous unité entière, se déplacent en réponse à des stimuli spécifiques tels que la liaison d'une petite molécule.

#### **1.4. Approcher la dynamique des protéines**

Le repliement des protéines est un processus qui met en jeu différentes échelles temporelles. La formation des structures secondaires se fait sur des durées entre 10 ns et 1 ms. tandis que la formation des structures tertiaires et quaternaire natives peut prendre jusqu'à plusieurs heures. L'observation des dynamiques de repliement des protéines ou de leurs mouvements au cours de leur activité a fait l'objet de nombreuses expériences qui ont permis de développer plusieurs techniques. Le choix d'une technique d'observation dépend des objectifs de l'expérience, de la protéine à étudier, des informations que l'on souhaite obtenir dans les deux

échelles, cellulaire ou à l'échelle de la protéine. Les techniques, pour approcher et observer la dynamique des protéines, les plus utilisées sont résumées dans les tableaux 1.2 et 1.3.

**Table 1.2 Méthodes pour approcher la dynamique des protéines**

Méthodes d'approches à l'échelle cellulaire	Méthodes d'approches à l'échelle de la protéine
<u>1. Gènes rapporteurs</u> - Fusion transcriptionnelle (séquences régulatrices et/ou promotrices...) - Fusion traductionnelle (expression d'une protéine hybride) - Les protéines fluorescentes (GFP, DsRed, YFP...)	1. Diffraction aux Rayon X (Cristallographie)
<u>2. Photoblanchiment (Photobleaching)</u> - FRAP : Fluorescence Recovery After Photobleaching - FLIP : Fluorescence Loss Induced by Photobleaching	2. Résonance Magnétique Nucléaire (RMN)
<u>3. Mesure des interactions protéiques</u> - FRET : Fluorescence Resonance Energy Transfer - BRET : Bioluminescence Resonance Energy Transfer - PCA : Protein Complementation Assay	3. Modélisation / Simulation

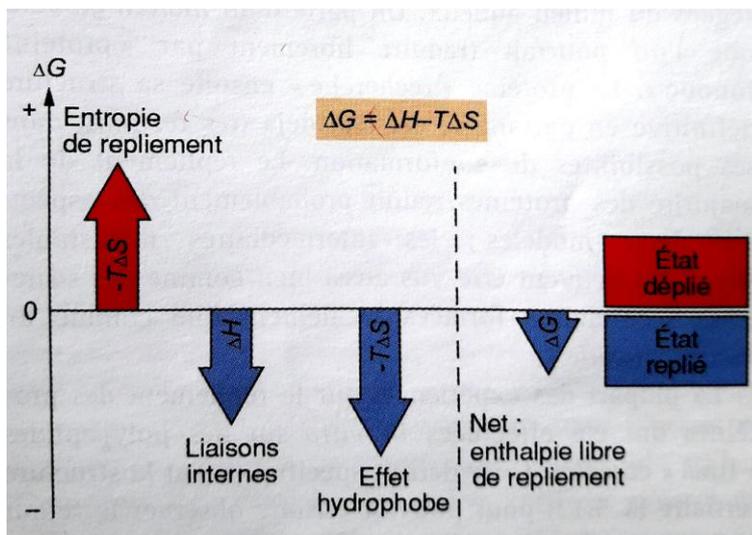
**Table 1.3 Principales techniques d'observation de dynamique moléculaire**

Technique	Type d'information	Détails
Absorption	Information sur le changement d'environnement d'un chromophore	Lorsque le chromophore se déplace au cours du repliement, il change d'environnement ce qui déplace ses bandes d'absorption.
Fluorescence	Selon la nature du fluorophore, on peut obtenir des informations locales ou globales sur le repliement de la protéine	1. L'observation de la fluorescence de certains acides aminés aromatiques renseigne sur les variations de leur environnement. 2. L'ajout d'un couple de fluorophores donneur/accepteur à deux extrémités de la protéine renseigne sur le mouvement relatif des deux zones de la protéine.
Absorption Infrarouge	Donne une information globale sur la structure secondaire	La bande d'absorption à $1650\text{ cm}^{-1}$ , impliquant les vibrations de la liaison C=O de la chaîne peptidique, subit un décalage lors d'une variation de structure secondaire.
Dichroïsme circulaire	Donne une information globale sur le contenu en structure secondaire	Le dichroïsme circulaire est utilisé pour sonder de manière quantitative la structure secondaire.
Spectroscopie Raman	Informations locales et globales	Permet d'observer la structure secondaire ou des changements de structures plus localisés selon le chromophore étudié.

## Chapitre 2. Thermodynamique de la stabilité et du repliement des protéines

### 2.1. Le code du repliement et la thermodynamique

Le code de repliement des protéines peut être vu comme étant le profil thermodynamique du processus d'interaction qui représente les différents types de forces entraînant le repliement des protéines, y compris les changements d'énergie libre (G), enthalpie (H) et entropie (S). La force prédominante qui entraîne le repliement des protéines est considérée comme l'interaction hydrophobe ou la désolvatation, qui est essentiellement l'effet entropique et qui est réparti à la fois localement et non localement dans la séquence. Les structures secondaires natives et les interactions finalement observées (telles que les interactions électrostatiques, Van der Waals et les liaisons hydrogène) sont la conséquence plutôt que la cause des forces de repliement, qui mènent à la recherche au minimum global d'énergie libre du système protéine-solvant par compensation d'entropie-enthalpie. La Force motrice du processus de repliement est la diminution d'énergie libre du système :  $\Delta G = \Delta H - T\Delta S$  (doit prendre une valeur négative, T = température, °K). Cette situation est le résultats de facteurs opposés :



**Fig. 2.1. Energie libre de repliement des protéines** (Müller-Esterl *et al.*, 2007). La limitation de la liberté de conformation de la chaîne polypeptidique travaille contre le repliement, les liaisons non covalentes et l'effet hydrophobes le favorise au contraire. En dernier ressort, la variation d'énergie libre  $\Delta G$  doit être négative pour que le repliement soit spontané.

d'un côté, un état enthalpiquement favorisé ( $\Delta H < 0$ ) est atteint grâce à la formation de nouvelles interactions non covalentes, liaisons ioniques, liaisons hydrogènes et interactions de Van der Waals. D'un autre côté, la limitation à une conformation ordonnée fait baisser l'entropie de la chaîne polypeptidique, ( $\Delta S < 0$ ) (Fig. 2.1).

Cependant, l'eau environnante gagne en entropie lorsque les chaînes latérales hydrophobes sont enfouies dans la protéine repliée. Le premier et le dernier facteur précipités favorisent donc le repliement, alors que la perte d'entropie du polypeptide agit dans l'autre sens.

## 2.2. Cinétique du repliement

En 1961, Anfinsen et ses collègues ont observé que la ribonucléase A complètement dénaturée (en rompant tous les ponts disulfures de la protéine) était capable de se replier spontanément en une enzyme complètement active. Anfinsen en a déduit un postulat qui porte son nom : Toute l'information nécessaire à l'acquisition de la structure native d'une protéine dans un environnement donné est contenue dans sa séquence. Cela signifie que les protéines acquièrent leur structure native spontanément. Peu de temps après, en 1968, Levinthal (Levinthal, 1969) a reconnu que les protéines ont des espaces conformationnels beaucoup trop vastes pour permettre une recherche approfondie de la structure repliée native dans un temps biologiquement rapide. Pour résoudre ce «paradoxe», Levinthal a proposé la notion de voie de repliement des protéines à travers laquelle les protéines peuvent rechercher et converger rapidement vers les états natifs.

Les protéines qui se replient relativement rapidement ont généralement plus d'éléments structuraux locaux tels que des hélices  $\alpha$  et des boucles serrées, tandis que les protéines qui se replient relativement lentement ont plus d'éléments structuraux non locaux tels que des feuillets  $\beta$ . Cependant, existe-t-il une seule voie ou plusieurs voies parallèles par lesquelles une protéine se replie? Comment caractériser les voies de repliement des protéines reliant les états dénaturés et les états natifs? Ce sont deux aspects du problème de la cinétique du repliement des protéines. Les réponses complètes à ces questions requièrent l'introduction du concept de «l'entonnoir de repliement des protéines» basé sur la théorie du paysage énergétique libre. En bref, le repliement des protéines peut être vu comme descendant du paysage énergétique en entonnoir par de multiples voies parallèles allant de la grande majorité

des conformations non-natives individuelles aux états natifs autour du bas de l'entonnoir. A tout moment, la protéine existe sous la forme d'un ensemble de conformations et peut être piégée de manière transitoire dans de nombreux puits minimums d'énergie locale.

Il faut garder à l'esprit que le temps et les voies de repliement sont des concepts liés à la cinétique, qui définit les barrières énergétiques entre les différents états conformationnels; tandis que l'énergie libre correspond aux états conformationnels par rapport à la thermodynamique, qui définit les probabilités / populations / durées relatives des différents sous-états conformationnels. Partant de ces postulats, on peut se demander alors comment il est possible que certaines protéines ne se replient pas correctement. Se pose donc la question de savoir comment la protéine arrive à se replier dans sa structure native sans aide extérieure, et quels mécanismes sont mis en jeu au cours du repliement.

### **2.3. Modèles et hypothèses de repliement des protéines**

En raison de la complexité des problèmes de repliement des protéines, plusieurs modèles simplifiés ont été proposés pour explorer les espaces conformationnels et séquentiels, décrire la séquence des événements dans le processus de repliement, et explorer, examiner et découvrir le mécanisme de repliement des protéines. En essayant de répondre aux questions suivantes: quelle force conduit le repliement, et quel est l'ordre d'apparition des différentes structures au cours de la réaction de repliement?

#### **2.3.1. Modèle de diffusion-collision**

Le modèle de diffusion-collision, a été développé par Karplus et Weaver, en 1976, sur des bases théoriques et a été réévalué en 1994 à la lumière des données expérimentales plus récente (Fig. 2.2 A). Le modèle de collision de diffusion est basé sur la division du repliement d'une molécule de protéine en parties telles que les informations stockées dans la séquence de chaque partie peuvent être utilisées indépendamment. Ces parties divisées, appelées microdomaines, peuvent être des parties de structures secondaires naissantes (telles que des

hélices  $\alpha$ , des brins  $\beta$  et des boucles) ou des agrégats hydrophobes (c'est-à-dire des acides aminés hydrophobes proches les uns des autres dans la séquence d'acides aminés). Dans ce modèle, le processus de repliement pourrait commencer avec une chaîne polypeptidique dans un état enroulé aléatoire étendu: les microdomaines sont formés en premier; puis ils se déplacent de façon diffusive avec des collisions qui se produisent entre eux, conduisant à la coalescence dans les intermédiaires multi-microdomaines ou agrégats supérieurs; et enfin, ces intermédiaires multi-microdomaines forment la structure tertiaire exacte par des réarrangements mutuels. La formation de microdomaines ou de structures secondaires (causées par des interactions locales) et leurs interactions plutôt que des interactions entre des acides aminés individuels réduiraient considérablement les temps de repliement à des valeurs raisonnables. Cependant, un tel modèle simplifié ne peut ni caractériser les mouvements détaillés impliqués dans le processus de repliement au niveau atomique ni fournir des informations complètes sur les barrières énergétiques et les populations d'états conformationnels individuels. Par conséquent, le modèle de collision de diffusion est seulement utile pour une analyse semi-quantitative de la dynamique du repliement des protéines en se concentrant sur les microdomaines.

### **2.3.2. Modèle de nucléation-condensation**

Le mécanisme de nucléation-condensation (Fig. 2.2B) suggère que la réaction est initiée par la formation d'un noyau faiblement stable qui agit comme une matrice "ou un noyau" à partir de laquelle la structure native se propagerait par condensation rapide d'autres structures. Une caractéristique essentielle du modèle de nucléation-condensation est que l'apparition du noyau nécessite la formation simultanée d'interactions secondaires et tertiaires. Cela contraste le modèle de diffusion-collision qui, comme décrit ci-dessus, implique un processus hiérarchique d'assemblage structurel où les éléments de structure secondaires, guidés par des contacts locaux, sont initialement formés indépendamment de la structure tertiaire, puis la

structure tertiaire native. Par conséquent, la différence essentielle entre ces deux modèles de repliement est de savoir si la réaction de repliement peut se dérouler de manière concertée (comme dans le modèle de nucléation-condensation) ou par étapes (comme dans le modèle de diffusion-collision).

### 2.3.3. Modèle de fermeture "zipping" et assemblage (ZA)

Le modèle ZA (également appelé "hydrophobic zipper model") a été proposé par (Dill *et al.*, 1993) pour expliquer comment une protéine trouve son état natif sans une recherche globalement exhaustive. Dans ce modèle, la réaction de repliement des protéines commence avec les formations indépendantes de pièces structurales locales le long de la chaîne, puis ces pièces se développent (zip) ou coalesce (assembler) avec d'autres structures pour former le repliement natif. À première vue, un tel processus de repliement est très similaire à celui du modèle de diffusion-collision puisque tous les deux décrivent le processus hiérarchique de l'assemblage structurel. Cependant, ces deux modèles diffèrent les uns des autres dans certains aspects. Dans le modèle ZA, il y a un effondrement "collapse" rapide qui est entraîné par des interactions hydrophobes non locales conduisant aux formations simultanées des clusters hydrophobes, hélices  $\alpha$ , des feuilletts  $\beta$ , et le processus de fermeture "zipping" hydrophobe conduit à un ensemble d'états intermédiaires compacts qui ont beaucoup de structures secondaires mais peu de structures tertiaires; le processus lent consécutif consiste à casser / décompresser des contacts hydrophobes-hydrophobes incorrects (non natifs) pour avancer vers la structure native. Le ZA souligne donc que le repliement est dominé par des interactions non locales, c'est-à-dire que l'effondrement est la force motrice et que les structures secondaires sont une conséquence de l'effondrement plutôt que de sa cause.

Le modèle de diffusion-collision suppose que les interactions locales sont des facteurs importants dans la réduction de la recherche conformationnelle, impliquant que les forces motrices principales du repliement des protéines sont les interactions hydrogène,

électrostatique et de Van der Waals menant aux formations précoces des structures secondaires (hélices  $\alpha$  et feuillets  $\beta$ ) et l'effondrement subséquent et l'assemblage dans la structure tertiaire. Le ZA est donc similaire, dans une certaine mesure, à la nucléation-condensation puisque dans les deux modèles, les interactions non locales dominent la formation de l'effondrement ou du noyau. Un avantage du modèle ZA est qu'il indique explicitement la cause de l'effondrement de la force hydrophobe. De cette façon, le modèle ZA est essentiellement similaire au modèle d'effondrement hydrophobe parce que les deux modèles suggèrent que c'est la force hydrophobe à médiation par le solvant qui entraîne la première étape de l'effondrement, dans le processus de repliement des protéines.

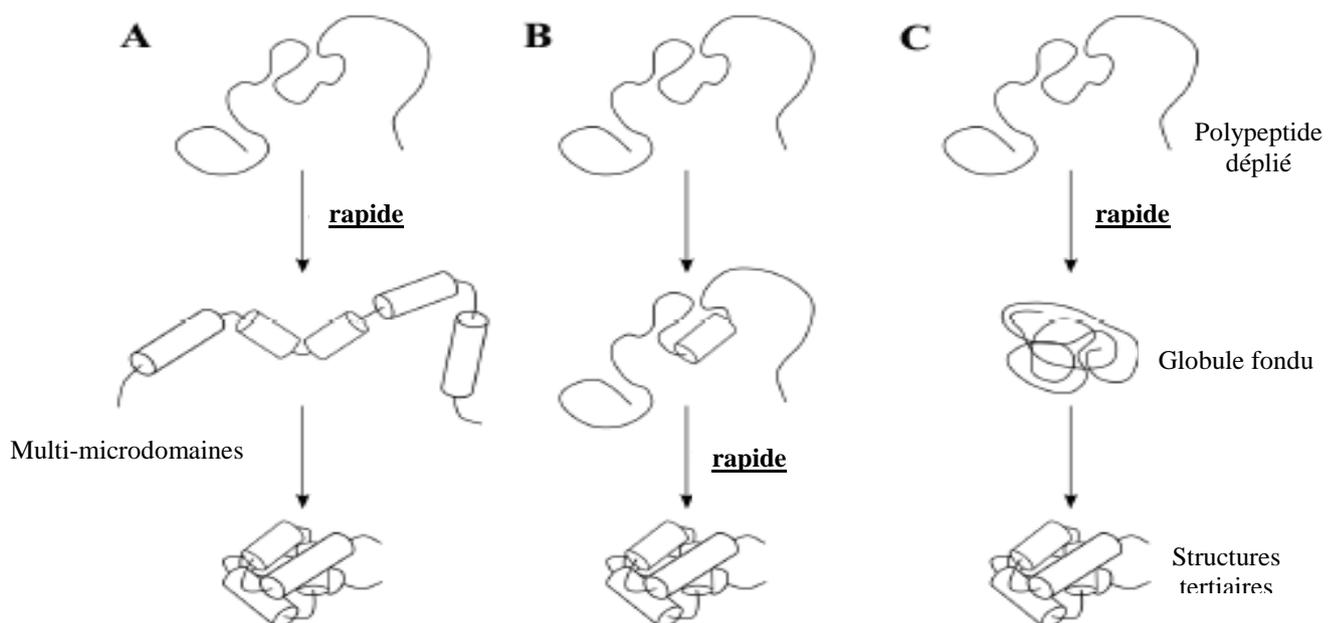
#### **2.3.4. Modèle de puzzle de Jigsaw**

Le modèle de puzzle, un mécanisme qualitatif pour expliquer le repliement des domaines compacts des protéines, a été proposé par Harrison & Durbin en 1985. Ils ont fait valoir que les protéines se replient par un grand nombre de voies parallèles différentes plutôt que par une seule séquence définitive d'événements, rendant ainsi le repliement plus robuste aux mutations qui n'affectent pas défavorablement la structure native. Ici, l'existence de multiples voies de repliement des protéines vers la structure native ressemble aux multiples voies pour atteindre une solution unique dans l'assemblage d'un puzzle. Selon ce modèle, l'identification des intermédiaires représente une description cinétique plutôt qu'une description structurale puisque chaque intermédiaire est constitué d'espèces hétérogènes en équilibre rapide. Ce modèle présente quelques similitudes avec le modèle diffusion-collision car sa cinétique pourrait être obtenue par les diffusions et les collisions des "microdomaines" si tous les microdomaines élémentaires des propriétés similaires et des intermédiaires multi-microdomaines de même taille ont des vitesses de repliement et de non repliement similaires. Le repliement des voies multiples mis en évidence par le modèle de puzzle est en fait cohérent avec une «nouvelle vue» du concept de l'entonnoir de repliement des protéines basé sur la

théorie du paysage énergétique: l'état natif d'une protéine à son minimum énergétique global est situé en bas du paysage énergétique en forme d'entonnoir; et chaque molécule peut suivre une voie microscopique différente du haut vers le bas.

### 2.3.5. Modèle d'effondrement hydrophobe et hypothèse d'entonnoir de repliement

Le modèle d'effondrement hydrophobe "hydrophobic collapse" (Fig. 2.2 C), proposé pour la première fois par (Dill, 1985), repose principalement sur l'observation que les états natifs des protéines globulaires contiennent souvent le cœur hydrophobe de chaînes latérales d'acides aminés apolaires à l'intérieur des protéines, tout en laissant la plupart des résidus polaires ou chargés sur la surface de la protéine exposée au solvant. Par conséquent, ce modèle considère que le repliement d'une protéine globulaire commence avec un effondrement rapide de la chaîne entraînée par la force hydrophobe, entraînant la formation des intermédiaires effondrés, c'est-à-dire le globule fondu "molten globule", dans lequel la structure secondaire et/ou tertiaire est formée ultérieurement.

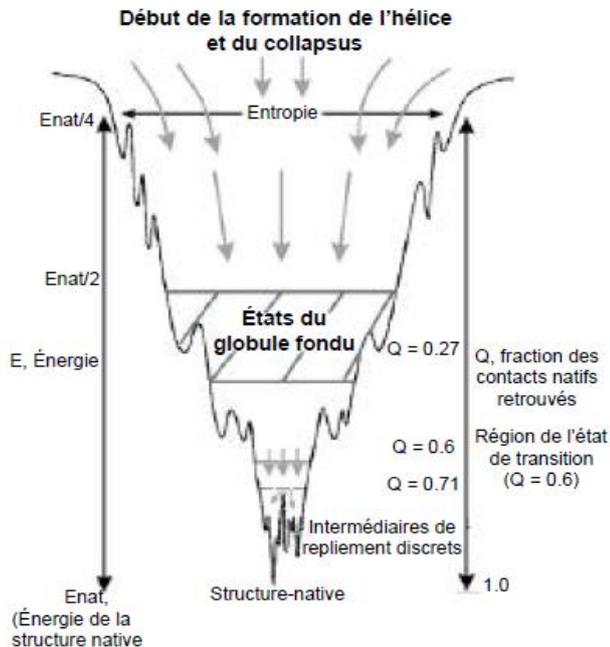


**Fig. 2.2. Représentation schématique de plusieurs modèles de repliement de protéines** (adaptée selon Liu *et al.*, 2012). (A) Modèle de diffusion-collision (B) Modèle de nucléation-condensation. (C) Modèle d'effondrement hydrophobe. L'étape rapide du repliement est indiquée.

Ce modèle souligne que le «premier événement» du repliement est que les chaînes latérales des résidus hydrophobes sont entraînées ensemble par l'action de l'eau, ce qui ressemble beaucoup à l'agrégation des gouttelettes d'huile dans l'eau. Lors de l'égouttage de l'huile dans un solvant aqueux ou lors de la récupération des protéines, la force hydrophobe non locale et non spécifique médiée par les molécules d'eau force la chaîne polypeptidique ou l'huile à se regrouper rapidement pour minimiser la surface des agrégats et maximiser la l'entropie du solvant, abaissant ainsi l'énergie libre totale de Gibbs des systèmes. Dans le cas du repliement des protéines, l'énergie libre est davantage abaissée par des contacts énergétiques favorables tels que l'isolement de chaînes latérales chargées électrostatiquement sur la surface de la protéine et la neutralisation des ponts salins à l'intérieur de la protéine. L'étape suivante du repliement de la protéine est un lent recuit vers la structure native: réarrangement mutuel et ajustement des blocs de construction déjà formés par rupture et reformation de diverses liaisons non covalentes, ce qui abaisse encore l'énergie libre du système via une contribution enthalpique. Comme mentionné ci-dessus, le modèle d'effondrement hydrophobe est similaire au modèle ZA "hydrophobic zipper", les deux mettent en évidence que le repliement est dominé par l'interaction hydrophobe médiée par l'eau.

Comparé au modèle d'effondrement hydrophobe, l'hypothèse de l'entonnoir de repliement caractérise un scénario plus complet de la façon dont une population de chaînes peptidiques dépliées descend du sommet du paysage énergétique en entonnoir vers le bas des états natifs via plusieurs voies. L'effondrement hydrophobe n'est que l'un des événements les plus importants survenant au début du repliement. D'autres événements aussi ont lieu tels que la formation du globule fondu, le piégeage des intermédiaires de repliement dans les minima locaux, le dépassement des barrières énergétiques locales, la formation de l'ensemble des états de transition, la reconfiguration des interactions tertiaires et les aspects thermodynamiques et cinétiques de la structure protéique. caractérisé par l'entonnoir de repliement. Dans la

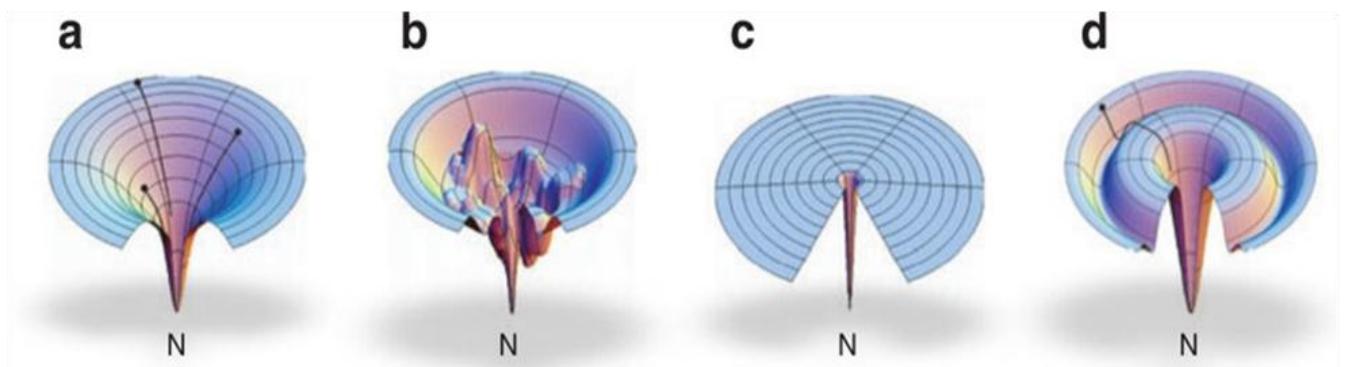
représentation canonique de l'entonnoir de repliement, la profondeur des puits représente la stabilisation énergétique de l'état natif (minimum global) par rapport à l'état dénaturé;



**Fig. 2.3. Repliement en entonnoir schématique d'une protéine** (Benhablès *et al.*, 2000). Au cours du repliement la protéine descend dans l'entonnoir jusqu'à atteindre la structure native. La largeur de l'entonnoir représente l'entropie, la profondeur.

la largeur des puits représente l'entropie conformationnelle du système; et la surface à l'extérieur du puit est montrée relativement plate (Fig. 2.3). Les intermédiaires de repliement sont des irrégularités dans les parois de l'entonnoir. Ainsi, une même protéine peut se replier en passant par différents intermédiaires selon le chemin parcouru à la surface de l'entonnoir (Fig 2.4).

Cette représentation permet d'expliquer les cinétiques de repliement monophasiques et multiphasiques ainsi que l'existence de phénomènes d'agrégation (qui correspondraient à des minima locaux dans la paroi de l'entonnoir).



**Fig. 2.4. Représentation tridimensionnelle d'entonnoirs de repliement protéique** (Dill *et al.*, 1997). (a) Entonnoir idéalisé : le repliement se fait de la même façon quel que soit la trajectoire de la protéine à sa surface, Plusieurs voies menant à un résultat stable. (b) Entonnoir de repliement réel avec des barrières d'énergie, des pièges cinétiques ; le repliement peut alors avoir plusieurs états intermédiaires et suivre différents mécanismes. (c) Repliement immédiat menant à un résultat stable. (d) Intermédiaires Significativement stables à surmonter dans la trajectoire avant le repliement final.

### 2.3.6. Hypothèse de repliement des protéines pilotée par la stœchiométrie

L'hypothèse du repliement des protéines par la stœchiométrie a été proposée par Mittal *et al.*, en 2010, dont les analyses statistiques de plus 3700 structures protéiques repliées ont révélé un principe unifiant de l'organisation de la chaîne principale. Une des découvertes intéressantes est que le nombre total de contacts possibles pour l'atome C $\alpha$  d'un acide aminé donné est parfaitement corrélé avec son pourcentage d'occurrence dans les séquences d'acides aminés, conduisant à la conclusion que le repliement des protéines est une conséquence directe d'une bande étroite des événements stœchiométriques des acides aminés dans les séquences primaires, indépendamment de la taille et du repliement d'une protéine. D'autre part, les résultats statistiques révèlent qu'il n'y a pas «d'interactions préférentielles» entre les acides aminés, conduisant ainsi à la conclusion que les «interactions préférentielles» entre les acides aminés n'entraînent pas le repliement des protéines.

En conclusion générale, les modèles ou hypothèses de repliement des protéines décrits ci-dessus, qui ont été construits à partir de résultats expérimentaux, ne sont pas indépendants et s'excluent mutuellement, mais ils sont inextricablement liés et tentent généralement de saisir différents aspects du repliement des protéines.

### 2.4. Méthodes d'investigation de stabilité de protéine

De nombreuses méthodes de biophysique permettent d'obtenir des informations sur les intermédiaires de repliement et étudier de façon générale le repliement protéique sont explicitées dans le tableau 2.1. Ces études ont bénéficié des avancées technologiques : les techniques de cinétique rapide, telles que les méthodes de flux (Stopped Flow) de cinétiques rapides avec des constantes de temps de l'ordre de la milliseconde, couplées à des méthodes physiques (fluorescence, dichroïsme circulaire "CD", puis RMN) ont été largement utilisées.

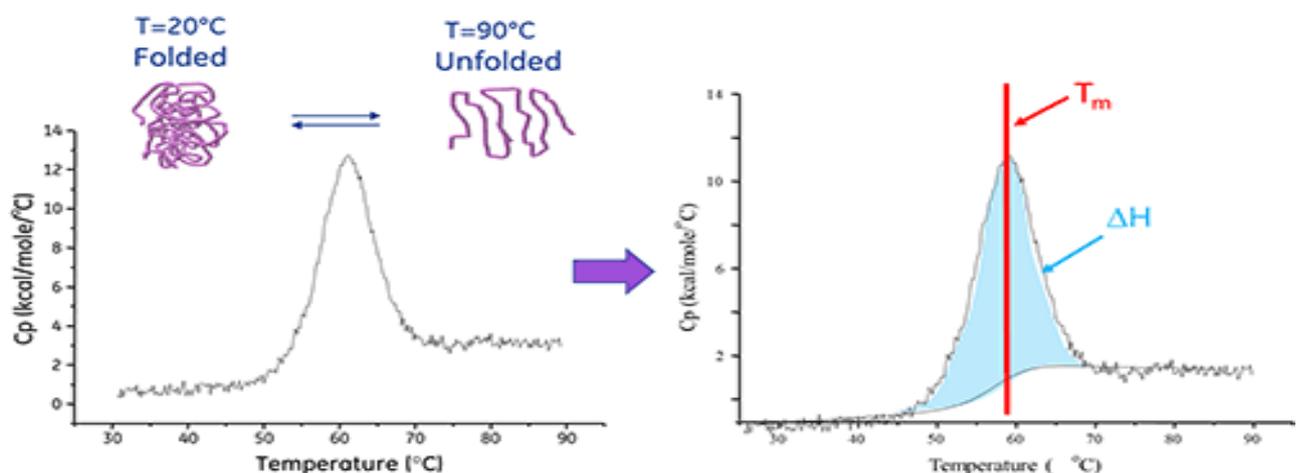
Plus récemment, sont apparues des méthodes ultrarapides avec des constantes de temps de l'ordre de la microseconde et même de la nanoseconde. Ainsi, l'observation des structures secondaires dans les étapes précoces du repliement a été réalisée pour la plupart des protéines. Par ailleurs, parmi les méthodes d'investigation les plus utilisées pour étudier la stabilité des protéines, la calorimétrie différentielle à balayage et le dichroïsme circulaire que nous allons détailler dans ce chapitre.

**Table 2.1 Techniques biophysiques utilisées pour étudier le repliement des protéines** (Benhabiles *et al.*, 2000).

Propriétés	Techniques	Résolution	Mesures
Compactage du cœur de la protéine	Fluorescence intrinsèque	< 1 ms	Orientation et environnement de la chaîne latérale du tryptophane (Engelhard, Evans, 1996a)
	Absorption aux UV	1 ms	Orientation et environnement de la chaîne latérale de la tyrosine (Udgaonkar, Baldwin, 1995)
	Fluorescence extrinsèque (ANS)	1 ms	Formation et interruption de segments hydrophobes (Engelhard, Evans, 1996b)
	Extinction de fluorescence	1 ms	Localisation de la chaîne latérale du tryptophane (Engelhard, Evans, 1996a)
	“Quenching CysteinyI”	10 s	Protection des chaînes latérales des cystéines (Ballery <i>et al.</i> , 1993)
Dimensions moléculaires	Anisotropie de fluorescence	1 ms	Mobilité de la chaîne du tryptophane et dimensions moléculaires globales (Jones <i>et al.</i> , 1995)
	Transfert d'énergie de fluorescence	1 ms	Distance scalaire entre un tryptophane et un fluorophore attachés covalamment (Rischel, Poulsen, 1995)
	Diffusion des rayons X aux petits angles	< 100 ms	Rayon de giration moyen (Eliezer <i>et al.</i> , 1995)
	Diffusion de la lumière quasi élastique	1 s	Rayon de giration moyen (Feng, Widom, 1994)
Structures secondaires et liaisons hydrogènes persistantes	Dichroïsme circulaire dans l'UV lointain	1 ms	Conformation du squelette, moyennée sur la séquence et la population étudiée (Evans, Radford, 1994)
	RMN en marquage pulsé	5–10 ms	Formation de liaisons hydrogènes stables du squelette peptidique (Baldwin, 1993)
	Spectroscopie de masse en marquage pulsé	5–10 ms	Formation de liaisons hydrogènes persistantes pour les intermédiaires du repliement (Miranker <i>et al.</i> , 1996)
Contacts tertiaires et structure native	Activité biologique	1 ms–1 s	Formation de la structure tertiaire native et du site actif (Evans, Radford, 1994)
	Repliement interrompu	10 ms	Vitesse de dépliement des intermédiaires discrets (Schreiber, Fersht, 1993)
	Dichroïsme circulaire en proche UV	1 ms	Formation de contacts tertiaires stables (entre aromatiques, pont disulfures) (Evans, Radford, 1994)
	RMN en temps réel	1 s	Formation de contacts tertiaires entre chaînes latérales spécifiques (Balbach <i>et al.</i> , 1995)
	Ingénierie des protéines		Contribution énergétique des chaînes latérales aux intermédiaires (Fersht, 1995c)

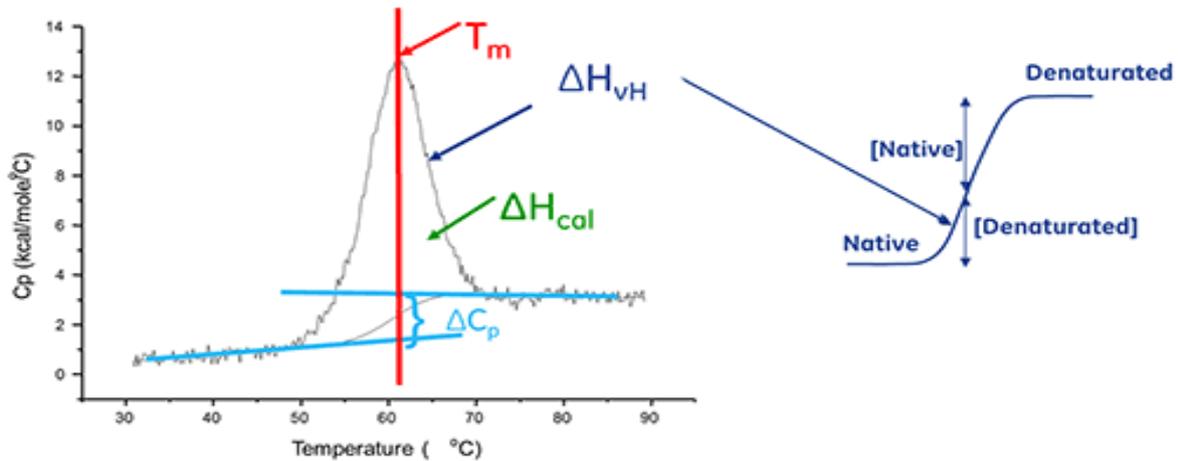
### 2.4.1. Calorimétrie différentielle à balayage (DSC)

La Calorimétrie Différentielle à Balayage (Differential Scanning Calorimetry ou DSC) est une technique utilisée pour caractériser la stabilité d'une protéine ou d'une autre biomolécule directement sous sa forme native. Cette caractérisation exploite la mesure de la variation de chaleur associée à la dénaturation thermique de la molécule lorsque celle-ci est chauffée à vitesse constante. En solution, une biomolécule est en équilibre entre ses conformations native (repliée) et dénaturée (dépliée). Plus la température médiane de transition thermique ( $T_m$ ) est élevée, plus la molécule est stable. La DSC mesure l'enthalpie ( $\Delta H$ ) qui résulte d'un changement de conformation induit par une élévation de température. Elle peut également être utilisée pour déterminer la variation de capacité calorifique ( $\Delta C_p$ ) associée à la dénaturation. La DSC permet d'identifier les facteurs contribuant au repliement et à la stabilité des biomolécules natives. Ces facteurs incluent les interactions hydrophobes, les liaisons hydrogène, l'entropie conformationnelle et l'environnement physique. Les changements conformationnels des protéines sous l'effet de la chaleur. Ces réarrangements structuraux se traduisent par une absorption de chaleur due à la redistribution des liaisons non covalentes. La calorimétrie différentielle à balayage mesure cette absorption de chaleur (Fig. 2.5).



**Fig. 2.5.** Explication conceptuelle des données DSC pour  $T_m$  et enthalpie (Yoo *et al.*, 2017).

L'enthalpie de dénaturation d'une protéine est la surface sous le pic de DSC normalisée par la concentration et s'exprime en calories (ou Joules) par mole. Dans certains cas, il est possible d'adapter des modèles thermodynamiques aux données pour en extraire l'énergie libre de Gibbs ( $\Delta G$ ), l'enthalpie calorimétrique ( $\Delta H_{cal}$ ), et l'enthalpie de van't Hoff ( $\Delta H_{vH}$ ), l'entropie ( $\Delta S$ ) et la variation de capacité calorifique ( $\Delta C_p$ ) associées à la transition (Fig. 2.6).



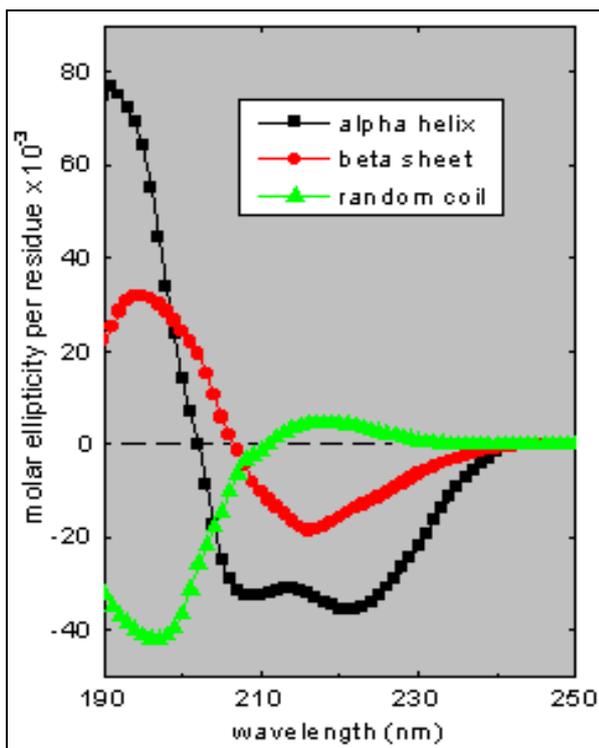
**Fig. 2.6. L'enthalpie de dénaturation d'une protéine (Stabilité) mesurée par DSC (Yoo *et al.*, 2017).**

La DSC est largement utilisée pour la recherche et le développement de composés biothérapeutiques afin de caractériser et sélectionner les candidats protéiques potentiels les plus stables, étudier les interactions ligand / récepteur...etc.

#### 2.4.2. Dichroïsme circulaire (CD)

Le dichroïsme circulaire (CD) est une technique spectroscopique d'analyse largement utilisée pour étudier la conformation des acides nucléiques et des protéines. Dans les cas des protéines, les spectres enregistrés au cours d'une expérience de CD contiennent des informations qui renseignent sur la nature des différents types de structures secondaires. L'allure d'un spectre CD dépend de l'organisation locale de la liaison peptidique, ce qui permet donc de distinguer les hélices  $\alpha$ , les feuilletts  $\beta$  (parallèles ou antiparallèles), " coudes " et structures désordonnées. Le traitement des données par des logiciels appropriés permet une estimation de la quantité respective des différentes conformations qui participent à

l'architecture tridimensionnelle. C'est une technique non destructive qui permet donc d'étudier les changements de conformation des protéines dans différents environnements (pH, dénaturation par les détergents, les agents réducteurs et la température), d'étudier le repliement des protéines en suivant l'apparition de la structure secondaire et pour l'étude des intermédiaires de repliement. La stabilité thermodynamique des protéines peut être suivie par le CD en déterminant les différentes constantes thermodynamiques; la température médiane de transition thermique ( $T_m$ ), l'enthalpie ( $\Delta H$ ) et l'entropie ( $\Delta S$ ). Le CD est basé sur la capacité des molécules chirales d'absorber différemment la lumière polarisée circulairement à droite et la lumière polarisée circulairement à gauche. Les coefficients d'extinction molaires pour la lumière gauche et droite sont différents pour ces molécules chirales.



**Fig. 2.7. Spectre caractéristique des structures secondaires protéiques (hélice  $\alpha$ , feuillet  $\beta$  et coude).**

Le spectre CD correspond à la différence d'absorbance mesurée en milli degrés (m deg) pour chaque longueur d'onde. En étudiant l'évolution du spectre CD au cours du temps, on peut donc suivre directement les changements de conformations moléculaires d'une protéine en le comparant au spectre de référence caractéristique des différentes structures secondaires protéiques (Fig. 2.7), et ainsi mieux comprendre certains mécanismes naturels, tels que la dynamique protéique et la stabilité protéique.

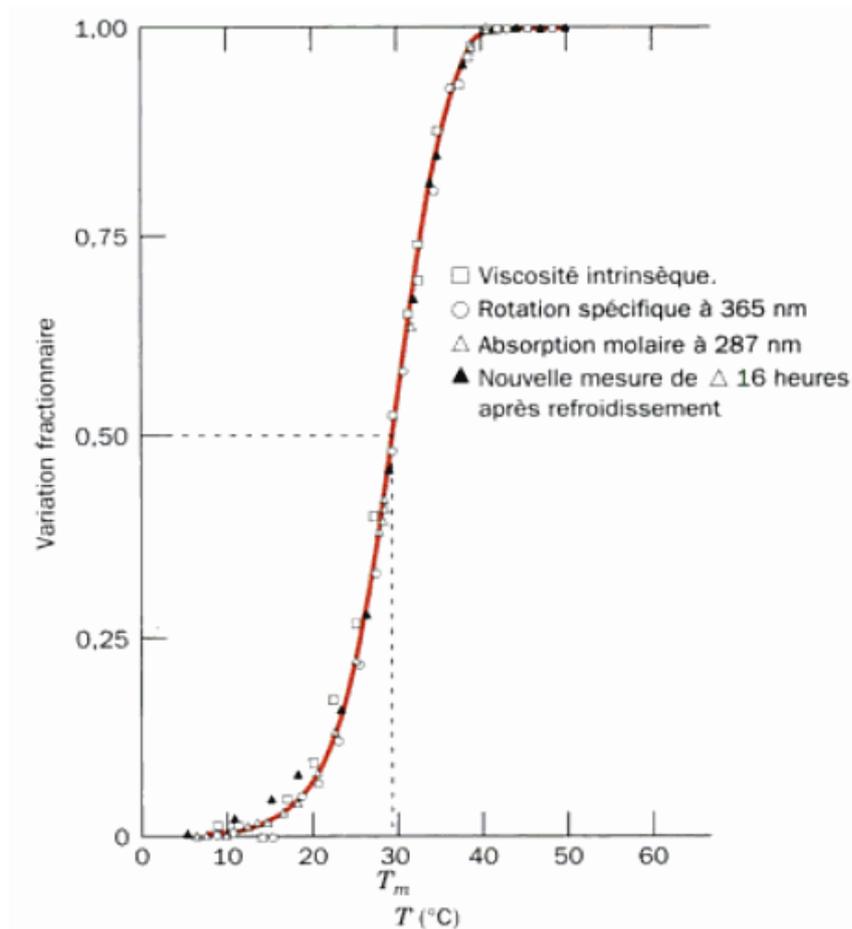
En effet, une étude dynamique en présence de dénaturant des protéines peut être réalisée grâce au dichroïsme circulaire. Lorsque la température augmente et que les structures sont détruites, le signal spectroscopique associé à ces structures protéiques tend à disparaître.

### Chapitre 3. Dénaturation des protéines

On peut définir la dénaturation protéique comme étant tout processus physique ou chimique qui modifie la structure native de la protéine et la rend incapable de remplir sa fonction biologique normale. Les premières études démontrant la réversibilité du processus de dénaturation des protéines ont apparues à partir de 1930. Que la dénaturation soit réversible dépend de la protéine et de l'ampleur de la dénaturation. Par exemple: chauffer un blanc d'œuf est la dénaturation irréversible alors que l'ondulation permanente des cheveux est la dénaturation réversible. D'un point de vu structural, si les modifications structurales sont discrètes, la dénaturation peut être réversible, mais si la protéine est incapable de reprendre la conformation native, la dénaturation est alors irréversible.

Les critères de dénaturation protéique sont : la perte d'activité biologique, l'insolubilisation de la protéine due à l'agrégation en amas et l'élévation de la viscosité des solutions protéiques. Selon le contexte biologique dans lequel se trouve une protéine, la dénaturation de cette dernière peut avoir des conséquences biologiques qui sont : la perte d'activité enzymatique (mort), la destruction des toxines, la digestibilité améliorée, la perte de solubilité et les changements de texture.

Les stabilités conformationnelles fragiles des protéines natives les rendent sensibles à la dénaturation par modification de l'équilibre des forces faibles non liantes qui maintiennent la conformation native. Quand on chauffe une protéine en solution, ses propriétés qui dépendent de la conformation, comme le pouvoir rotatoire, la viscosité et l'absorption en UV, changent brusquement dans une zone étroite de température (Fig. 3.1). La température de demi-dénaturation est appelée température de fusion,  $T_m$  de la protéine, par analogie avec la fusion d'un solide. La plupart des protéines ont des valeur de  $T_m$  bien au dessus de 100 °C. Outre les températures élevées, différentes conditions et diverses substances dénaturent les protéines :

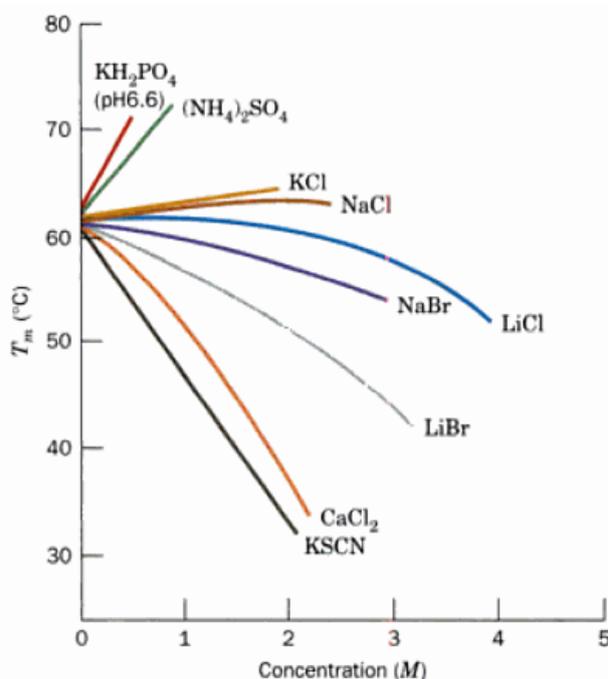


**Fig. 3.1. Dénaturation d'une protéine** (Voet *et al.*, 2016).

- Les variations de pH modifient les états d'ionisation des chaînes latérales d'acides aminés, qui changent la répartition des charges et les besoins en liaisons hydrogène de la protéine.
- Les détergents, dont certains perturbent sérieusement les structures protéiques à des concentrations aussi faibles que  $10^{-6}$  M, s'associent par interactions hydrophobes avec les résidus non polaires de la protéine, interférant ainsi avec les interactions hydrophobes impliquées dans la structure native de la protéine.
- Des concentrations élevées de substances organiques solubles dans l'eau, comme les alcools aliphatiques, interfèrent avec les forces hydrophobes qui stabilise les structures protéiques et ce à cause de leurs propres interactions hydrophobes avec l'eau. Les substances organiques ayant plusieurs groupes hydroxyles, comme l'éthylène glycol ou le saccharose, sont cependant de médiocres agents dénaturants car leur possibilité d'établir des liaisons hydrogène les rend

moins aptes à déstabiliser la structure de l'eau. Les agents dénaturant peuvent être classés selon leur nature physique ou chimique. Les effets des agents physiques sont : (i) l'élévation de la température : rompt les liaisons hydrogène (cuisson, stérilisation, ...), (ii) les radiations UV: photolyse des ponts disulfures (radiations ionisantes) et (iii) pH extrêmes : rupture de liaisons électrostatiques et hydrogène (précipitation par les acides).

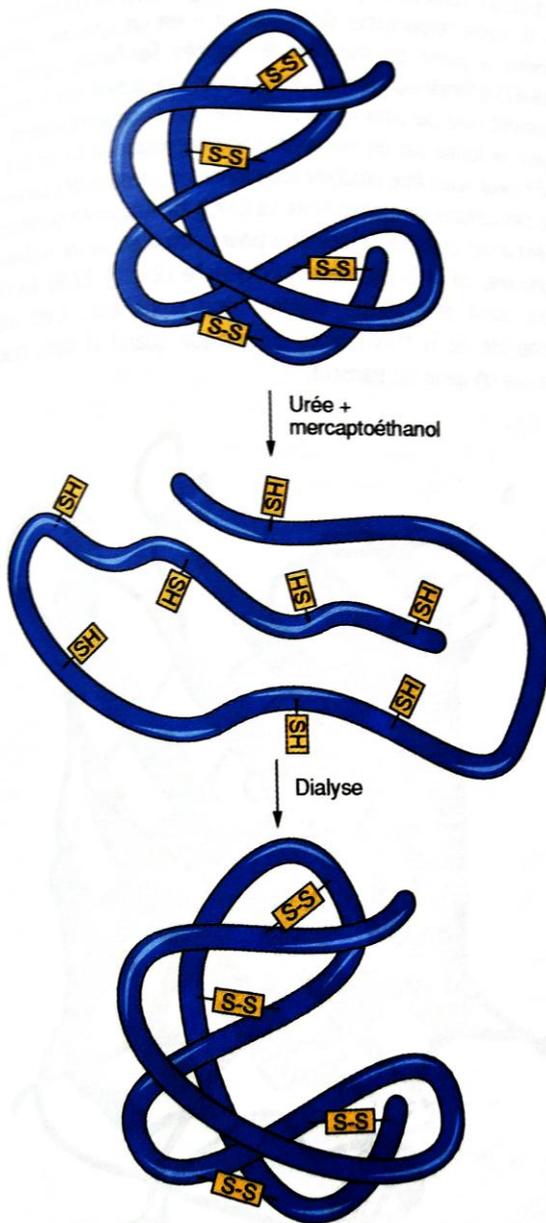
Les effets des agents chimiques sont : (i) les solvants organiques ( $\text{CCl}_4$ ), (ii) les réactifs rompant les ponts disulfures : qui peuvent être réversibles (réducteurs : BME, DTT) ou irréversibles (oxydants : chlorates) et (iii) l'urée et Chlorure de guanidinium : ce sont des petites molécules en solution concentrée qui induisent une dénaturation réversible (formation de liaisons H) et (iiii) les détergents : induisant une dissociation des structures tertiaire et quaternaire (effet réversible sans précipitation). la concentration de dénaturant nécessaire à la déstructuration de la protéine permet d'en estimer la stabilité. Plus la protéine est stable, plus elle est difficile à dénaturer et plus la concentration en agent dénaturant doit être grande.



**Fig. 3.2. Température de fusion de la Ribonucléase A (ARNase A) en fonction des concentrations en différents sels** (Voet *et al.*, 2016).

Par ailleurs, l'influence des sels sur les protéines est plus aléatoire. L'expérience de dénaturation par les sels réalisée sur la protéine Ribonucléase A (RNAase A) de pancréas de bœuf, qui est une enzyme de structure globulaire clivant les acides ribonucléiques, a montré que certains sels, comme  $(\text{NH}_4)_2\text{SO}_4$  et  $\text{KH}_2\text{PO}_4$ , stabilisent la structure de la protéine native (la  $T_m$  augmente); d'autres, comme KCl et NaCl, ont peu d'effet; d'autres encore, comme KSCN et LiBr, la déstabilisent (Fig. 3.2).

Une autre expérience réalisée sur la même enzyme, mais cette fois-ci en chauffant à 100°C ou en lui ajoutant de l'urée, on peut abolir la structure spatiale ordonnée de la protéine native. La chaîne polypeptidique est alors sous la forme d'une pelote statistique (random coil) parfaitement désordonnée (dénaturation protéique) (Fig. 3.3).



**Fig. 3.3. Dénaturation de la ribonucléase A** (Voet *et al.*, 2016).

L'urée est un chaotrope courant : elle augmente la solubilité des groupes hydrophobes dans l'eau et favorise le dépliement des protéines globulaires. Seuls les ponts disulfures marquent encore les points de contact des segments polypeptidiques de la structure ordonnée; on peut les détruire en ajoutant un réducteur comme par exemple le  $\beta$ -mercaptoéthanol. Après dénaturation, la RNAase perd toute activité enzymatique. Si l'on élimine l'urée et le  $\beta$ -mercaptoéthanol par dialyse, il y a repliement spontané vers la conformation native : l'activité enzymatique est alors régénérée. Une protéine reprendra donc sa conformation authentique; cette information doit être contenue dans la structure primaire (renaturation). Dans le cas de nombreuses protéines, la dénaturation est dans la pratique irréversible : l'exemple le plus connu est le blanc d'œuf dur. Le rôle déterministe de la structure primaire explique aussi pourquoi des changements en une position de la chaîne polypeptidique peuvent gêner le processus de repliement si durablement.

## Chapitre 4. Repliement des protéines *in vivo*

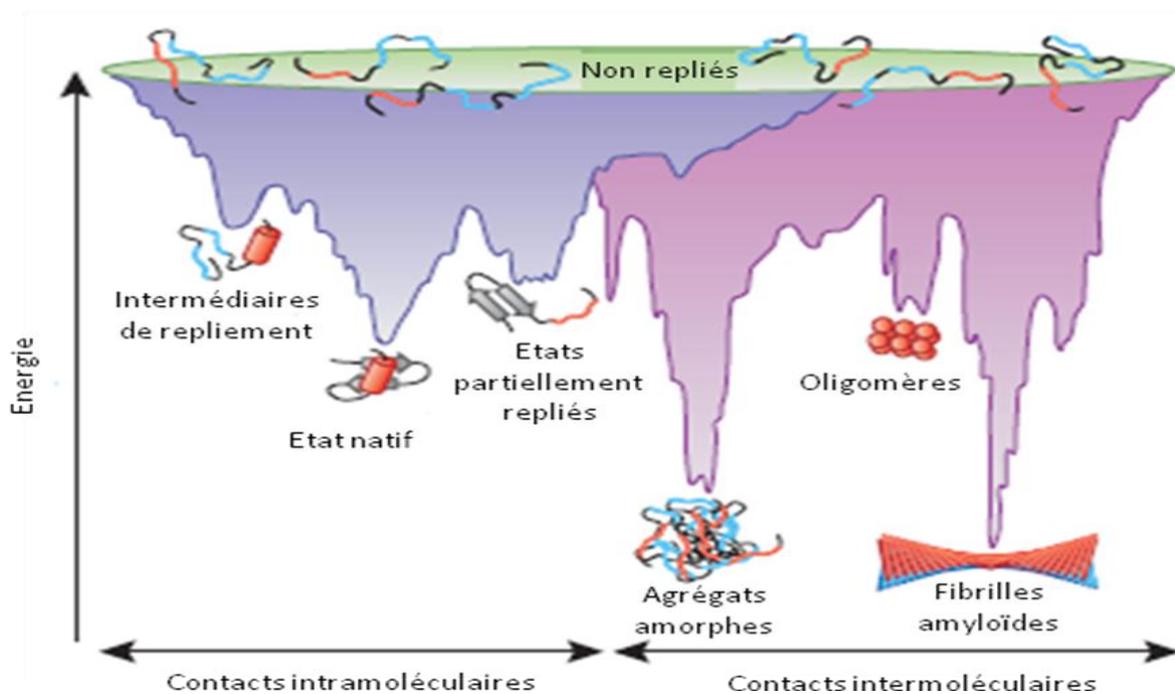
### 4.1. Mécanisme de repliement

Il a été démontré que de nombreuses protéines se replient spontanément *in vitro*, confirmant ainsi l'analyse d'Anfinsen selon laquelle la séquence linéaire de la chaîne polypeptidique contient toutes les informations nécessaires pour spécifier la structure tridimensionnelle d'une protéine. Bien que le repliement des protéines ait fait l'objet d'études approfondies pendant près de 50 ans, la manière dont le repliement final et le processus de repliement est déterminé par la séquence d'acides aminés reste l'un des problèmes les plus importants en biologie.

En outre, des études ont montré récemment que, dans la cellule, une grande partie des protéines nouvellement synthétisées ont besoin de l'aide de chaperons moléculaires pour atteindre leurs états repliés de manière efficace et dans un délai biologiquement pertinent. Il est clair que les protéines dans l'éprouvette et dans la cellule sont soumises aux mêmes lois de la physique. Quelle est donc la particularité du repliement dans des conditions cellulaires et pourquoi des chaperons sont-ils nécessaires? La disponibilité croissante de techniques biophysiques très sensibles pour étudier le repliement *in vitro* et dans les systèmes cellulaires fournit maintenant de nouvelles informations sur ces questions. Ces études ont également mis en lumière sur le processus d'agrégation (mal conformation), une réaction potentiellement dangereuse en dehors de la trajectoire qui peut provoquer une maladie et qui doit être prévenue par des chaperons moléculaires.

Les intermédiaires de repliement sont la règle pour les protéines les plus grandes de >100 acides aminés (environ 90% de toutes les protéines d'une cellule), qui ont une plus grande tendance à s'effondrer rapidement en solution aqueuse dans des conformations compactes non natives. Comme cela a été démontré récemment par une combinaison de techniques spectroscopiques sensibles, même les petites protéines qui se replient sur une échelle de temps inférieure à la seconde peuvent passer par des intermédiaires structurels en route vers l'état

natif. De tels intermédiaires représentent soit des «tremplins» sur le chemin vers l'état natif, soit des conformations mal repliées, cinétiquement stables, pouvant nécessiter une réorganisation importante avant que l'état natif ne puisse être atteint. La formation d'interactions métastables non indigènes au cours du repliement est interprétée comme une conséquence de la robustesse du paysage énergétique du repliement (Fig. 4.1), que les protéines soient supposées se replier par de multiples voies de descente dans l'entonnoir ou par des voies préférentielles définies par l'assemblage séquentiel d'unités de repliement élémentaires, appelées «foldons».



**Fig. 4.1. Schéma de paysage énergétique du repliement et de l'agrégation des protéines** (adapté selon Hartl *et al.*, 2009). La surface mauve montre la multitude de conformations «en train de s'amener» à l'état natif via des contacts intramoléculaires et la zone rose montre les conformations se dirigeant vers des agrégats amorphes ou des fibrilles amyloïdes via des contacts intermoléculaires. Les deux parties de la surface d'énergie se chevauchent. La formation d'agrégats peut se produire à partir d'intermédiaires peuplés lors du repliement *de novo* ou par déstabilisation de l'état natif en états partiellement repliés. Elle est normalement empêchée par les chaperons moléculaires. Des oligomères toxiques pour les cellules peuvent apparaître en tant qu'intermédiaires hors de la voie de formation de fibrilles amyloïdes.

Des exemples de tels motifs de nucléation minimaux sont les motifs à hélice double brin trouvés dans les protéines du domaine  $\alpha/\beta$ . Un comportement de repliement multi-états avec

des intermédiaires peuplés serait observé lorsque plusieurs foldons sont séparés et n'agissent pas de manière coopérative ou en cas de mauvais assemblage des foldons. La tendance à mal replier augmente avec les types de plis topologiquement complexes qui sont stabilisés par des interactions à longue portée (par exemple, les architectures de domaine  $\alpha/\beta$ ) ou lorsque les protéines contiennent plusieurs domaines séparés à l'état natif mais pouvant interagir pendant le repliement.

Les états partiellement repliés ou mal repliés ont souvent tendance à s'agréger, en particulier lorsqu'ils représentent des pièges cinétiques majeurs dans la voie du repliement. Cela est dû au fait que ces formes exposent généralement des résidus d'acides aminés hydrophobes et des régions du squelette polypeptidique non structuré, des caractéristiques qui sont pour la plupart enfouies à l'état natif. Comme le repliement intramoléculaire, l'agrégation (l'association de deux molécules de protéine non natives ou plus) dépend en grande partie des forces hydrophobes et conduit principalement à la formation de structures amorphes. L'agrégation peut conduire à la formation d'agrégats fibrillaires hautement ordonnés appelés amyloïdes, dans lesquels les brins  $\beta$  s'étendent perpendiculairement au long axe de la fibrille (structure croisée  $\beta$ ). Bien qu'apparemment limitées à un sous-ensemble de protéines dans des conditions physiologiques, ces structures hautement stables sur le plan thermodynamique sont accessibles à de nombreuses protéines dans des conditions dénaturantes, largement indépendantes de la séquence, ce qui suggère que leur formation est une propriété inhérente de la chaîne polypeptidique. La formation de fibrilles amyloïdes est généralement toxique pour les cellules et peut provoquer certaines des maladies neurodégénératives.

#### 4.2. Comparaison entre le repliement des protéines *in vitro* et *in vivo*

Les protéines se replient initialement *in vivo* lors de leur biosynthèse. Par conséquent, le premier environnement auquel ils sont soumis est créé par les enzymes et les chaperons associés au ribosome et au ribosome. De plus, les chaînes peuvent se replier de manière co-

traductionnelle avant que toute la chaîne ait été faite. Par contre, le repliement *in vitro* de protéines est généralement initié à partir d'un ensemble non replié dans lequel une population de chaînes complètes (ou dans le cas d'expériences sur une seule molécule, un polypeptide) est soumise à des conditions de repliement. Ainsi, la possibilité de repliement en co-traduction constitue une différence majeure entre la réaction de repliement *de novo* dans l'éprouvette et dans un organisme vivant.

Deuxièmement, les protéines *in vitro* échantillonnent leur état non replié dans un équilibre dynamique régi par leur stabilité thermodynamique. Que les protéines passent beaucoup de temps dépliées *in vivo* est pas clair. De nombreux facteurs peuvent être défavorables à l'accumulation de toute population importante à l'état non replié, notamment la liaison de chaperons, la dégradation en cours et les barrières cinétiques.

Troisièmement, des expériences de repliement de protéines *in vitro* sont effectuées à haute dilution. Alors qu'*in vivo*, les concentrations de macromolécules vont de 200 à 400 mg/ml et les surfaces présentes tout autour d'une chaîne de repliement sont hautement interactives.

Quatrièmement, alors que les protéines se replient sur elles-mêmes *in vitro*, une fraction significative d'entre elles ont des "auxiliaires" *in vivo*. On ne sait toujours pas dans quelle mesure et comment les chaperons modifient l'énergie de repliement fondamentale et le paysage des protéines.

Cinquièmement, les protéines sont vulnérables aux réactions d'agrégation intermoléculaires concurrentes dans une mesure qui dépend assez facilement *in vitro* de la concentration d'espèces sujettes à l'agrégation.

Enfin, les réactions de repliement *in vivo* sont organisées dans l'espace, de telle sorte que certaines interactions seront préférées à d'autres. *In vitro*, il est très difficile d'imiter un environnement hétérogène organisé dans l'espace. De même, le repliement des protéines dans

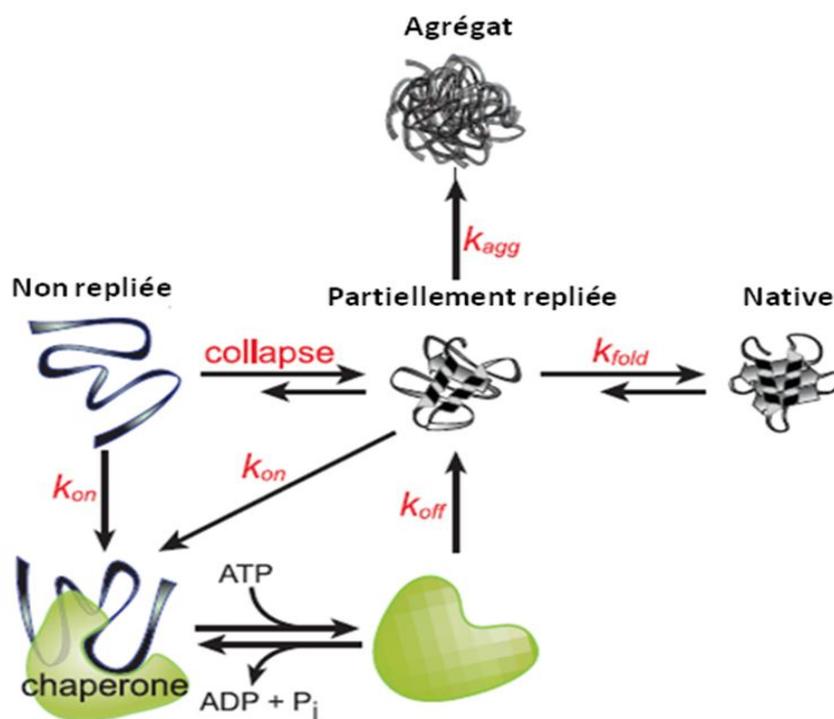
les organites, en particulier le réticulum endoplasmique et la voie de sécrétion, dépend de manière critique de la compartimentation et de l'organisation sub-cellulaire.

Bien que la structure primaire d'une protéine impose sa structure tridimensionnelle, de nombreuses protéines ont besoin de l'aide de protéines auxiliaires, comme les protéines chaperons, la protéine disulfure isomérase (PDI) et la protéine (Dsb) qui sont des enzymes assurant la formation ou le clivage des ponts disulfure et les peptidyl prolyl *cis-trans* isomérases qui assurent l'isomérisation des résidus prolines, pour se replier et s'assembler en structure native quaternaire assurant ainsi la fonction cellulaire.

### 4.3. Protéines chaperons

Les protéines chaperons ou chaperons moléculaires sont un groupe fondamental de protéines qui sont des composants essentiels du mécanisme de qualité des protéines dans la cellule. Cette machinerie assure que le processus de repliement de toute chaîne polypeptidique nouvellement synthétisée entraîne la formation d'une protéine correctement repliée et que la protéine repliée soit maintenue dans une conformation active pendant toute sa durée de vie fonctionnelle. Il a été démontré que les protéines chaperons jouent un rôle essentiel dans la viabilité cellulaire, à la fois dans des conditions normales et dans des conditions de stress. Les chaperons peuvent également aider au déploiement et à la dégradation des protéines mal repliées ou régulatrices et à la désagrégation des agrégats de protéines préformés. Les chaperons sont impliqués dans d'autres fonctions cellulaires, notamment la translocation de protéines à travers les membranes, des événements de fusion vésiculaire, la sécrétion de protéines et l'assemblage d'oligomères. Il en résulte un intérêt croissant pour le rôle des protéines chaperons dans plusieurs maladies, notamment les cancers. Ces dernières années, notre compréhension de la biologie, de la biochimie et de la biophysique de la fonction des chaperons moléculaires a considérablement progressé. De plus, les récents développements techniques dans les domaines de la protéomique et de la génomique nous ont permis d'obtenir

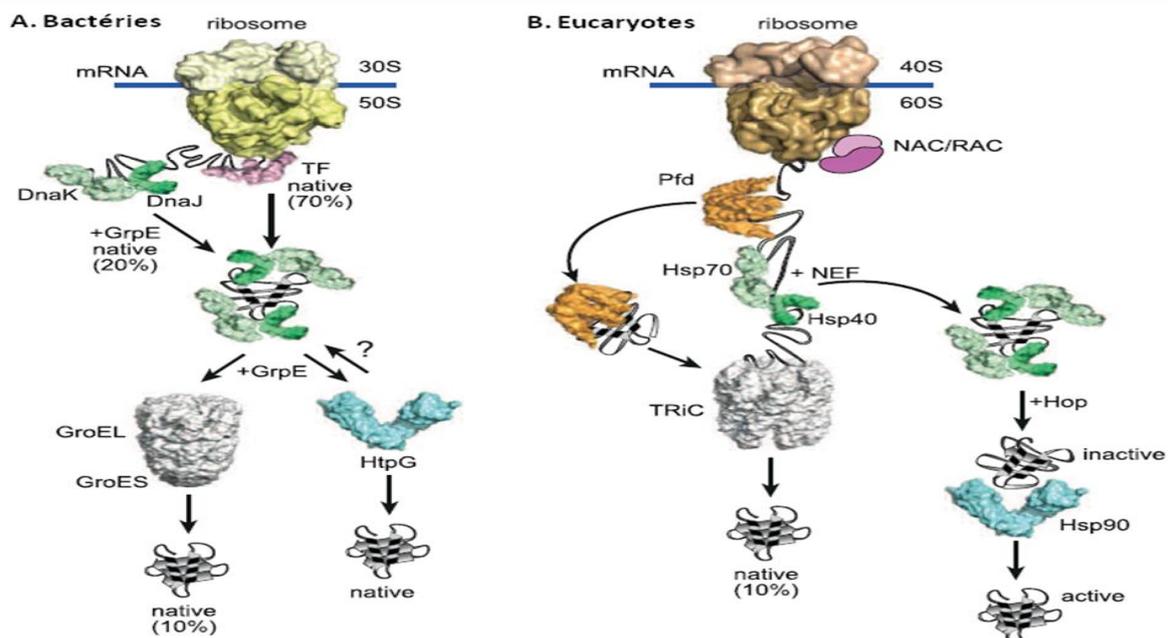
une vision globale des réseaux d'interaction des protéines chaperons. Les chaperons moléculaires qui fonctionnent largement dans le repliement *de novo* comprennent les chaperonines (Hsp60) et les systèmes Hsp70 et Hsp90, les protéines de choc thermique. Ils reconnaissent les résidus d'acides aminés hydrophobes exposés par des protéines non natives et favorisent le repliement au travers de cycles de liaison et de libération de protéines régulés par l'adénosine 5'-triphosphate (ATP). Dans ce mécanisme de partitionnement cinétique, la libération d'éléments hydrophobes permet le repliement, tandis que la (re)liaison des protéines non natives bloque l'agrégation et peut inverser les états mal repliés (Fig. 4.2).



**Fig. 4.2. Mécanisme de chaperon favorisant le repliement par partitionnement cinétique** (adaptée selon Balchin *et al.*, 2016).

Les chaperons moléculaires favorisent le repliement des protéines par un mécanisme de partitionnement cinétique des états non natifs. De nombreux chaperons utilisent la liaison ATP et l'hydrolyse pour passer d'un état de faible à haute affinité aux intermédiaires de repliement (déplié, partiellement replié), exposant des résidus d'acides aminés hydrophobes. Relier les blocs de chaperons à l'agrégation et réduire la concentration des molécules agrégées-sujettes. Le rejet de chaperons permet le repliement (enfouissement de résidus hydrophobes). Un repliement efficace se produit lorsque la constante de vitesse du repliement ( $k_{fold}$ ) est plus rapide que la liaison des intermédiaires du repliement au chaperon ( $K_{on}$ ) et l'agrégation ( $k_{agg}$ ) est plus lente que  $K_{on}$ .

Plusieurs familles de chaperons moléculaires conservées au cours de l'évolution guident les protéines le long de voies de repliement productives, en évitant et en inversant parfois les erreurs de repliement et d'agrégation. Leurs membres sont souvent appelés protéines de stress ou protéines de choc thermique (Hsps) car ils sont régulés positivement dans des conditions de stress conformationnel. Les principales familles de chaperons sont classées par poids moléculaire (Hsp40, Hsp60, Hsp70, Hsp90, Hsp100 et le petit Hsp). Les principes organisationnels des voies cytosoliques des chaperons sont hautement conservés. Pour la plupart des protéines trop grosses pour se replier en association avec le ribosome, les chaperons retardent le compactage de la chaîne, prévenant les erreurs de repliement, jusqu'à ce que suffisamment d'éléments structuraux soient disponibles pour que le repliement soit productif. Les chaperons liant les ribosomes [Facteur de déclenchement (TF), dans les bactéries; le complexe associé au ribosome (RAC) et le complexe associé à la chaîne naissante (NAC) chez les eucaryotes] interagissent d'abord avec le polypeptide naissant, suivis par les chaperons n'ayant pas d'affinité directe pour le ribosome, y compris le système classique Hsp70 (DnaK / DnaJ chez les bactéries) ; Hsp70 / Hsp40 chez les eucaryotes) (Fig. 4.3). La fin du repliement peut être réalisé par Hsp70 ou nécessiter un transfert vers la chaperonine (GroEL / ES chez les bactéries; TRiC chez eucaryotes) ou le système Hsp90 (HtpG chez les bactéries). Les différentes machineries chaperones interagissent directement ou utilisent des protéines adaptatrices spécifiques pour faciliter le transfert des protéines. De cette manière, la protéine de repliement est constamment protégée des interactions aberrantes et de la dégradation prématurée du cytosol. Les protéines de petit choc thermique (sHsps) indépendantes de l'ATP agissent dans une agrégation tampon et coopèrent avec Hsp70 dans des réactions de repliement. En plus de ces systèmes chaperons espacés, des chaperons d'assemblage spécifiques peuvent aider à la formation de complexes protéiques oligomères en interagissant avec leurs sous-unités repliées.



**Fig. 4.3. Les voies chaperon du repliement des protéines dans le cytosol**

(adapté selon Balchin *et al.*, 2016). Pour environ 70% des protéines bactériennes (A) et eucaryotes (B), le ribosome et les facteurs associés [facteur déclencheur (TF) chez les bactéries; le complexe associé aux chaînes naissantes (NAC) et le complexe associé aux ribosomes (RAC) chez les eucaryotes] fournissent une aide au repliement. En aval du ribosome, Hsp70 (DnaK chez les bactéries) est la plaque tournante du réseau chaperon. Hsp70 coopère avec les Hsp40 (DnaJ chez les bactéries) et les facteurs d'échange de nucléotides (NEFs; GrpE chez les bactéries) pour replier environ 20% du protéome. Les 10% restants du protéome sont transmis aux chaperonines pour le repliement: GroEL / ES chez les bactéries et TRiC chez les eucaryotes. Chez les eucaryotes, certaines protéines naissantes sont transférées directement vers TRiC par prefoldin (Pfd). Cette voie alternative domine dans les archées, qui manquent des protéines Hsp70. Les eucaryotes utilisent également le système chaperon Hsp90 pour catalyser l'activation de protéines métastables telles que les kinases et les facteurs de transcription. Les protéines sont transférés vers Hsp90 via le système Hsp70 et le cochaperone Hop. Le Hsp90 bactérien, HtpG, agit plus généralement dans le repliement et fonctionne sans cochaperones connues.

La vision actuelle des mécanismes de l'activité des chaperons tend à se conformer à l'hypothèse du paysage énergétique en forme d'entonnoir. Les chaperons sont considérés comme des auxiliaires qui aident les protéines à se replier en empêchant leur capture dans les pièges cinétiques non natifs. Les intermédiaires de repliement piégés cinétiquement sont la règle pour les protéines plus grosses, ce qui présente un risque de repliement dû à la formation d'interactions non négatives. Les étapes lentes du repliement menant à l'accumulation des intermédiaires de repliement incluent l'isomérisation du prolyle et la formation de liaisons disulfure, qui peuvent être accélérées par les prolyl-isomérases et la protéine disulfure isomérases, respectivement.

#### 4.4. Formation des ponts disulfure

Les protéines possédant des ponts disulfures peuvent être soit des protéines dont la liaison cystéine-cystéine est indispensable à la stabilité de leur conformation native ou bien des protéines qui possèdent un couple de cystéines alternant entre une forme oxydée (présence du pont) et une forme réduite.

La formation d'un pont disulfure entre deux groupements sulfhydriles (-SH) est une réaction d'échange de deux électrons entre un oxydant et un accepteur d'électron. Bien que les ponts disulfures peuvent se former spontanément en présence d'oxygène mais le processus de formation de ponts disulfures est extrêmement lent (> de 15min) *in vitro* et extrêmement plus rapide et précis *in vivo* (quelques minutes). L'observation de ces différences entre les vitesses de réaction observées *in vivo* et *in vitro* a permis de découvrir le premier catalyseur de la formation des ponts disulfures, la Protein Disulfide Isomerase (PDI). PDI est une protéine eucaryote présente dans le réticulum endoplasmique. Elle fait partie du réseau complexe des protéines impliquées dans la formation des ponts disulfures.

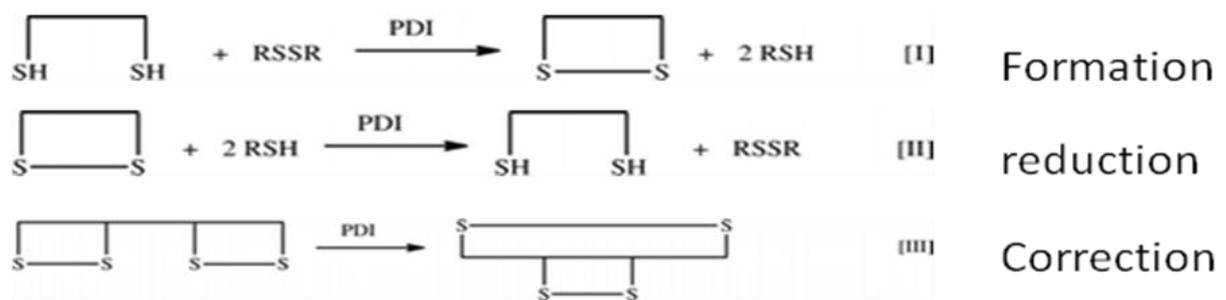
La formation des ponts disulfure est une réaction d'échange thiol-disulfure entre des thiols libres (-SH) et une molécule avec un pont disulfure (S-S) déjà formé. Cet échange peut avoir lieu entre une protéine et un substrat contenant des groupements sulfhydriles, telles de petites molécules contenant des thiols comme le glutathion ou des protéines contenant des ponts disulfures. La vitesse de la réaction va dépendre de l'accessibilité des sites actifs, de la différence de potentiels redox entre les partenaires redox et de la probabilité des atomes de soufre à pouvoir se placer à la distance nécessaire pour que l'échange thiol-disulfure puisse avoir lieu. Cette réaction a lieu dans le réticulum endoplasmique chez les cellules eucaryotes et dans l'espace périplasmique chez les cellules procaryotes.

Parmi les enzymes impliquées dans les échanges thiol-disulfure, la famille des thiorédoxines qui est composée de thioltransférases dont les mieux caractérisées sont les Thiorédoxine

TRX1 et TRX2 et les Glutarédoxine GRX1, GRX2, GRX3. Cette famille comprend également des protéines eucaryotes qui appartiennent à la famille des PDI, comme Erp57 et des protéines bactériennes, comme les enzymes DsbA, DsbC, DsbD, DsbG et DsbE d'*E.coli*, impliquées dans la synthèse de ponts disulfures des protéines extracytoplasmiques.

Les enzymes PDIases appartiennent à un grand groupe ubiquitaire de protéines possédant les propriétés suivantes:

- il y a peu de similarité de séquence entre les différentes protéines.
- activité enzymatique globale: formation de liaisons disulfure correctes.
- Formation des ponts disulfure (I) et réduction des ponts disulfure (II) et correction des liaisons disulfure (III) (Fig. 4.4).



**Fig. 4.4. Activité enzymatique des Protein Disulfide Isomerases (PDI).**

#### 4.5. Isomérisation des prolines

En 1973, il a été constaté que la RNase A complètement non repliée (dénaturée) utilisant 6M de Chlorure de guanidinium (GdmCl) suivait des réactions de repliement rapides et lentes avant d'arriver à l'état de protéine native (renaturation). Il a été conclu qu'il existait deux formes différentes de RNase A non pliée; repliement rapide ( $U_f$ ) et repliement lent ( $U_s$ ). Il n'était pas clair pourquoi ces formes existaient et quelle était la différence entre elles.

John Brandts et ses collaborateurs avaient proposé un modèle d'isomérisation de la proline pour étudier les formes de repliements rapide ( $U_f$ ) et lente ( $U_s$ ) de la protéine RNase A. La RNase A native comporte deux résidus proline *trans* et deux *cis*. Il a été suggéré que, lors du dépliage rapide de la RNase A, il s'est formé une forme de repliements rapide ( $U_f$ ) qui avait

un état *cis* alors qu'il était à l'état *trans* en une forme de repliements lent ( $U_s$ ). La *trans*-isomérisation de la proline était un processus très lent qui était responsable de la formation de deux états non repliés. Ce modèle était très attrayant, mais il ne pouvait pas fournir d'autres preuves à l'appui. Par la suite, il a été découvert que le repliement de la RNase A était beaucoup plus complexe et impliquait un certain nombre d'états à repliement lent et rapide dans la voie de repliement.

Par la suite en 1978, grâce aux expériences d'étude de repliement menée par Schmid *et al.*, pour mesurer le rapport  $[U_f] / [U_s]$  à différents moments après l'ajout de l'acide chlorique ( $HClO_4$ ). Ses expériences ont confirmé que la différence entre  $U_f$  et  $U_s$  n'était due qu'à l'isomérisation de la proline. Il avait également découvert que le repliement des protéines pouvait être accéléré en ajoutant du sel de stabilisation tel que l'Ammonium sulfate,  $(NH_4)_2SO_4$ . Ses expériences ont été conclues avec les faits que : les espèces à repliement lent ( $U_s$ ) de la RNase A ont été repliées en intermédiaires de type natif ( $I_N$ ), tandis que les espèces à repliement rapide ( $U_f$ ) ont été repliées à leur état (N). L'état natif (N) et intermédiaire de type natif ( $I_N$ ) se lient à 2'CMP (l'inhibiteur spécifique de la RNase A, utilisé dans ces expériences) avec des spécificités similaires. Ces expériences ont prouvé la présence d'intermédiaires structurels de repliement lors du repliement de protéines. Par conséquent, on peut en conclure que le repliement des protéines peut se faire par d'autres voies qui confortent l'argument de Levinthal selon lequel le repliement se fait par la voie la plus rapide si des voies de substitution sont disponibles.

Toutes les liaisons peptidiques adoptent la configuration *trans*, à l'exception de certaines liaisons X-Pro, où X représente un résidu quelconque, qui peuvent prendre la configuration *cis*. La *cis*-proline a un rôle structural important car elle favorise la formation de coudes. L'isomérisation spontanée de la liaison X-Pro est lente en raison du caractère de double liaison partielle de cette dernière et elle serait une étape limitante du repliement des chaînes

polypeptidiques si, *in vivo*, elle n'était accélérée par des peptidyl prolyl isomérases qui déforment la liaison peptidique de telle façon que les atomes C, O, et N ne soient plus coplanaires et que la barrière d'activation soit abaissée. De plus, la formation de liaisons disulfure correctes dans les protéines naissantes est accélérée, *in vivo*, par la disulfure isomérase, enzyme de large spécificité pour des substrats contenant des résidus cystéine. Ce dernier, en brassant les liaisons disulfure permet aux protéines naissantes de trouver rapidement les appariements thermodynamiquement les plus favorables.

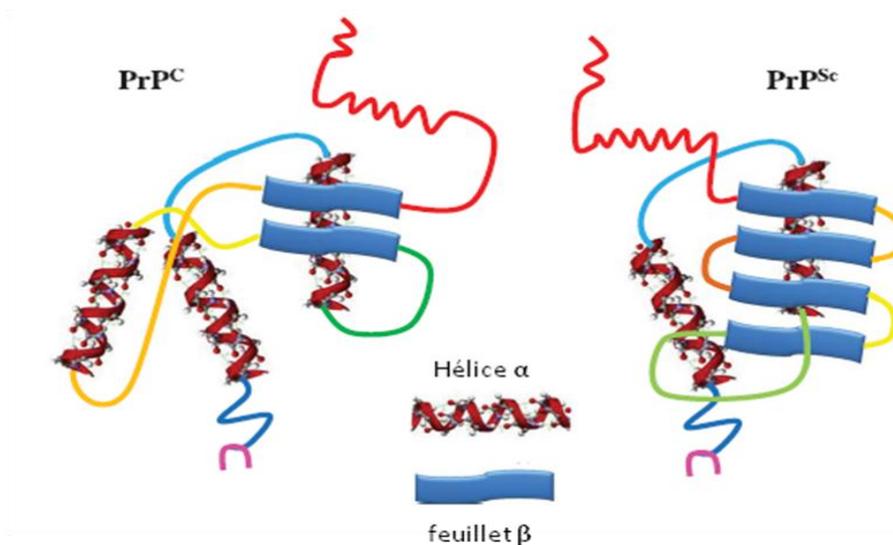
Enfin, au sein de la cellule, milieu très encombré, deux classes de chaperons moléculaires, les Hsp70 et les Hsp60 "chaperonines", orientent au niveau moléculaire le repliement des chaînes polypeptidiques en empêchant des conformations incorrectes ou des liaisons inappropriées avec d'autres constituants du milieu. Isomérases et chaperons ne font qu'assister le repliement des chaînes polypeptidiques; elles ne le dirigent pas et leur présence ne contredit en rien le dogme de la morphogenèse autonome qui seule détermine la conformation des protéines, comme celle des autres structures complexes des êtres vivants.

#### 4.6. Maladies conformationnelles

Le mauvais repliement des protéines a une importance pathologique fondamentale dans de nombreuses maladies. Il existe plus de dix huit maladies létales qui sont associées à l'altération de la structure secondotertiaire des protéines (repliement incorrect des protéines). Dans cette partie, nous montreront, à titre d'exemples, des maladies conformationnelles les plus connues, mais c'est loin d'être exhaustif. Dans la maladie neurodégénérative d'Alzheimer, on trouve des agrégats de protéines insolubles dans le cerveau des patients. A cause de leur ressemblance avec l'amidon (amylose), on les appelle "amyloïdes". L'agrégation d'un peptide  $\beta$ -amyloïde en polymères insolubles riche en feuillets  $\beta$  aboutit à la formation de plaques amyloïdes. Les conditions dans lesquelles le peptide  $\beta$ -amyloïde s'agrège sont encore peu claires. Autres maladies neurodégénératives comme la maladie de Creutzfeld-Jacob (maladie

de la vache folle), la tremblante du mouton et l'encéphalopathie spongiforme bovine (ESB) ont un rapport étroit avec le repliement incorrect et l'agrégation des protéines. Il existe un autre type de maladies lié à l'absence de protéine chaperon, tel que les  $\beta$ -thalassémies causés par des défauts génétiques qui touchent la synthèse de l'une des sous-unités polypeptidique de l'hémoglobine et en absence de la protéine chaperon spécifique, protéine  $\alpha$  stabilisatrice de l'hémoglobine (AHSP), les sous-unités de l'hémoglobine  $\alpha$  s'agrègent causant des effets cytotoxiques sur le développement érythrocytaire.

La protéine prion PrP impliquée dans l'encéphalopathie spongiforme transmissible comporte 219 acides aminés. Elle porte de nombreuses modifications, parmi lesquelles une ancre de glycoposphatidylinositol, qui maintient la protéine dans le feuillet extérieur de la membrane cellulaire. Le monomère «normal» du prion PrP<sup>C</sup> est essentiellement formé d'hélice  $\alpha$ ; une partie importante de sa forme «pathologique» PrP<sup>Sc</sup> est faite de feuillets  $\beta$  et forme des multimères insolubles, qui ont des effets toxiques. Au cours de ce processus, PrP<sup>Sc</sup> mal repliée peut pousser la PrP<sup>C</sup> «normale» à adopter la forme pathologique PrP<sup>Sc</sup> : nous avons ici affaire à un effet de dominos (Fig. 4.5).



**Fig. 4.5. Structure de la forme cellulaire normale de la protéine prion (PrP<sup>C</sup>) et de la forme pathologique (PrP<sup>Sc</sup>) (adaptée selon Balchin *et al.*, 2016).**

## Chapitre 5. Ingénierie de nouvelles interactions protéine-ligand

### 5.1. Interactions protéine-ligand et mécanismes de liaison

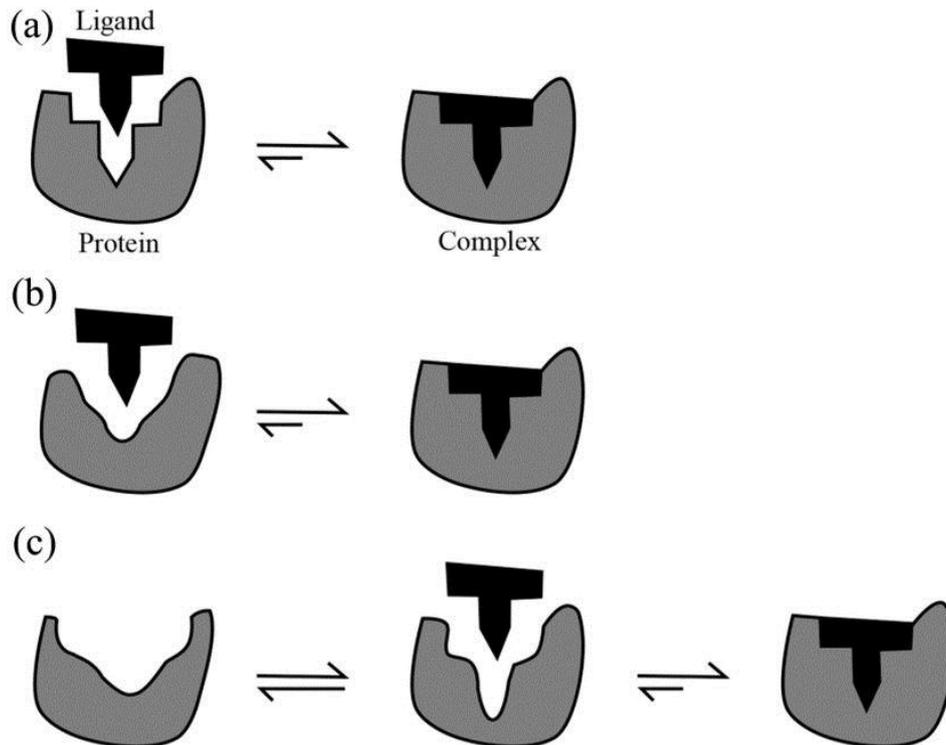
Un principe fondamental de tous les processus biologiques est l'organisation moléculaire et la reconnaissance, par exemple, les interactions protéine-ligand et la liaison. Essentiellement, les protéines remplissent leur fonction par interaction avec des molécules telles que des protéines et des peptides, des acides nucléiques, des ligands et des substrats, et d'autres petites molécules telles que l'oxygène ou les ions métalliques. Ainsi, une compréhension détaillée des processus biologiques nécessite d'étudier non seulement le mécanisme de repliement des protéines mais aussi le mécanisme de liaison protéine-ligand. Un ligand, par définition, est toute substance qui peut former un complexe non-covalent avec une biomolécule.

La reconnaissance moléculaire est un processus par lequel les macromolécules biologiques interagissent les unes avec les autres ou avec de petites molécules pour former un complexe spécifique. Il a deux caractéristiques importantes: la spécificité et l'affinité. La spécificité est que les macromolécules biologiques sont capables de distinguer le ligand hautement spécifique d'autres moins spécifiques. Afin de rendre une interaction spécifique, les liens entre les partenaires corrects devraient être forts, alors que pour les autres partenaires ne présentant que des différences mineures de structure, ils devraient être faibles ou même répulsifs. Par conséquent, une fois qu'un ligand correct se lie à sa molécule d'accepteur, l'affinité entre eux devrait être suffisamment forte pour empêcher la dissociation. L'aspect le plus important de la reconnaissance moléculaire est que la reconnaissance n'est généralement pas un processus en soi mais un élément d'un mécanisme plus complexe et fonctionnellement important comme la régulation allostérique de l'activité enzymatique, la transduction du signal, le repliement des protéines ou la formation de structures multi-sous-unités et supramoléculaires. Cela nécessite des changements importants et parfois spectaculaires dans les propriétés des partenaires en interaction.

## 5.2. Mécanismes de liaison protéine-ligand

Les mécanismes de liaison protéine-ligand ont évolué depuis les premiers «serrure et clé» jusqu'à « l'ajustement induit » et aux modèles de « sélection conformationnelle » (Fig. 5.1).

Le modèle de serrure et clé (Fig. 5a) a été proposé par (Fischer, 1894) pour expliquer la liaison d'un seul substrat à l'enzyme. Dans cette analogie, la serrure est l'enzyme et la clé est le substrat. La condition préalable de ce modèle est que l'enzyme et le substrat soient à la fois rigides et que leurs surfaces soient complémentaires, et donc seule la clé (substrat) correctement dimensionnée s'adapte dans le trou clé (site actif) de la serrure (enzyme); les clés de taille incorrecte ou mal positionnées (molécules de substrat de forme ou de taille incorrecte) ne rentrent pas dans la serrure (enzyme) car la clé et la serrure ne peuvent pas changer de forme et de taille. Cependant, toutes les preuves expérimentales ne peuvent pas être expliquées de manière adéquate par le modèle «serrure et clé». Par conséquent, une modification appelée le modèle d'ajustement induit a été proposée pour compenser la protéine et le ligand rigides, structurellement invariables. Le modèle d'ajustement induit (Fig. 5b) suppose que le ligand joue un rôle dans la détermination de la forme finale de la protéine, c'est-à-dire que la liaison du ligand induit un changement conformationnel de la protéine (Koshland, 1958). Cependant, il semble que le mécanisme d'ajustement induit ne convient que pour les protéines présentant un changement conformationnel mineur ou modéré après les liaisons du ligand. Bosshard *et al.*, en 2001 a suggéré que l'ajustement induit n'est possible que si la correspondance entre les sites en interaction est suffisamment forte pour fournir au complexe initial suffisamment de force et de longévité pour que l'ajustement induit prenne lieu dans un délai raisonnable. Par conséquent, le mécanisme d'ajustement induit seul ne peut pas bien expliquer l'association entre le ligand et les protéines subissant de grands changements conformationnels après la liaison.



**Fig. 5.1. Mécanismes de liaison protéine-ligand** (Tobi *et al.*, 2005).

(a) modèle de serrure et de clé. (b) ajustement induit. (c) sélection conformationnelle suivie d'ajustement induit.

Par la suite, le modèle de sélection conformationnelle (également appelé équilibre préexistant, sélection de la population, ajustement de la fluctuation ou ajustement choisi), (Fig. 5c) est apparu comme une alternative à l'ajustement induit et devient de plus en plus répandu. Essentiellement, un tel mécanisme est l'extension du modèle en entonnoir de repliement, postulant que l'état natif d'une protéine n'est pas défini par une seule conformation rigide mais plutôt par un ensemble de sous-produits conformationnels étroitement liés coexistant en équilibre, et le ligand peut sélectionner de se lier au conformère (isomères de conformation) le plus approprié parmi ces sous-états conformationnels, déplaçant l'équilibre vers la formation complexe. Comme décrit ci-dessus, au bas de l'entonnoir de repliement, le paysage énergétique accidenté à proximité de l'état natif contient plusieurs minima correspondant à ces sous-états conformationnels. Plus la protéine est souple, plus la hauteur de la barrière entre les substrats est faible, et plus l'ensemble des conformations quasi-natives acceptables peut être

grand. Généralement, les changements de conformation dans ce modèle sont au-delà des réarrangements de chaînes latérales locales près du site de liaison, mais affichent plutôt de grands mouvements concertés ou des mouvements de domaine entiers.

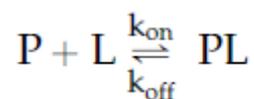
Les premières études de (Berger *et al.*, 1999) et (Foote & Milstein, 1994) démontrent que la sélection conformationnelle est responsable de la formation du complexe antigène-anticorps. Récemment, de plus en plus de données issues de la cristallographie à rayons X et des images de microscopes électroniques, des études de cinétique, de fluorescence et surtout de RMN montrent qu'une protéine peut avoir un répertoire d'états conformationnels, y compris des conformations de forme liée.

### 5.3. Mécanismes physicochimiques de l'interaction protéine-ligand

Afin de mieux comprendre la reconnaissance moléculaire entre une protéine et son ligand, il est nécessaire de comprendre les mécanismes physico-chimiques sous-jacents à l'interaction protéine-ligand.

#### 5.3.1. Cinétique de liaison protéine-ligand

La cinétique de liaison protéine-ligand décrit le processus sous-jacent à l'association entre la protéine et le ligand, en se concentrant particulièrement sur la vitesse à laquelle ces deux partenaires se lient l'un à l'autre. Dans un cas simple, lorsqu'une molécule de protéine P et une molécule de ligand L ayant une affinité mutuelle sont mélangées dans une solution, l'association en fonction du temps entre elles peut être formulée de la manière suivante :



où PL représente le complexe protéine-ligand,  $k_{\text{on}}$  et  $k_{\text{off}}$  sont les constantes de la vitesse cinétique qui représentent la réaction d'association (liaison) et la réaction de dissociation.

Les unités de  $k_{on}$  et  $k_{off}$  sont  $M^{-1}.s^{-1}$  et  $s^{-1}$ , respectivement. À l'équilibre, la réaction de liaison  $P + L \rightarrow PL$  doit être contre balancée par la réaction de dissociation  $PL \rightarrow P + L$ , et ca s'écrit:

$$k_{on}[P][L] = k_{off}[PL]$$

où les crochets représentent la concentration à l'équilibre de toute espèce moléculaire. La constante de liaison "binding"  $K_b$  (en unité de  $M^{-1}$ ) est alors définie par:

$$K_b = \frac{k_{on}}{k_{off}} = \frac{[PL]}{[P][L]} = \frac{1}{K_d}$$

où  $K_d$  (en unité de  $M$ ) est appelée constante de dissociation. Par conséquent, le taux de liaison rapide accompagné par une vitesse de dissociation lente donnera une liaison "association" élevée / faible / constante de dissociation et, par conséquent, une forte affinité de liaison.

### 5.3.2. Thermodynamique de l'interaction protéine-ligand

La compréhension complète de la reconnaissance moléculaire des protéines et des ligands nécessite la caractérisation de l'énergétique de liaison. Comme la situation du repliement des protéines, une description quantitative des forces qui entraînent la liaison aux protéines nécessite également la caractérisation et la détermination du profil «thermodynamique complet», y compris l'énergie libre de liaison ( $\Delta G$ ), l'enthalpie ( $\Delta H$ ) et l'entropie ( $\Delta S$ ) de liaison et le changement de capacité thermique ( $\Delta C_p$ ). Ces données thermodynamiques contiennent des informations cruciales non seulement pour élucider le mécanisme de liaison, mais aussi pour la conception rationnelle du médicament "drug design" en reliant les données thermodynamiques aux données structurales, qui seules ne peuvent décrire complètement les forces motrices de la liaison ou prédire avec précision l'affinité de liaison.

Le seul moyen direct de mesurer le changement de chaleur lors de la formation d'un complexe à température constante est mis en œuvre par la méthode de Calorimétrie par Titration Isotherme (ITC). Dans ce procédé, un partenaire de liaison est titré dans une solution

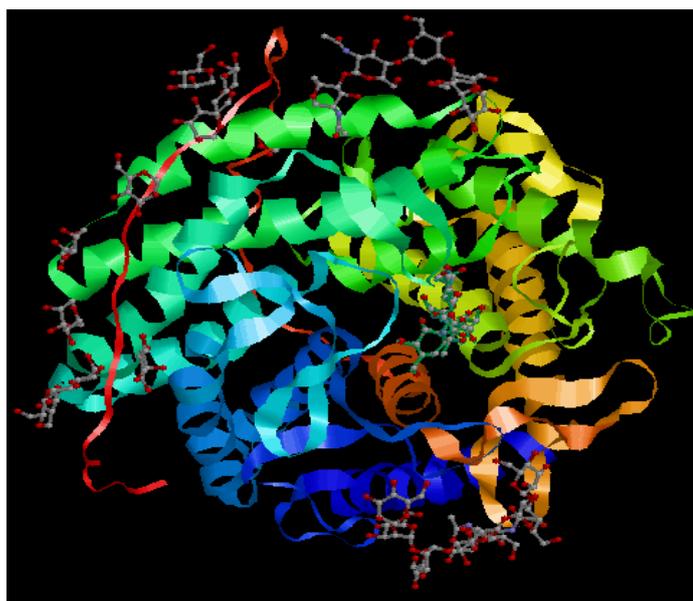
contenant le partenaire d'interaction, générant ou absorbant ainsi de la chaleur. Cette chaleur, reflétée par la différence de température entre une cellule échantillon et une cellule de référence, est directement mesurable et peut être quantifiée par le calorimètre. Dans des conditions appropriées, l'analyse des données provenant d'une seule expérience de l'ITC permet la détermination simultanée de la constante de liaison à l'équilibre ( $K_b$ ), du changement d'énergie libre standard de Gibbs ( $\Delta G$ ), des changements d'enthalpie ( $\Delta H$ ) et d'entropie ( $\Delta S$ ), ainsi que la stœchiométrie ( $n$ ) de l'événement d'association. De plus, des expériences effectuées à différentes températures donnent le changement de capacité thermique ( $\Delta C_p$ ) de la réaction de liaison.

## 5.4. Prédiction de l'interaction protéine-ligand

### 5.4.1. Docking (Amarrage moléculaire)

Le docking (ancrage) est l'expression donnée aux simulations moléculaires dans lesquelles différentes approches chimique, physique et biologique sont combinées pour étudier au niveau atomique les interactions entre deux molécules :

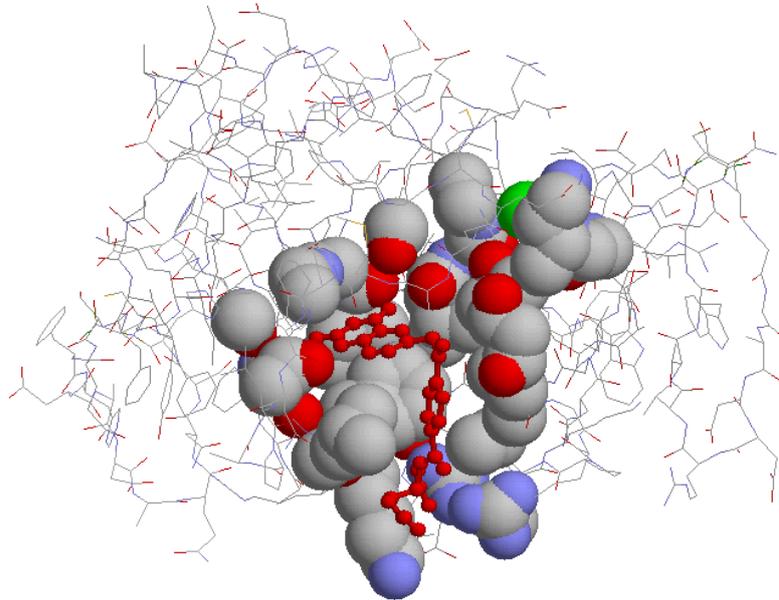
- L'une, la cible, est une protéine possédant un ou plusieurs sites actifs spécifiques avec une structure tridimensionnelle connue (structure-based design) (Fig. 5.2).



**Fig. 5.2. Docking de la structure de Glucoamylase avec l'acarbose (ligand).**

Le ligand est présenté sous forme de bâtonnets gris. Le code pdb de la protéine, 1AGM.

- L'autre, le ligand, est une petite molécule flexible d'origine exogène. Le docking permet la détermination des conformations les plus favorables pour la fixation du ligand à la cible conduisant à la formation du complexe le plus stable (Fig. 5.3).



**Fig. 5.3. La structure de la Dihydrofolate réductase complexée avec le Methotrexane.** l'inhibiteur est en représentation de boules et bâtonnets rouges, et les atomes de la protéine distants de moins 0,5 nm sont représentés en sphères de Van der Waals. Le code pdb de la protéine, 1RG7.

Tout programme repose sur la réalisation de deux étapes principales: la première consiste à rechercher toutes les orientations possibles dans la protéine (searching) et la seconde permet d'évaluer l'énergie de formation du complexe par la fonction de score (scoring). Le modèle de prédiction est élaboré à partir d'une ou plusieurs familles de molécules connues pour une activité biologique. Ce sont les méthodes basées sur les propriétés du ligand (ligand-based design). Le représentant de cette stratégie est le QSAR (Quantitative Structure Activity Relationship ou relation structure activité quantitative). Il s'agit d'établir une corrélation entre les variables de propriétés moléculaires, encodées par des descripteurs moléculaires, et de l'activité biologique (ou l'affinité) pour un ensemble de composés.

#### 5.4.2. Criblage virtuel

Le criblage virtuel s'apparente à une succession de techniques dont le point de départ est le pré-filtrage des chimiothèques à cribler. Souvent, la sélection des composés avant criblage

virtuel suit des règles spécifiques aux médicaments déjà connus et mis sur le marché (règles «drug-like»). Par exemple, la règle de Lipinski ou encore «règle des 5» peut être utilisée pour la sélection de composés « drug-like »:

- masse molaire  $\leq 500$  g/mol
- Log P  $\leq 5$
- Accepteurs de liaisons H  $\leq 10$
- Donneurs de liaisons H  $\leq 5$

Tous les composés ne validant pas au moins trois de ces conditions sont susceptibles de poser des problèmes d'absorption par voie orale. Des adaptations peuvent être faites afin de rendre la «règle des 5» plus stricte. C'est la notion de «lead-like». La masse molaire requise est diminuée (Masse molaire  $\leq 460$  g/mol). De même, le Log P doit être compris entre 4 et 4,2. Les composés «lead-like» peuvent ainsi subir des optimisations ou des ajouts de fragments induisant une augmentation de la masse molaire ainsi que du Log P tout en restant dans des limites compatibles avec les règles «drug like». Les composés respectant les critères «lead-like» ne sont pas pour autant certains de devenir des médicaments. En effet, beaucoup de molécules actives sont abandonnées du fait qu'elles possèdent une pharmacocinétique en dehors des normes. Afin d'évaluer le métabolisme d'une molécule, des méthodes de prédiction *in silico* basées sur la modélisation des cytochromes P450 ont vu le jour.

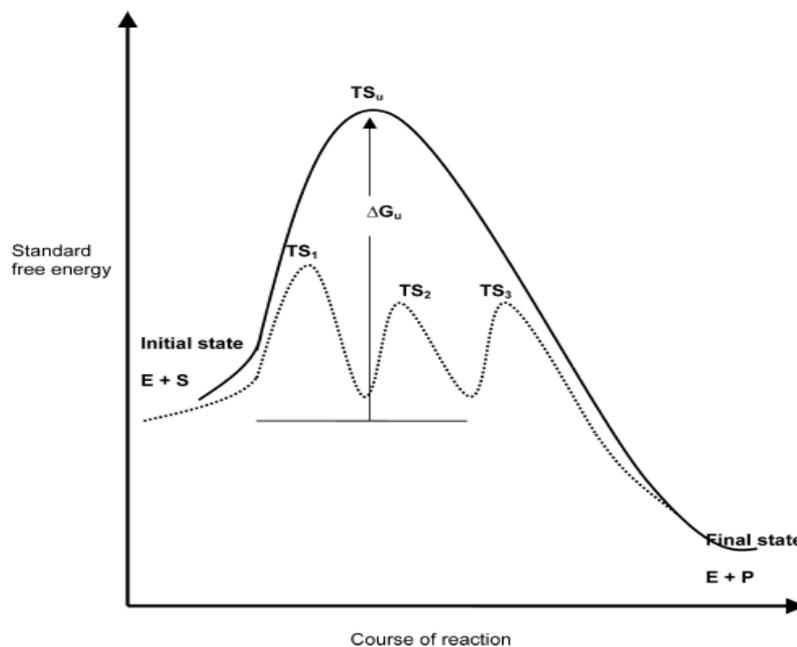
De manière plus générale, l'ADME-Tox (Absorption, Distribution, Métabolisation, Elimination - Toxicité) des médicaments est un facteur à prendre en compte dans l'optimisation de candidats prometteurs. Il serait également judicieux d'élaborer des filtres basés sur les propriétés ADME mais également sur la toxicité potentielle, éliminant ainsi tous les composés qui induiraient des effets secondaires. L'élaboration de tels filtres est aujourd'hui encore très délicate car les phénomènes de toxicité sont difficiles à modéliser, du fait des multiples paramètres à prendre en compte.

## Chapitre 6. Ingénierie des enzymes

### 6.1. Rappels sur les mécanismes de la catalyse enzymatique:

Avant d'aborder l'ingénierie des enzymes, il est nécessaire de faire un rappel sur les mécanismes de la catalyse enzymatique.

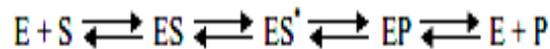
En effet, pour qu'une réaction se produise, les molécules du réactif doivent posséder suffisamment d'énergie pour franchir une barrière d'énergie potentielle, appelée énergie d'activation (Fig. 6.1).



**Fig. 6.1.** le profil d'énergie libre du déroulement d'une réaction catalysée par une enzyme impliquant la formation de complexes enzyme-substrat (ES) et enzyme-produit (EP). La voie de réaction catalysée passe par les états de transition  $TS_1$ ,  $TS_2$  et  $TS_3$ , avec l'énergie libre d'activation  $\Delta G_c$ , tandis que la réaction non analysée passe par l'état de transition  $TS_u$  avec l'énergie libre standard d'activation  $\Delta G_u$ .

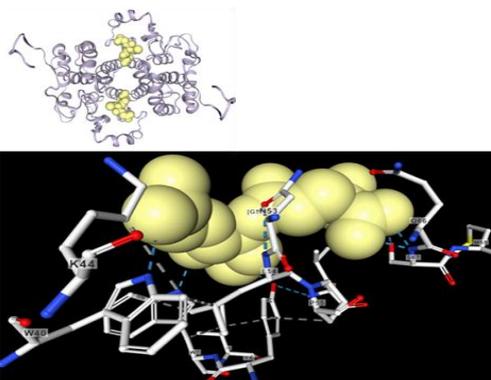
Toutes les molécules réactives ont différentes quantités d'énergie, mais seule une faible proportion d'entre eux a suffisamment d'énergie pour traverser l'énergie d'activation de la réaction. Plus l'énergie d'activation est faible, plus le substrat est capable de traverser l'énergie d'activation. Le résultat est que la vitesse de réaction est augmentée. La catalyse enzymatique nécessite la formation d'un complexe spécifique réversible entre le substrat et l'enzyme. Ce complexe est connu sous le nom de complexe enzyme-substrat (ES) et fournit toutes les

conditions favorables à l'événement catalytique. Les enzymes accélèrent les réactions en réduisant l'énergie nécessaire à la formation d'un complexe de réactifs capable de produire les produits de réaction. Ce complexe est connu sous le nom de complexe d'état de transition de la réaction et se caractérise par une énergie libre inférieure à celle que l'on trouverait dans la réaction non analysée.



(ES) doit passer à l'état de transition (ES\*). Le complexe d'état de transition doit évoluer vers un complexe enzyme-produit (EP), qui se dissocie en enzyme libre et en produit (P). Le chemin de cette réaction passe par les états de transition TS1, TS2 et TS3. La quantité d'énergie nécessaire pour atteindre l'état de transition est réduite; par conséquent, une plus grande proportion des molécules dans la population peut atteindre l'état de transition et traverser l'énergie d'activation. Les enzymes accélèrent les réactions directes et inverses proportionnellement, de sorte qu'elles n'ont aucun effet sur la constante d'équilibre des réactions qu'elles catalysent. Le substrat est lié à l'enzyme par des forces non covalentes relativement faibles. Les forces attractives intermoléculaires entre enzyme et substrat sont généralement de trois types:

liaisons ioniques, liaisons hydrogène et interactions de Van der Waals.



**Fig. 6.2. Le site de liaison de Glutathione S-transférase.** Des résidus sont impliqués directement dans la catalyse et sont considérés comme des résidus catalytiques, code PDB (1GNE).

La partie spécifique de la structure de la protéine qui interagit avec le substrat est appelée site de liaison au substrat (Fig. 6.2). Le site de liaison au substrat est une entité tridimensionnelle conçue de manière appropriée en tant que poche ou fente pour accepter la structure tridimensionnel du substrat. Les résidus de liaison sont définis comme tout résidu avec n'importe quel atome dans un rayon de 4 Å d'un substrat lié. Ces résidus de liaison qui participent à

l'événement catalytique sont appelés résidus catalytiques et forment le site actif. un résidu est défini comme catalytique si l'une des situations suivantes se produit:

1. Participation directe au mécanisme catalytique, par exemple comme nucléophile.
2. exercer un effet facilitant la catalyse sur un autre résidu ou molécule d'eau, qui est directement impliqué dans le mécanisme catalyseur.
3. Stabilisation d'un intermédiaire proposé à l'état de transition.
4. Exercer un effet facilitant la catalyse sur un substrat ou un cofacteur.

En effet, il se produit toujours plusieurs mouvements et changements de conformation des différentes régions. L'ampleur de ces mouvements dépend de nombreux facteurs, notamment la température, les propriétés du milieu de solvation, la présence ou l'absence de substrat et de produit. Les changements de conformation subis par l'enzyme jouent un rôle important dans le contrôle du cycle catalytique. Dans certaines enzymes, il y a des mouvements importants des résidus de liaison, généralement sur des boucles en surface, et dans d'autres cas, des changements de conformation plus importants. La catalyse se déroule sous forme fermée et l'enzyme s'ouvre à nouveau pour libérer le produit. Dans cette hypothèse d'ajustement induit (voir chapitre 5), l'interaction initiale entre l'enzyme et le substrat induisent rapidement des changements conformationnels sous la forme du site actif, ce qui donne une nouvelle forme du site actif qui rapproche les résidus catalytiques du substrat. Lors de la liaison du substrat pour que l'enzyme ait lieu, l'ajustement de forme déclenche la catalyse en générant des complexes d'états de transition. Cette hypothèse aide à expliquer pourquoi les enzymes ne catalysent que des réactions spécifiques (Notion de spécificité).

## 6.2. Histoire de l'ingénierie enzymatique

Les microorganismes étaient largement utilisés chez les peuples anciens. La fabrication de fromages, de pains, de boissons alcoolisées et de nombreuses autres applications dépend de microorganismes. Les enzymes étaient les composants principaux des microorganismes pour

la fabrication d'aliments et d'autres applications. Cependant, les applications anciennes reposaient sur des observations et des expériences répétées plutôt que scientifiques et technologiques. La dernière décennie a été marquée par une augmentation spectaculaire de l'utilisation d'enzymes comme des catalyseurs vert et durable dans les applications pharmaceutiques et industrielles. Les tableaux 6.1 et 6.2 indiquent les étapes historiques de l'ingénierie enzymatique et certains lauréats du prix Nobel au cours de cette période.

**Table 6.1 Histoire de l'ingénierie enzymatique**

Année	Événement important / découverte
1890	Le modèle «clé et serrure» a été proposé (Fisher)
1893	La définition du terme «catalyseur» (Ostwald) a été introduite
1897	L'explication que les enzymes ne nécessitent pas de cellule (Buchner)
1926	Enzyme s'est avéré être une protéine (Sumner)
1952	Réalisation de la biotransformation de stéroïdes
1958	La proposition du modèle « ajustement induit » (Koshland)
1963	L'identification de la séquence en acide amine de Ribonucléase
1965	La proposition du "modèle allostérique" d'enzyme (Monod)
1970	Des enzymes immobilisées ont été utilisées pour la production de HFCS
1980	La synthèse enzymatique de composés chiraux et de polymères en solvant organique
1990	L'ingénierie des protéines, y compris l'évolution dirigée, a été utilisée
2000	La conception bioinformatique de l'enzyme a été introduite
2010	Les enzymes d'ingénierie métabolique deviennent populaires

**Table 6.2 Prix Nobel sur les enzymes / protéines**

Année	Gagnant du prix	Découverte scientifique
1907	Eduard Buchner	Respiration alcoolique avec extrait sans cellules
1909	Wilhelm Ostwald	Définition du mot «catalyseur»
1946	James B. Sumner	Enzyme Uréase cristallisée à partir des haricots
1958	Frederick Sanger	Structure de l'Insuline
1972	Stanford Moore and	Relation entre la structure chimique et activité catalytique de la
1993	Michael Smith	Mutagenèse dirigée pour enzyme
	Kary B. Mullis	Invention de la réaction en chaîne de la polymérase (PCR)
2002	John B. Fenn, Koichi	Outils (Spectroscopie de Masse) pour l'identification et l'analyse
2008	Martin Chalfie, Roger	Green fluorescent protein (GFP)
2013	Martin Karplus, Michael Levitt et Arieh Warshel	Modèles multi-échelles pour systèmes chimiques complexes (mécanique quantique / mécanique moléculaire)

Après l'explication et la définition par les chercheurs des différentes notions fondamentales de l'enzymologie et après la biotransformation de stéroïdes dans les années 1950, Il y a eu historiquement quatre étapes d'avancées technologiques qui ont eu un impact important sur les

enzymes en tant que catalyseurs alternatifs sur les bioconversions, à savoir: l'isolation d'enzymes, l'immobilisation d'enzymes, les enzymes dans des milieux non conventionnels et la technologie de l'ADN recombinant. Le développement technologique pour la séparation des enzymes des micro-organismes est l'un des progrès technologiques importants. Cependant, l'isolement de ces enzymes intracellulaires était assez coûteux; ainsi leurs applications avaient été limitées. Leur ingénierie pour un usage répété, une activité et une stabilité supérieures a permis des processus plus réalisables.

L'immobilisation d'enzyme est une technique qui convertit une enzyme hydrosoluble en une forme insoluble pouvant être facilement récupérée et réutilisée. L'immobilisation peut améliorer les applications des enzymes de nombreuses manières, telles que la production de produits chimiques et pharmaceutiques, le biocapteur enzymatique, etc. Dans les années 1960, plusieurs technologies industrielles utilisant des enzymes immobilisées ont été développées. Un exemple en est l'isomérisation enzymatique du glucose en fructose pour la production de sirop de maïs à haute teneur en fructose (HFCS). Généralement, les réactions enzymatiques industrielles avant le milieu des années 1970 utilisent des substrats et des produits solubles dans des solutions aqueuses. A partir de 1975, divers procédés enzymatiques utilisant des solvants organiques ont été mis en place de manière industrielle (Table 6.3).

**Table 6.3 Bioconversions industrielles de substrats peu solubles dans l'eau**

Processus	Enzyme	En fonctionnement depuis
Interestérisation de graisse	Lipase	1979
Hydrolyse de l'ester	Lipase	1988
Transestérisation	Lipase	1990
Synthèse d'aspartame	Thermolysine	1992
Acylation	Lipase	1996

Le prochain progrès technologique est la technologie de l'ADN recombinant. Cette technologie, associée à l'ingénierie des protéines, permet de produire rapidement de nouveaux et meilleurs variants d'enzymes. Depuis 1990, l'évolution dirigée est utilisée comme un outil puissant pour l'ingénierie enzymatique. Cette méthode a beaucoup contribué à la mise au

point de procédés enzymatiques pour les applications industrielles afin de tirer parti des capacités des enzymes naturelles, car dans la plupart des cas, les enzymes naturelles ne sont pas optimisées pour les conditions de réaction industrielles. Actuellement, les avancées récentes en génie enzymatique concernent le design "conception" computationnel. Le design computationnel de l'enzyme est une autre méthode mise au point vers l'an 2000. Elle est réalisée à l'aide de modèles informatiques permettant de suggérer des séquences et des structures pouvant agir sur les propriétés souhaitées (voir chapitre 8). Actuellement, l'étude des enzymes est l'une des questions importantes pour la communauté scientifique et le secteur industriel en général. Les enzymes artificielles, les anticorps catalytiques sont des exemples de problèmes actuels en ingénierie enzymatique.

### **6.3. Technologie et ingénierie des enzymes**

Dans la technologie des enzymes, de nombreux produits tels que les aliments, la chimie fine et les produits pharmaceutiques ont été et sont fabriqués en utilisant des enzymes comme biocatalyseurs. De plus, les enzymes sont utilisées à des fins d'analyse et de diagnostic. Ils sont également utilisés dans de nombreux domaines, y compris l'assainissement de l'environnement. Ces dernières années, les enzymes pour la synthèse chimique remplaçant les produits pétrochimiques et pour l'utilisation du dioxyde de carbone sont de nouveaux sujets de la société dans l'ingénierie des enzymes.

Les processus biotechnologiques utilisent une ou plusieurs enzymes comme biocatalyseurs en fonction de l'état du processus requis. Comparé aux processus de fermentation, enzymatique la catalyse avec des enzymes isolées ou immobilisées présente les avantages suivants :

(1) la formation de sous-produits est réduite au minimum par d'autres enzymes dans les cellules, (2) pas besoin d'un milieu nutritif complexe (par exemple, le carbone, l'azote et d'autres sources de nutriments essentielles à la croissance cellulaire), et (3) des réacteurs plus petits peuvent être utilisés, car une productivité plus élevée peut être obtenue qu'avec des

cellules vivantes. Cependant, la fermentation utilisant les cellules vivantes sont encore appropriées que les enzymes isolées lorsque plusieurs enzymes en série et la régénération des cofacteurs sont impliqués.

Les enzymes sont plus sélectives que les catalyseurs chimiques conventionnels. Une sélectivité élevée est l'un des principaux avantages, notamment une réduction des réactions secondaires et donc des séparations plus faciles. De plus, la catalyse enzymatique peut être réalisée dans des conditions de réaction douces et elle a, dans de nombreux cas, des constantes catalytiques,  $k_{cat}$ , "turnover numbers" plus élevées par rapport à la catalyse chimique conventionnelle. La catalyse chimique est généralement effectuée à une température et une pression beaucoup plus élevées. Cependant, la catalyse enzymatique pour la synthèse chimique soulève également plusieurs problèmes, dont celui de la difficulté à isoler les enzymes. Les enzymes sont inhibées par des concentrations relativement faibles de produits finis, une température élevée, des conditions de pH extrêmes, certains ions métalliques et des solvants. Par ailleurs, certaines enzymes restent très coûteuses pour une utilisation industrielle et peuvent nécessiter des cofacteurs coûteux pour la catalyse. La catalyse enzymatique a tendance à être trop spécifique pour une applicabilité générale. Le tableau 1.4 résume les

**Table 6.4 Avantages et inconvénients de la technologie de l'enzyme par rapport aux procédés chimiques**

Avantages	Inconvénients
- Haut degré de sélectivité - Réactions dans des conditions douces	- Coûteuse
- Ecologique	- instable
- Catalyser large spectre de réactions	- Faible productivité
- Moins de sous-produits	
- Non toxique et non-inflammable	

avantages et les inconvénients de la technologie enzymatique. L'ingénierie enzymatique est également une terminologie largement utilisée couvrant la technologie traditionnelle liée aux enzymes et l'ingénierie des enzymes qui est très importante de nos jours pour améliorer l'applicabilité des enzymes.

#### 6.4. Classification des enzymes

Environ 5 000 enzymes ont été caractérisées jusqu'à présent, tandis que plus de 300 enzymes sont disponibles dans le commerce et fournies par des fabricants d'enzymes. En fonction des réactions qu'ils catalysent, les enzymes sont regroupées en fonction du rapport du Comité de la nomenclature de l'Union internationale de biochimie (1984). Les six classes distinctes sont présentées dans le tableau 6.5. Par ailleurs, International Union of Biochemistry (IUB) a attribué à chaque enzyme un nom et un numéro à quatre niveaux. Les numéros de la Commission des enzymes (CE) divisent les enzymes en six groupes principaux en fonction des réactions qu'ils catalysent, comme indiqué dans le tableau 6.5.

**Table 6.5 Classification des enzymes**

Classe	Sous-classes	Type de réaction
Oxydoréductases	Oxydases, oxygénases, peroxydases, déshydrogénases	Transfert d'électrons (oxydation et réduction)
Transférases	Glycosyltransférases, méthyltransférases, transaldolases, transkétolases, acyltransférases, alkyltransférases, transaminases, sulfotransférases, phosphotransférases, nucleotidyltransférases	Réaction de transfert de groupe
Hydrolases	Estérases, lipases, glycosidases, protéases, sulfatases, phosphatases, aminoacylases, endo- et exo-nucléases, halohydrolases	Réaction d'hydrolyse
Lyases	Décarboxylases, aldolases, kétéolases, hydratases, déhydratases, Polysaccharide lyases, ammonia lyases	Ajout de groupes aux doubles liaisons
Isomérases	Racémases, épimérases, isomérases	Transfert de groupes vers des formes isomères
Ligases	Synthétases, carboxylases	Formation des liaisons: C-C, C-S, C-O et C-N

#### 6.5. Application industrielle des enzymes

Les enzymes ont de nombreuses applications dans les industries alimentaire, médicale, chimique et pharmaceutique. Les industries ont connu une croissance rapide au cours des dernières décennies et devraient poursuivre leur croissance. Le tableau 6.6 présente les applications des biocatalyseurs et leur échelle de production. Pour ces processus enzymatiques, diverses enzymes ont été appliquées.

**Table 6.6 Application industrielle des enzymes et son échelle de production**

Échelle de production	Produit	Enzyme	Compagnie
>1,000,000	Sirop de maïs à haute teneur en fructose (HFSC)	Glucose isomérase	Divers
>100,000	Lait sans lactose	Lactase	Divers
>10,000	Acrylamide	Nitrilase	Nitto Co.
	Le beurre de cacao	Lipase (CRL)	Fuji oil
>1000	Aspartame ®	Thermolysine	Tosoh/DSM
	Nicotinamide	Nitrilase	Lonza
>100	Ampicilline	Penicilline amidase	DSM-Gist Brocades
	(S)-methoxyisopropylamine	Lipase	BASF

Les biocatalyseurs dans les industries sont généralement utilisés pour produire leurs produits et dérivés naturels. Les glucides et les dérivés d'acides gras sont principalement utilisés dans l'industrie alimentaire, tandis que d'autres types de composés sont principalement utilisés dans les industries pharmaceutiques et agro-alimentaires. Le secteur pharmaceutique domine particulièrement les applications des biocatalyseurs. Actuellement, plus de 5000 enzymes sont connues. Environ plus de 200 enzymes d'origine microbienne sont utilisées dans le commerce. La production commerciale d'enzymes a augmenté au cours des dernières décennies en réponse à la demande croissante et à l'application d'applications d'enzymes. La fabrication d'enzymes est relativement concentrée dans quelques pays tels que le Danemark, la Suisse, l'Allemagne, les Pays-Bas, les États-Unis, le Japon, la Russie et la Corée.

En effet, de nombreux chercheurs universitaires et industriels recherchent toujours plus d'applications et de meilleures technologies. La relation entre la structure et la fonction a fait l'objet de recherches approfondies, mais elle reste aujourd'hui l'une des questions les plus répondues en matière de génie enzymatique. Récemment, CRISPER (Clustered Regularly Interspaced Short Palindromic Repeats) est largement étudié comme un outil puissant et novateur pour étudier la régulation des gènes, l'expression des gènes et le criblage à l'échelle du génome après l'introduction des ADN. Les problèmes d'activité et de stabilité enzymatique par rapport à la structure peuvent trouver une solution pour de nouvelles fonctions et applications. De nouvelles applications des enzymes actuelles sont également

considérées comme un problème important. Par exemple, dans l'industrie cosmétique, des ingrédients plus naturels et non toxiques sont particulièrement nécessaires, pouvant être obtenus à partir de produits naturels ou d'une synthèse enzymatique à la place d'une synthèse chimique dans laquelle des sous-produits sont souvent formés.

Les enzymes sont utilisées dans les formulations de détergents depuis 1960 pour lutter contre l'eutrophisation de l'eau provoquée par les détergents au phosphore. Des protéases, des amylases, des lipases et des cellulases ont été utilisées pour dégrader les taches de protéines, de glucides et de lipides sur les vêtements. En utilisant des enzymes dans l'industrie des détergents, on peut économiser de l'énergie en lavant à basse température avec une efficacité comparable à celle à haute température et en utilisant des tensioactifs traditionnels.

En résumé, les enzymes de l'industrie des détergents présentent les avantages suivants:

- Coût réduit depuis utilisé à des concentrations de détergent faibles.
- Acceptable pour l'environnement: biodégradabilité et aucun impact négatif sur les processus de traitement des eaux usées.
- Une plus grande efficacité dans l'élimination des taches.
- Moins d'utilisation de polluants tels que les phosphates, l'eau de Javel et les caustiques.

Les enzymes liées aux détergents ont traditionnellement été isolées de la nature, mais sont aujourd'hui conçues pour offrir de meilleures propriétés répondant aux conditions de formulation et aux processus de lavage nécessitant parfois des conditions de température et de pH élevés.

## **6.6. Production des enzymes**

### **6.6.1. Sources des enzymes**

La première étape de la production d'enzymes est la sélection de la source d'enzymes. Les enzymes peuvent être dérivées de micro-organismes par des processus de fermentation, ainsi que de sources végétales et animales. Le tableau 6.7 présente les enzymes d'importance

industrielle et leurs sources. Les microorganismes sont des sources attractives d'enzymes car ils peuvent être cultivés à grande échelle. Les enzymes produites à partir de micro-organismes ont, dans de nombreux cas, de meilleures propriétés, telles qu'une stabilité élevée, que les enzymes de sources végétales ou animales. L'identification des micro-organismes appropriés pour la production d'enzymes commence par le criblage d'une grande variété d'organismes généralement reconnus comme sûrs, si possible. D'autre part, le dépistage des micro-organismes compte tenu de leur condition de croissance peut également être une possibilité.

Les plantes ne sont généralement pas considérées comme des sources d'enzymes industrielles, car elles sont saisonnières et intracellulaires, ce qui nécessite des processus supplémentaires pour rompre la paroi cellulaire ferme. Cependant, malgré ces limitations, un certain nombre d'enzymes de sources végétales ont été utilisées dans l'industrie. Par exemple, la papaïne, protéase de cystéine peut être obtenue à partir du fruit vert et des feuilles de la plante *Carica papaya*. Les tissus et les sécrétions animales sont également une source potentielle d'enzymes. La pepsine, également connue sous le nom de chymosine, est l'une des enzymes d'origine animale importantes dans l'industrie. La présure, protéase aspartique, obtenue à partir de l'estomac ou des abomasums de veaux, est utilisée pour la production de fromage et comme aide à la digestion.

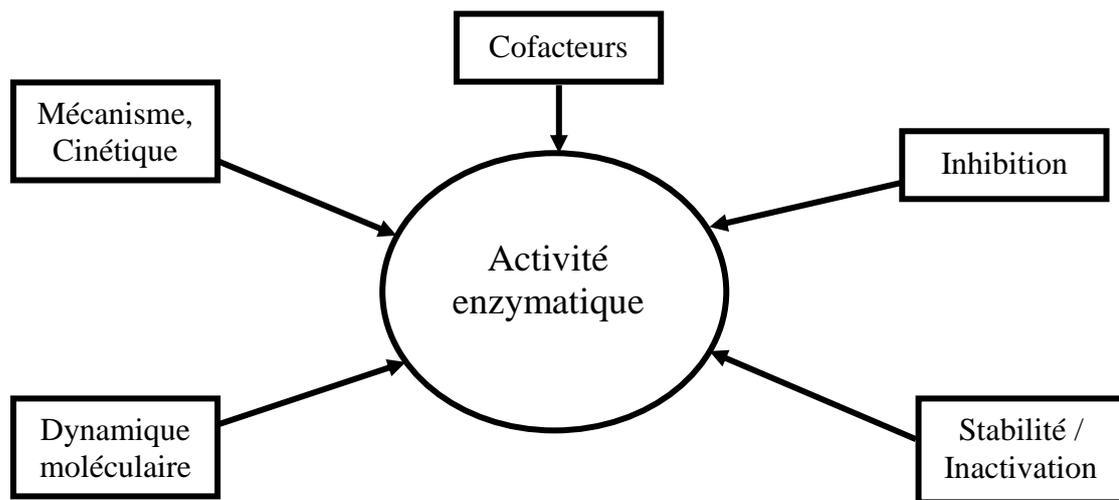
**Table 6.7 Enzymes d'importance industrielle et leurs sources**

Enzyme	Source	Application
<b>Dérivés de microbes:</b>		
Protéases	<i>Bacillus amyloliquefaciens, B.subtilis, Aspergillus oryzae, Streptomyces spp.</i>	Hydrolyse enzymatique des protéines
$\alpha$ -amylase	Divers <i>bacilli</i> et <i>Aspergillus oryzae</i>	Hydrolyse de l'amidon
$\beta$ -amylase	Divers <i>bacilli</i> , orge	Dégradation de l'amidon
Glucoamylase	<i>Aspergillus niger, Rhizopus spp.</i>	Amidon hydrolysé en glucose sirop.
Glucose isomerase	Divers <i>bacilli, Streptomyces</i> spp.	Glucose isomérisé en fructose
Invertase	<i>Saccharomyces spp.</i>	Saccharose hydrolysé en glucose et fructose
Penicilline acylase	<i>E. coli, divers bacilli, Streptomyces spp.</i>	Production de semi-synthétique pénicillines
Pectinases	<i>Aspergillus niger</i>	Hydrolyse enzymatique de la pectine
Lipase	<i>Candida antarctica</i>	Hydrolyse des triglycérides
<b>Dérivés de plantes:</b>		
Papaine	<i>Carica papaya</i>	Attendrissement de la viande
Bromélaïne	Ananas (fruit / tige)	Mise au froid des bières
Actinidine	Kiwi	Attendrissement de la viande

Lipoxygenase	Soja	Pain et production d'arômes
<b>Dérivé de sources animales:</b>		
Chymosine (présure)	la caillette (le quatrième estomac)	Fabrication de fromage
Ancrod	Venin de serpent	Anticoagulant
Urokinase	Urine	Thrombolytique
Trypsine	Pancréas animal	Aide digestive
Pancreatine (amylase, protease, lipase)	Extrait pancréatique	Aide digestive
Pepsine	Estomac des animaux	Aide digestive
Acetyl-cholinesterase	Erythrocytes bovines	Analyse de composés organophosphorés
Cholesterol esterase	Pancréas porcin	Détecter les niveaux de cholestérol sérique

### 6.6.2. Criblage des enzymes

Les enzymes sont essentielles car leur pouvoir catalytique facilite les processus de la vie. Cependant, plusieurs facteurs doivent être pris en compte, tels que la spécificité du substrat enzymatique, l'activité, la sélectivité, la stabilité et la recyclabilité pour déterminer l'utilité de l'enzyme pour la biocatalyse. L'activité des enzymes, l'une des propriétés les plus importantes, est affectée par leur structure et leur environnement (Fig. 6.3).



**Fig. 6.3. Facteurs affectant l'activité enzymatique**

De nos jours, les recherches portent sur la modification de l'enzyme afin d'améliorer son potentiel catalytique pour une utilisation industrielle. Étant donné que l'application industrielle des enzymes dépend de leur activité, de leur stabilité et de leur coût; il est très intéressant de concevoir des enzymes (par exemple, de sélectionner de nouvelles enzymes et de trouver de nouvelles fonctions des enzymes) répondant aux critères des producteurs et des utilisateurs finaux. L'une des stratégies les plus récentes et les plus efficaces pour découvrir de

nouvelles enzymes consiste à rechercher des micro-organismes présentant une diversité caractéristique. Une autre stratégie est l'approche basée sur le métagénome, la méthode indépendante de la culture utilisant des données génomiques microbiennes (métagénomique).

Le processus se compose de quatre étapes principales:

1. Obtenir des échantillons de l'environnement,
2. Isolement de l'ADN et manipulation du matériel génétique,
3. Construction de la librairie métagénomique, et
4. Criblage d'une nouvelle enzyme de fonction et séquençage du matériel génétique de la librairie métagénomique.

Certaines techniques sont envisagées pour le ciblage d'un microorganisme pour la production d'enzymes : le microorganisme devrait se développer avec des nutriments pas chers et facilement disponibles, l'enzyme doit être produite avec un rendement élevé dans un temps relativement court et le microorganisme doit être non pathogène.

Un certain nombre de techniques ont été développées pour obtenir de nouvelles enzymes.

Ceux-ci incluent les méthodes suivants:

1. ingénierie au niveau génétique pour muter une enzyme préexistante,
2. combinant des catalyseurs chimiques et enzymatiques,
3. appliquer les enzymes existantes sur un nouveau substrat non naturel, et
4. utilisation de différentes conditions de réaction, telles que le solvant.

### **6.6.3. Processus de fermentation**

La production d'enzymes est toujours un domaine important de la biotechnologie. Les brevets et les articles de recherche sont en augmentation et les ventes d'enzymes avoisineraient un milliard de dollars par an. Les processus de fermentation pour la production microbienne d'enzymes peuvent être réalisés par culture à l'état solide ou par culture immergée de micro-organismes. La plupart des fabricants d'enzymes ont produit des enzymes en utilisant des

techniques de fermentation en immersion. La fermentation en immersion est définie comme une fermentation en présence d'un excès d'eau. Cette technique de fermentation permet de mieux suivre et contrôler les paramètres du processus, tels que la température, le pH, l'aération et la dispersion, pour une croissance efficace. Un autre avantage sera la facilité de la manipulation et la mise à grande échelle possible. Ainsi, presque toutes les grandes installations de production d'enzymes utilisent cette technologie. Cependant, les titres de cette technologie sont relativement bas. Le coût de récupération du produit est inversement proportionnel à la concentration du produit dans un bouillon de fermentation. Par conséquent, la concentration dans la fermentation en immersion doit être optimisée pour une production d'enzymes rentable. La fermentation à l'état solide où les microorganismes sont cultivés sur un substrat solide humidifié est encore largement utilisée pour les enzymes bon marché et où les conditions météorologiques sont bonnes et la main-d'œuvre peu coûteuse. Les microorganismes, tels que les champignons, la levure, certaines bactéries, ou une combinaison de ceux-ci, peuvent être utilisés. La fermentation à l'état solide présente de nombreux avantages par rapport à la fermentation en immersion en termes de fonctionnement simple, de faible investissement en capital et de faible consommation d'énergie. Les déchets agricoles peuvent être utilisés en tant que substrats sans prétraitement important en la fermentation à l'état solide. Cependant, le développement de cette dernière est plus lent que celui de la fermentation en immersion, car celle-ci rencontre des difficultés dans le contrôle du processus et la mise à grande échelle. Le tableau 6.8 montre la différence entre les deux techniques.

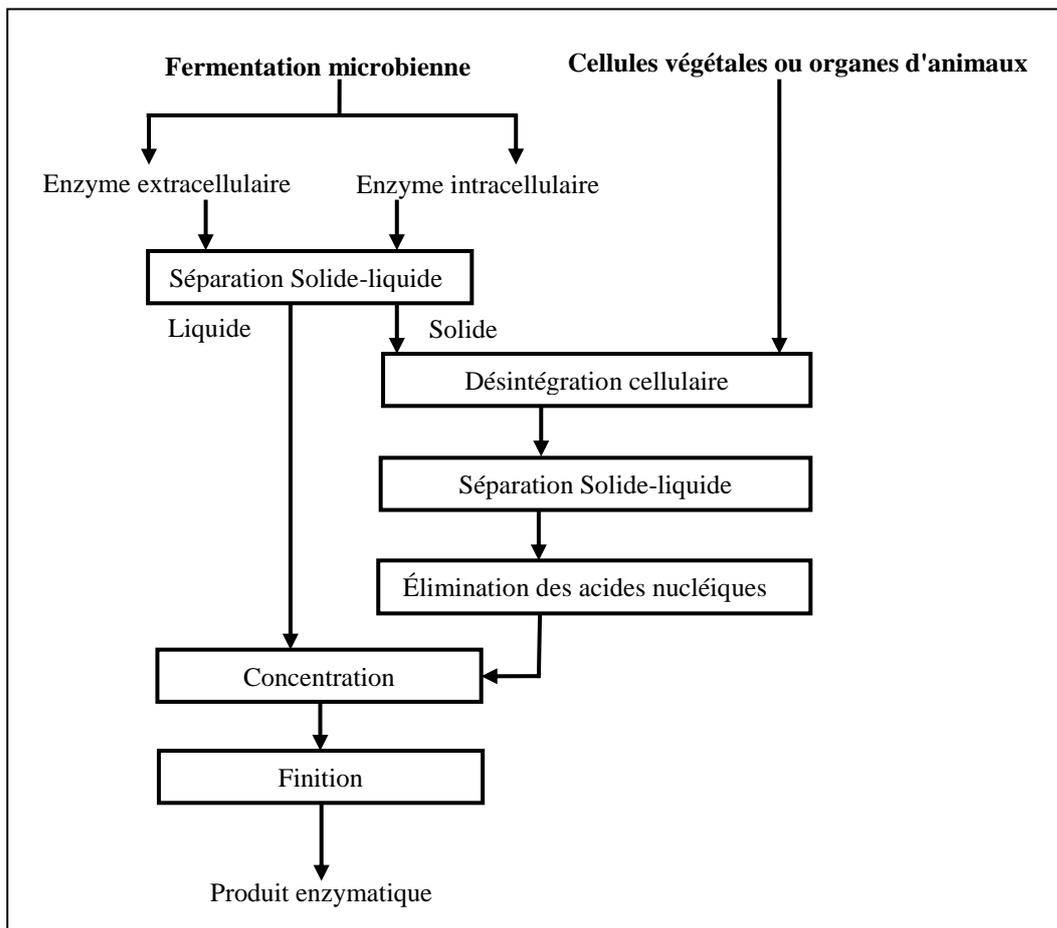
**Table 6.8 Différentes méthodes de culture**

<b>Fermentation à l'état solide</b>	<b>Fermentation en immersion</b>
Nécessite beaucoup d'espace pour les plateaux	Utilise des fermenteurs compacts et fermés
Nécessite beaucoup de travail	Nécessite un minimum de travail
Peu de puissance requise	Besoin de puissance pour les compresseurs d'air et les agitateurs

Un minimum de contrôle est nécessaire	Nécessite de contrôler les conditions de culture
Pas beaucoup de séparation pour les applications	Diverses technologies de séparation sont nécessaires

#### 6.6.4. Séparation et purification des enzymes

Au préalable, les enzymes doivent être extraites de leurs sources biologiques. Voir la figure 6.4 pour les procédures d'extraction. Chaque procédure d'extraction ou de purification entraîne une augmentation du coût de production et diminue donc le rendement en enzyme. L'activité enzymatique spécifique peut être augmentée en ajoutant des étapes de purification qui nécessitent également un coût supplémentaire.



**Fig. 6.4. Séparation des enzymes**

De la source biologique, la teneur en protéines de l'enzyme devrait être élevée afin de faciliter les processus en aval. Un autre facteur important à prendre en considération est l'activité enzymatique. Une activité maximale doit être conservée lors de la préparation des enzymes.

La chaleur, la protéolyse, le pH, l'oxydation et la perte de cofacteurs font partie des facteurs qui réduisent l'activité enzymatique. La chose la plus importante parmi ces facteurs est probablement l'inactivation de la chaleur, en particulier si un refroidissement insuffisant est disponible pendant l'extraction et la purification. La contamination microbienne par les protéases est également une autre cause d'inactivation enzymatique. Une protéolyse par protéase peut survenir au début de l'extraction d'enzymes. Les enzymes produites de manière extracellulaire sont parfois utilisées sans isolement ni purification. Par contre, une désintégration cellulaire est nécessaire pour le système intracellulaire.

Les enzymes doivent être libérées des cellules dans la solution avant purification. La séparation solide-liquide est nécessaire pour la séparation de la masse cellulaire et la séparation ultérieure des débris cellulaires après la désintégration de la cellule. Une désintégration cellulaire est nécessaire pour extraire les enzymes intracellulaires. Cela implique de donner des forces de cisaillement aux cellules ou d'ajouter des produits chimiques pour la lyse des cellules. Pour les forces de cisaillement, le degré d'énergie nécessaire pour casser les cellules dépend de l'organisme. Les enzymes intracellulaires peuvent contenir des acides nucléiques susceptibles d'interférer avec les procédures de purification des enzymes en raison d'une augmentation de la viscosité. Les acides nucléiques doivent être éliminés par addition de nucléases ou par précipitation. Cependant, la précipitation d'acide nucléique à l'aide de sulfate d'ammonium peut également éliminer certaines protéines précieuses en même temps. Le sulfate d'ammonium en tant qu'agent précipitant est largement utilisé en raison de sa solubilité élevée, de son prix avantageux, de son absence de toxicité pour les enzymes. Les préparations enzymatiques qui ont été concentrées peuvent être encore purifiées par chromatographie. Le tableau 6.9 présente les processus en aval pour la séparation et la purification des enzymes.

**Table 6.9 Exemple de procédés en aval pour la séparation et la purification d'enzymes**

Méthodes	Description
<b>Séparation Solide-liquide:</b>	
Centrifugation	Séparer sur la base de la différence de granulométrie et de densité entre les phases liquide et solide. L'efficacité dépend de la fraction volumique de solides, de la surface de clarification efficace.
Filtration	Séparer sur la base de la taille des particules. L'efficacité est limitée par la forme et la compressibilité de la particules, la viscosité de la phase liquide et le maximum pressions admissibles.
<b>Désintégration cellulaire:</b>	
Désintégration des cellules par ultrasons	Utilise le mouvement sinusoïdal d'une sonde dans le liquide. Problème de bruit en laboratoire et bon pour le fonctionnement en laboratoire.
Le broyeur à billes	Les suspensions cellulaires sont agitées en présence de billes en verre ou d'acier. Les cellules sont brisées par les forts gradients de cisaillement et par la collision avec les billes.
Homogénéisateur haute pression	Le principal facteur perturbateur est la pression appliquée et la chute de pression qui en résulte à travers la vanne.
Méthode lytique	La lyse enzymatique a été largement utilisée à l'échelle du laboratoire. La lyse par un acide, un alcali, un tensioactif et des solvants peut être efficace en libérant des enzymes, à condition que les enzymes soient suffisamment robustes.
<b>Concentration:</b>	
Précipitation	Méthode bon marché et première étape dans la purification des enzymes. L'augmentation de la force ionique de la solution qui provoque une réduction de l'effet répulsif de charges identiques entre des molécules identiques d'une protéine et les forces qui maintiennent la couche de solvation autour de la protéine pour qu'elle soit précipitée.
Chromatographie échangeuse d'ions	Enzymes chargées positivement (par exemple, pH inférieur au point isoélectrique) se lient aux échangeurs de cations. Enzymes chargées négativement (par exemple, pH supérieur au point isoélectrique) se lient aux échangeurs d'anions. La force de liaison est déterminée par le pH et la force ionique de la solution et par les structures de l'enzyme et de l'échangeur d'ions.
Chromatographie d'affinité	Séparer les enzymes sur la base d'une interaction réversible entre l'enzyme et un ligand spécifique immobilisé sur une matrice de chromatographie. Haute sélectivité et capacité pour la protéine d'intérêt.
Chromatographie d'exclusion	Sépare les enzymes en fonction de leur taille moléculaire et la forme. Plus la molécule est grosse, plus elle a du mal à passer à travers et pénétrer les pores du gel.

### 6.7. Enzyme immobilisée

L'un des facteurs qui entrave l'utilisation des enzymes dans les processus industriels est leur coût. Comme les enzymes sont des catalyseurs, elles ne sont pas consommées lors des réactions. Si les enzymes sont utilisées sous forme libre, elles doivent être récupérées après la

réaction pour pouvoir être réutilisées. Les enzymes restantes peuvent également contaminer le produit s'il n'est pas séparé au cours des étapes de purification. La séparation de l'enzyme du produit pendant ou après la réaction en utilisant un système biphasique aqueux / non aqueux peut être une solution. Une phase contient l'enzyme et l'autre phase contient le produit. Un exemple de système à deux phases est constitué par les enzymes immobilisées: en fixant des enzymes sur ou dans des supports.

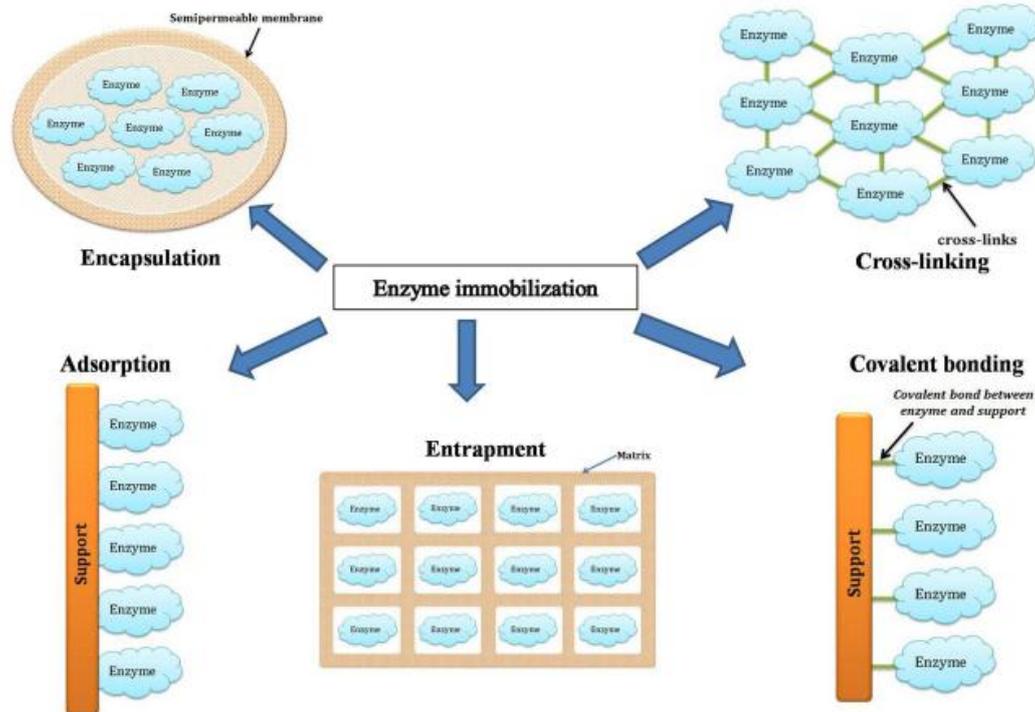
Les avantages importants dérivés de l'immobilisation des enzymes sont (1) la facilité de séparation du produit du milieu réactionnel et (2) la possibilité de réutilisation de l'enzyme. La facilité de séparation de l'enzyme du flux de produit permet aux applications industrielles de l'enzyme d'être plus efficaces. La contamination du produit par la réaction enzymatique peut être prévenue en utilisant une enzyme immobilisée qui aide à réduire le coût de traitement en aval. La réutilisation de l'enzyme peut offrir des avantages de coût qui constituent souvent une partie essentielle pour que les processus de catalyse de l'enzyme commerciale soient économiquement réalisables. La comparaison entre les enzymes immobilisées et les enzymes libres est donnée dans le tableau 6.10.

**Table 6.10 Comparaison des enzymes immobilisées avec des enzymes libres**

Enzyme libre	Enzyme immobilisée
Difficile de séparer et de réutiliser	Pas besoin de se séparer du flux de produits
Les enzymes sont généralement chères	Peut être réutilisé plusieurs fois, donc le coût pour l'enzyme peut être réduite
L'activité est influencée par la température, le pH, et autres conditions de fonctionnement	Les propriétés peuvent être changées: activité, température profil, pH optimum

Il existe plusieurs méthodes d'immobilisation d'enzymes: méthodes physiques et méthodes chimiques. Les méthodes physiques comprennent l'adsorption, le piégeage "entrapment" et l'encapsulation; les méthodes chimiques comprennent la réticulation "cross-linking" et la liaison covalente. Quelle que soit la méthode d'immobilisation, les matériaux à utiliser pour l'immobilisation de l'enzyme doivent être insolubles dans le milieu réactionnel.

L'immobilisation de l'enzyme doit offrir de bonnes propriétés physiques pertinentes pour les applications industrielles, même si le matériau de support est nécessaire pour faciliter la diffusion du substrat et du produit (Fig. 6.5).



**Fig. 6.5. Méthodes d'immobilisation enzyme / cellule** (V.L. Sirisha *et al.*, 2016).

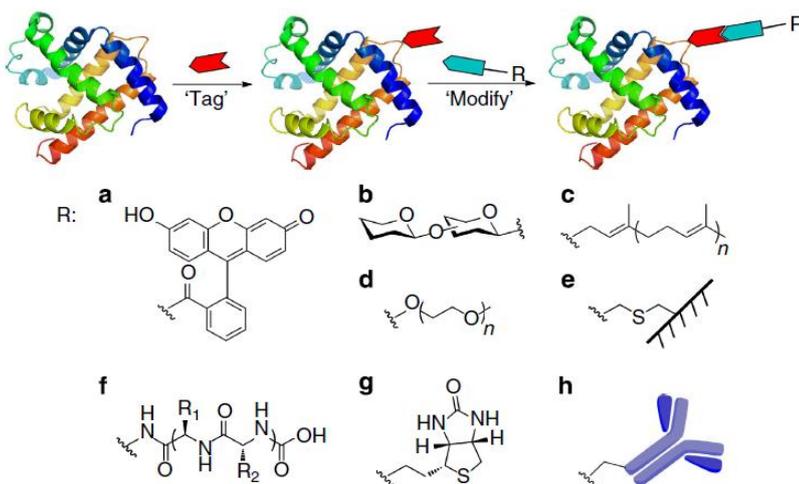
Ces enzymes immobilisées ont une large gamme d'applications. Celles-ci incluent des applications dans les industries du sucre, du poisson et du vin, où elles sont utilisées pour éliminer les composés organiques des eaux usées. Ces enzymes immobilisées sont utilisées dans des biocapteurs sophistiqués pour le contrôle des métabolites et la mesure *in situ* des polluants environnementaux. Les enzymes immobilisées trouvent également une application importante dans le métabolisme des médicaments, la production de biodiesel et d'antibiotiques, la bioremédiation et l'industrie alimentaire. L'utilisation répandue des enzymes immobilisées est due en grande partie au fait qu'elles sont moins chères, respectueuses de l'environnement et beaucoup plus faciles à utiliser par rapport à des technologies équivalentes. En effet, Au cours des dernières années, des progrès importants ont

été réalisés dans la conception de l'immobilisation d'enzymes, de la matrice de support à différentes tailles de pores et des modifications de la surface. Concevoir un matériau de support idéal en modifiant les caractéristiques structurelles spécifiques requises pour une enzyme cible est désormais possible grâce à de nouvelles simulations. De plus, comme la structure de l'enzyme et le mécanisme d'action sont connus, des méthodes d'immobilisation contrôlées peuvent être développées. En outre, l'identification et la caractérisation des métabolites essentiels et leur dépistage des maladies curatives ont fait l'objet d'une attention particulière. Ceci est principalement dû au développement de biocapteurs, de biopuces et de bioréacteurs économiques et disponibles. Par conséquent, l'avenir est très prometteur avec l'utilisation accrue d'enzymes immobilisées dans les domaines pharmaceutique, clinique, alimentaire, biotechnologique et autres.

En conclusion, la dernière décennie a été marquée par une augmentation spectaculaire de l'utilisation d'enzymes comme biocatalyseurs écologiques et durables dans les applications pharmaceutiques et industrielles. Cette tendance a été alimentée dans une large mesure par les progrès réalisés dans la capacité des scientifiques et des ingénieurs industriels à adapter les enzymes natives par ingénierie des protéines. L'ingénierie des protéines a généré des variants enzymatiques ayant une activité catalytique améliorée, une spécificité de substrat élargie ou modifiée, ainsi qu'une stéréosélectivité élevée ou inversée. Les enzymes ont été conçues pour conserver l'activité à des températures élevées et pour fonctionner en présence de solvants organiques, de sels et de valeurs de pH loin des conditions physiologiques.

### Chapitre 7. Modification chimique des protéines *in vitro* et *in vivo*

La modification des protéines est répandue dans la nature, augmentant la diversité de la structure des protéines. Cependant, notre capacité à imiter synthétiquement la capacité naturelle à installer de telles modifications est essentiellement limitée par la chimie disponible. La réaction sur un seul acide aminé ou un seul site, parmi une mer d'acides carboxyliques réactifs, d'amides, d'amines, d'alcools et de thiols, constitue un défi important et passionnant en chimio- et en régio-sélectivité. Les transformations potentielles, si elles doivent être pertinentes, sont conditionnées par la nécessité de conditions biologiquement ambiantes (c'est-à-dire  $<37^\circ\text{C}$ , pH 6-8, solvant aqueux) afin de ne pas perturber l'architecture et / ou la fonction des protéines. Idéalement, ceci devrait se faire avec une conversion presque totale pour générer des constructions homogènes. Les applications des protéines modifiées sont nombreuses. ils sont aussi variés que le suivi *in vivo* des conjugués protéine-fluorophore jusqu'à la polyéthylène glycol (PEG) des protéines thérapeutiques pour réduire l'immunogénicité, depuis la production de matériaux aux propriétés nouvelles pour sonder le mécanisme des enzymes pathologiques.



**Fig. 7.1. L'approche « Tag "étiquette" et modifier » de la modification des protéines** (Spicer *et al.*, 2014). Un «marqueur» d'acides aminés réactif unique est installé sur une surface de protéine pour permettre une réaction avec une «modification» souhaitée. Les exemples comprennent (a) les fluorophores (b) la glycosylation (c) la prénylation (d) la PEGylation (e) la fixation aux surfaces solides (f) les peptides (g) la biotinylation et (h) la conjugaison d'anticorps.

Au cours des trois dernières décennies, un certain nombre de méthodologies sont apparues

(Fig. 7.1) pour entreprendre la modification des résidus d'acides aminés naturels et non standard, *in vitro* et *in vivo*, en s'appuyant sur une boîte à outils auparavant limitée pour la modification. à la cystéine et à la lysine.

Bien que des progrès significatifs aient été réalisés dans les méthodes de marquage chimique des protéines *in vitro*, seules quelques-unes d'entre elles sont applicables au marquage *in vivo*. Cela représente un défi de taille pour le marquage chimique sélectif des protéines dans leur environnement d'origine. Les progrès récents dans les techniques de marquage "labeling" chimique *in vivo* impliquent la combinaison de la spécificité des étiquettes "tags" codées génétiquement et de la flexibilité des sondes à petites molécules. Ces séquences d'acides aminés comprennent des motifs de tétracystéine / tétraserine, des motifs de chélation de métaux, des étiquettes "tags" peptidiques pour modification enzymatique, des domaines de liaison de ligand et des domaines enzymatiques auto-marquants (Tableau 1). Comme le résume le tableau 1, toutes ces méthodes de marquage chimique *in vivo* ont leurs avantages et leurs inconvénients en termes de taille de marqueur, de cytotoxicité, de spécificité, de cinétique de réaction, de rapport signal sur bruit et d'accessibilité pour le marquage intracellulaire. De courtes étiquettes "tags" peptidiques sont possibles; cependant, la spécificité doit être conférée en utilisant des modifications par chélation ou catalysées par des enzymes, ce qui pose généralement les problèmes de cytotoxicité et de vitesse de réaction relativement lente, respectivement. Les systèmes de marquage utilisant des domaines de liaison au ligand et des domaines enzymatiques auto-marquants sont plus simples et mieux adaptés au marquage intracellulaire, alors que ces marqueurs relativement grands peuvent interférer avec les fonctions protéiques.

**Table 7.1. Méthodes de marquage chimique dans les cellules vivantes (Wu *et al.*, 2010).**

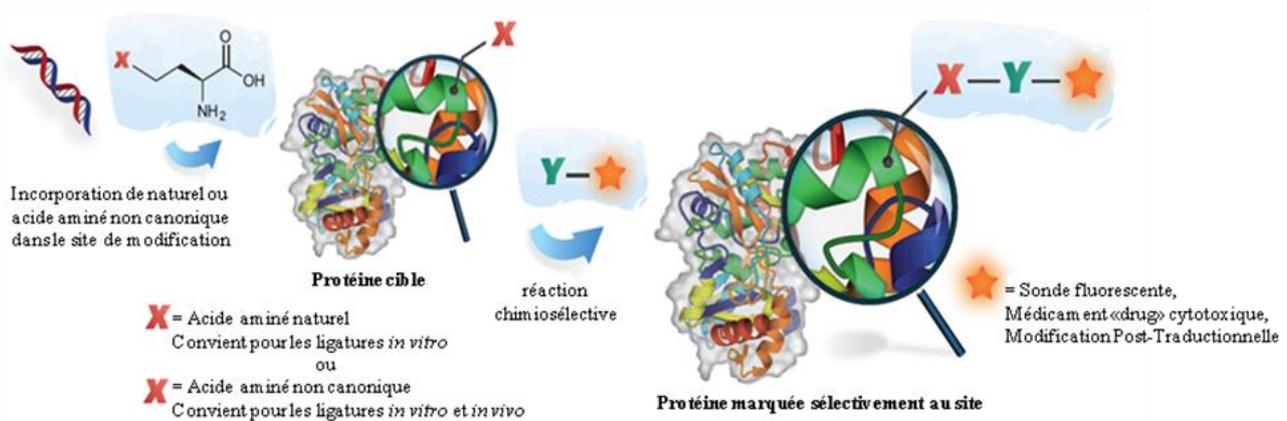
Tag (taille en amino acides)	Sonde de petite molécule	Enzyme requise	Marquage cellulaire	Commentaires
<b>Réactifs biarséniques et bisboroniques fluorogènes</b>				
CCXXCC (6) tetracysteine	Biarsenical: FIAsh, ReAsH	Non	Intra-cellulaire	Marquage fluorogénique; Non spécifique réactions et interactions; Cytotoxique dans certains cas.
SSXXSS (6) tétraserine	RhoBo Bisboronique	Non	Surface cellulaire	Fluorogène bas; Cytotoxique; Marquage hors cible de motif de tétraserine endogène.

Chélation de métaux				
6H (6)	Ni-NTA	Non	Surface cellulaire	Ni <sup>2+</sup> est toxique pour les cellules et inhibe la fluorescence "quenching"
6H (6)	Le complexe Zn <sup>2+</sup> HisZiFit	Non	Surface cellulaire	Non cytotoxique
DDDD (4)	Le complexe Zn <sup>2+</sup>	Non	Surface cellulaire	Changement spectroscopique lors de la chélation.
Modification enzymatique				
ACP (77)/A1 (12) ou PCP(80)/ybbR (11)	Dérivés de Coenzyme A (CoA)	Phospho-pantétheinyle transférases (AcpS ou Sfp)	Surface cellulaire	AcpS ou Sfp catalyse l'attachement d'une sonde activée par CoA au résidu sérine de l'étiquette "Tag" peptidique. k <sub>cat</sub> = 0,03–0,3 s <sup>-1</sup> , k <sub>cat</sub> / K <sub>m</sub> = 0,25–4,2 × 10 <sup>4</sup> M <sup>-1</sup> s <sup>-1</sup>
AP tag (15)	Analogue de biotine cétone et ligature ultérieure d'oxime	Biotine ligase (BirA)	Surface cellulaire	La cétobiotine fonctionne comme un substrat pour BirA et est incorporée dans le résidu lysine de tag peptidique; Temps minimal de marquage 20 min.
Q-tag (7)	Dérivés de Cadaverine	Transglutaminase (TGase)	Surface cellulaire	La TGase intervient dans la formation de la liaison amide entre l'amine de la cadavérine et le résidu de glutamate de tag peptidique.
LAP tag (13–22)	Acides azido-lipoïques et cycloaddition favorisée par les souches, ou la coumarine	Ligase d'acide lipoïque (LplA) et mutant LplA	Surface cellulaire et intra-cellulaire	k <sub>cat</sub> = 0,048 s <sup>-1</sup> pour la conjugaison de l'acide azido lipoïque et k = 4,3 × 10 <sup>-3</sup> M <sup>-1</sup> s <sup>-1</sup> pour la cycloaddition; Addition de coumarine par <sup>W37V</sup> LplA: k <sub>cat</sub> = 0,019 s <sup>-1</sup> , K <sub>m</sub> = 50 μM; <sup>W37I</sup> LplA: k <sub>cat</sub> = 0,016 s <sup>-1</sup> , K <sub>m</sub> = 261 μM.
CXPXR (6)	Sondes d'oxyamine ou d'hydrazide	Enzyme générant de la formylglycine (FGE)	Surface cellulaire	FGE catalyse la transformation d'un Cys de tag peptidique en une formylglycine.
LPXTG (5)	Sondes de peptide polyglycine	Sortase A (Srt A)	Surface cellulaire	SrtA clive le peptide au site Thr-Gly et attache un peptide polyglycine marqué; Aucune limitation de la taille de la modification introduite.
Auto-marquage des domaines enzymatiques				
Tags SNAP, CLIP (182)	Dérivés de benzylguanine	O6-alkylguanine-ADN alkyltransférase (hAGT) mutée	intra-cellulaire et surface cellulaire	hAGT est alkylée par transfert du groupe benzyle marqué de dérivés de O6-benzylguanosine.
Halo tag (296)	Sondes de chlorure d'alkyle	Haloalcane déshalogénase mutée (DhaA)	intra-cellulaire et surface cellulaire	DhaA muté est modifié de manière covalente par des sondes de chlorure d'alkyle. k <sub>obs</sub> = 2,7–8,5 × 10 <sup>6</sup> M <sup>-1</sup> s <sup>-1</sup>
Domaines de liaison de ligand				
eDHFR (159)	Les dérivés de Triméthylprim (TMP)	Non	intra-cellulaire	Liaison étroite et spécifique entre la Dihydrofolate réductase d' <i>Escherichia coli</i> (eDHFR) et le ligand de TMP; Marquage efficace dans les cellules dans les 10 minutes; Marquage réversible.

FKBP12 (108)	Conjugués SLF'	Non	intra-cellulaire	Liaison étroite et spécifique entre le mutant FKBP12 et le ligand du SLF'.
--------------	----------------	-----	------------------	--

Les défis associés au développement de réactions de modification de protéines sélectives au site sont multiples. Les réactions doivent être chimiosélectives pour une chaîne latérale d'acide aminé par rapport à toutes les autres sur la surface de la protéine. Les acides aminés qui sont peu abondants sur les surfaces protéiques et qui ont une réactivité unique se prêtent particulièrement bien aux chimies de modification sélectives au site. Les acides aminés naturels les plus importants répondant à ces critères se sont révélés être des résidus de cystéine (Cys) exposés en surface. Certains acides aminés naturels peuvent également avoir des propriétés uniques lorsqu'ils sont introduits à la position N-terminale d'une protéine. Par exemple, un Ser N-terminal peut être oxydé pour générer un aldéhyde qui peut ensuite être utilisé pour une modification sélective du site. De plus, à l'extrémité N-terminale, Cys peut réagir sélectivement avec les dérivés thioester pour donner une liaison peptidique native dans un processus appelé ligature chimique native (NCL). Les réactions de modification des protéines sélectives au site doivent en outre être compatibles avec les systèmes tampons aqueux, ne pas interférer avec le repliement correct des protéines et se dérouler à un pH presque neutre et à des températures modérées (c'est-à-dire 20 à 37 ° C). Comme les chances de rencontrer plusieurs protéines portant des acides aminés naturels même très peu abondants sur leur surface augmentent considérablement dans les mélanges complexes de protéines, il est généralement possible de marquer sélectivement les protéines recombinantes purifiées sur les acides aminés naturels. Un procédé permettant d'obtenir une modification sélective par site d'une protéine cible particulière dans un mélange complexe de protéines repose sur les progrès récents des techniques d'expansion du code génétique. Ces techniques permettent l'incorporation d'acides aminés non canoniques dans le ribosome dans des protéines présentant des fonctionnalités qui ne sont

normalement pas présentes dans les cellules vivantes (Fig. 7.2). Celles-ci comprennent des chaînes latérales comportant des groupes fonctionnels azide, alcyne, cétone, alcène ou tétrazine. Pour chacune de ces fonctionnalités, on peut envisager des réactions hautement chimiosélectives, voire mécaniquement spécifiques, de l'acide aminé non canonique par rapport à toutes les autres fonctionnalités présentes dans les cellules vivantes. Ce concept, appelé chimie bioorthogonale, permet aux chimistes d'installer des modifications sélectives sur les sites de protéines individuelles dans des mélanges biologiques complexes, y compris dans des organismes vivants. Les réactions de ligature bioorthogonale peuvent en outre être utilisées sur des préparations protéiques purifiées lorsque la stabilité de la liaison ou une précision de marquage supérieure à celle pouvant être obtenue avec des acides aminés natifs est requise (par exemple, la construction de conjugués anticorps-médicament "drug").



**Fig. 7.2. Stratégie conceptuelle pour obtenir une modification protéique sélective par site en utilisant des réactions chimiques biocompatibles *in vitro* et *in vivo*** (adaptée selon Krall *et al.* 2015). En règle générale, la modification chimique sélective de protéines de sites peut être réalisée soit (1) sur une chaîne latérale d'acides aminés naturelle à réactivité unique (par exemple, Cys) pour laquelle une copie est déjà présente sur la protéine d'intérêt ou peut être introduite par mutagenèse dirigée sur un site ou (2) sur un acide aminé non canonique présentant une chaîne latérale avec un groupe fonctionnel qui n'est normalement pas présent dans les systèmes vivants mais qui peut être introduit à l'aide de procédures de codage génétique (par exemple, azoture, alcyne, cétone, alcène et tétrazine). Dans le cas des chaînes latérales naturelles d'acides aminés, on se limite à la modification de protéines purifiées dans un tube à essai, de nombreuses autres biomolécules pouvant présenter la même fonctionnalité dans un mélange complexe de protéines. D'autre part, en utilisant des acides aminés non canoniques, il est possible de concevoir des réactions sélectives sur toutes les chaînes latérales naturelles des acides aminés, permettant ainsi la modification sélective en fonction du site de protéines individuelles dans des mélanges complexes et des organismes vivants. Lors du choix de la fonctionnalité à cibler chimiquement, la réaction avec une molécule d'intérêt modifiée appropriée (par exemple, une modification post-traductionnelle, un marqueur fluorescent ou un médicament cytotoxique) doit avoir lieu de manière chimiosélective dans des conditions physiologiques.

Les chimistes et les biologistes ont de plus en plus recours à des outils de modification de protéines sélectives au site pour étudier la fonction des modifications post-traductionnelles naturels. En outre, en s'attachant des modifications préparées synthétiquement qui ne sont généralement pas présentes dans la nature, ils tentent de conférer une fonctionnalité nouvelle à des protéines d'origine bactérienne ou eucaryote.

La prise de conscience de la nécessité de développer des méthodes de modification de protéines sélectives au site est encore récente. Malgré plusieurs avancées signalées au cours des dix dernières années, des réactions sites-sélectives supplémentaires et complémentaires, présentant une cinétique et des sélectivités améliorées, restent nécessaires.

## Chapitre 8. Ingénierie des protéines *in silico*

### 8.1. Prédiction de la structure des protéines

Les principales motivations pour la prédiction de la structure des protéines proviennent i) du désir fondamental des êtres humains d'explorer les lois naturelles et des besoins de tester ou de vérifier le mécanisme de repliement des protéines putatives; ii) les exigences pour déterminer les fonctions des séquences protéiques encodées dans l'ère post-génomique et pour rendre la découverte de médicaments plus rapide et plus efficace grâce à la substitution rapide pour des expériences de biologie structurale lente et coûteuse. Les méthodes de prédiction de la structure des protéines peuvent être classées en tant que méthodes fondées sur la physique et basées sur la bioinformatique. La base pour développer les approches purement physiques est la reconnaissance et la compréhension du mécanisme de repliement des protéines.

Un objectif idéal de telles méthodes est qu'elles sont capables de diriger une séquence d'acides aminés d'une protéine pour qu'elle se replie dans sa structure 3D native en imitant diverses forces motrices possibles sans connaissance dérivée de bases de données. Au contraire, les méthodes basées sur la bioinformatique reposent principalement sur diverses informations a priori dérivées de bases de données telles que les fonctions d'énergie statistique, la structure secondaire et les informations de classification des structures des protéines, en plus de l'information de la séquence d'acides aminés primaire.

#### 8.1.1. Méthodes basées sur la bioinformatique

L'évaluation critique des techniques de prédiction de la structure des protéines (CASP) est une expérience mondiale de prédiction de la structure des protéines basée sur l'ordinateur qui a lieu tous les deux ans depuis 1994. L'objectif principal de CASP est de faciliter le développement des approches pour prédire la structure de la protéine 3D à partir de sa séquence d'acides aminés. Actuellement, les résultats du classement montrent que la méthode bioinformatique la plus efficace et la plus précise est la modélisation par homologie, qui a non

seulement la vitesse de calculer des repliements approximatifs pour de grandes fractions de génomes entiers, mais aussi une précision suffisante pour les structures prédites avec des erreurs généralement supérieures à 3 Å si l'identité de la séquence entre la cible et le modèle est supérieure à 30% (par exemple, lancer un alignement en ligne BLAST "Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/>" d'une séquence protéique donnée contre la base de données, PDB "Protein Data Banque" et évaluer le degré d'homologie de cette séquence protéique par rapport aux structures déjà déposées dans PDB).

Pour les protéines globulaires à domaine unique ayant une longueur de moins de 90 acides aminés, plusieurs outils de serveur Web peuvent prédire les structures ayant souvent des valeurs  $C_{\alpha}$  RMSD "Root Mean Square Deviation", l'écart quadratique moyen d'environ 2 à 6 Å par rapport à leurs structures expérimentales. Récemment, une nouvelle méthode de prédiction de la structure basée sur l'homologie a obtenu une  $C_{\alpha}$  RMSD de 1 Å (moyenne de 1,6 Å) aux structures cristallines de référence du spectre complet des domaines de test dans les récentes expériences CASP. Cette approche est principalement basée sur la définition et l'extraction du PCAIN "Protein Core Atomic Interaction Network", le réseau d'interaction atomique à noyau protéique conservé dans des familles distinctes de domaines protéiques, qui se distinguent significativement parmi différentes familles de domaines et peuvent donc être considérées comme la «signature» du repliement natif d'un domaine protéique.

### 8.1.2. Méthodes basées sur la physique

En raison de l'existence objective de certaines limitations telles que la compréhension incomplète du mécanisme de repliement, la description inexacte des champs de force et la puissance de calcul insuffisante, il n'y a eu que quelques succès de méthodes basées sur la physique dans les structures informatiques de très petite taille. En dépit de ces succès limités de quelques études sur seulement de petites protéines simples, une fois que les approches basées sur la physique réussissent, les avantages seraient la capacité de prévoir des

changements conformationnels, tels que l'ajustement induit, un problème commun et important non résolu dans la découverte de médicaments computationnelle; la capacité de comprendre les mécanismes protéiques, les mouvements, les processus de repliement, les transitions conformationnelles et d'autres situations dans lesquelles le comportement des protéines nécessite plus qu'une simple connaissance de la structure native statique; la capacité de concevoir des protéines synthétiques pour de nouvelles applications ou de concevoir des polymères pliables à partir de squelettes non biologiques; et la capacité d'améliorer systématiquement la modélisation des protéines basée sur les lois de la physique. En effet, la combinaison de méthodes basées sur la physique et de prédictions structurelles basées sur la bioinformatique incubera des approches plus puissantes qui peuvent non seulement prédire avec précision et rapidité les structures protéiques, mais aussi faciliter la résolution des deux problèmes clés du repliement des protéines: le code de repliement et le taux de repliement.

## 8.2. Conception des protéines

Le développement des technologies de l'ADN recombinant et de la synthèse de peptide et petites protéines a permis de modifier spécifiquement les protéines et de les concevoir sur mesure. Le premier succès dans la conception d'une protéine *de novo* (*De Novo Protein Design*) se sont les travaux de Stephen Mayo *et al.*, 1997. Il s'agit de la synthèse d'un motif  $\beta\beta\alpha$  de 28 résidus dans le squelette devait avoir la conformation du motif en doigt à zinc de Zif268, une protéine de souris qui se lie à l'ADN, mais sans contenir d'ions métallique stabilisant. Ce polypeptide a été conçu sur ordinateur. Un autre exemple pratique, la subtilisine, une protéase qui dégrade efficacement les protéines qui salissent nos vêtements et donc ajoutée aux lessives. Cependant, ces lessives ne peuvent contenir d'eau de Javel, qui inactiverait totalement l'enzyme par oxydation d'une méthionine critique à la position 222. Si l'on remplace ce résidu par mutagénèse dirigée par un résidu alanine non réactif, l'enzyme survit aussi en présence d'eau de Javel. Le deuxième exemple concerne les enzymes

thermostables, qui sont fréquemment extraites des bactéries thermophiles: dans les sources volcaniques, les organismes vivants sont soumis à des températures de 100 °C et plus. Leurs protéines thermostables, comme par exemple l'ADN polymérase Taq de *Thermus aquaticus* sont d'une valeur biologique inestimable utilisés en génie génétique dans la PCR "réaction d'amplification des gènes en chaîne par polymérase" qui présente de nombreuses applications. L'imitation de ces enzymes naturelles a conduit à des variants plus thermostables de la subtilisine. On peut aussi obtenir des protéines ayant de nouvelles propriétés par fusion au niveau des gènes de protéines connues. Ces protéines chimériques peuvent réunir les propriétés indépendantes de deux protéines ou plus au sein d'un même polypeptide. Ces protéines peuvent servir comme des rapporteurs, exemple la protéine verte fluorescente, GFP qui en fusionnant avec la protéine d'intérêt devient détectable afin de suivre le devenir de cette protéine d'intérêt dans une condition biologique donnée.

### 8.3. La biologie synthétique

La biologie synthétique décrit l'ingénierie de parties biologiques et de systèmes entiers en modifiant des organismes naturels ou en construisant de nouveaux biosystèmes à partir de rien. Jusqu'à ce jour, la plupart des protéines utilisées en biologie synthétique proviennent de la nature. L'utilisation de protéines naturellement évoluées a conduit à de nombreuses applications réussies en biotechnologie. Néanmoins, ces applications bénéficient invariablement d'une optimisation des protéines naturelles originales par ingénierie des protéines. En revanche, la construction de protéines entièrement artificielles ne ressemblant pas aux protéines naturelles reste un défi majeur et est donc beaucoup moins répandue que l'ingénierie de protéines naturelles pour obtenir des propriétés nouvelles ou améliorées.

Avec le développement de la biologie synthétique, les enzymes devraient catalyser des réactions non reconnues ou des réactions non présentes dans la nature.

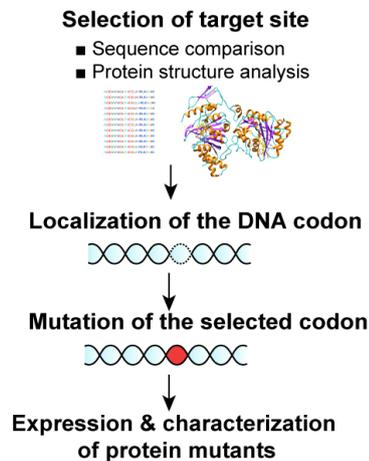
Actuellement, l'ingénierie enzymatique est principalement réalisée par évolution dirigée ou mutagenèse dirigée.

L'évolution dirigée est un outil puissant pour adapter les propriétés biomoléculaires. Pour les applications impliquant des protéines, l'approche nécessite une cartographie individuelle du phénotype et du génotype, puisque des sélections ou des analyses sont effectuées sur les protéines elles-mêmes, mais que l'amplification et l'identification des séquences désirées doivent être effectuées au niveau génétique. Cette liaison phénotype – génotype est facilement réalisée dans les cellules, qui peuvent naturellement compartimenter des séquences d'ADN individuelles et, par transcription et traduction, les protéines correspondantes. Cependant, l'utilisation de cellules réduit considérablement la taille des bibliothèques pouvant être échantillonnées, ce qui peut être vital, en particulier lorsque les fonctions biomoléculaires évoluent complètement *de novo*; En outre, l'évolution dirigée basée sur les cellules limite la portée des tampons, des solvants et des températures pouvant être utilisées, car la viabilité des cellules doit être maintenue.

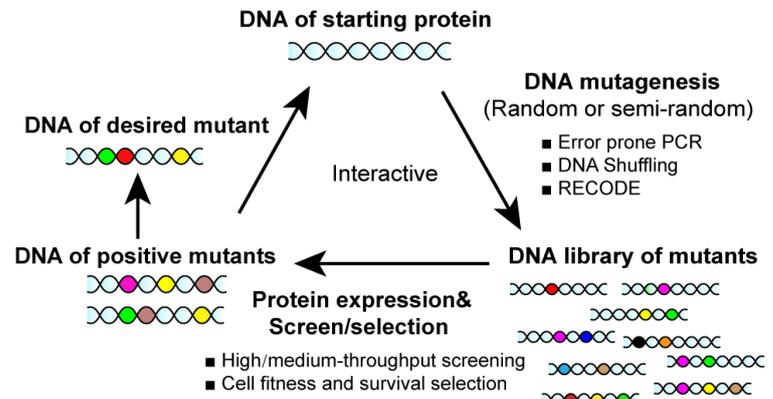
Les stratégies d'évolution dirigée ou de mutagenèse dirigée nécessitent un criblage approfondi et / ou la sélection d'un grand nombre de variants (mutants) ou d'expériences itératives pour intégrer différentes mutations positives (Fig. 8.1 A,B). Comparée à l'évolution dirigée ou à la mutagenèse dirigée, la conception de protéines *in silico* permet une évolution plus rapide des protéines. Comme la structure de la protéine dicte sa fonction, la conception de la protéine *in silico* cherche à construire de nouvelles structures qui correspondent à la fonction de la protéine souhaitée et à calculer des séquences dont la structure conçue est le repliement de l'énergie libre minimale (Fig. 8.1C). Il peut commencer à utiliser une structure protéique existante comme une chaîne principale (reconception de la protéine" protein redesign") ou à partir de zéro (conception de protéine *de novo*). L'application de la conception protéique

*in silico* a produit divers matériaux protéiques, biocapteurs, protéines thérapeutiques et biocatalyseurs.

### A Site-directed mutagenesis

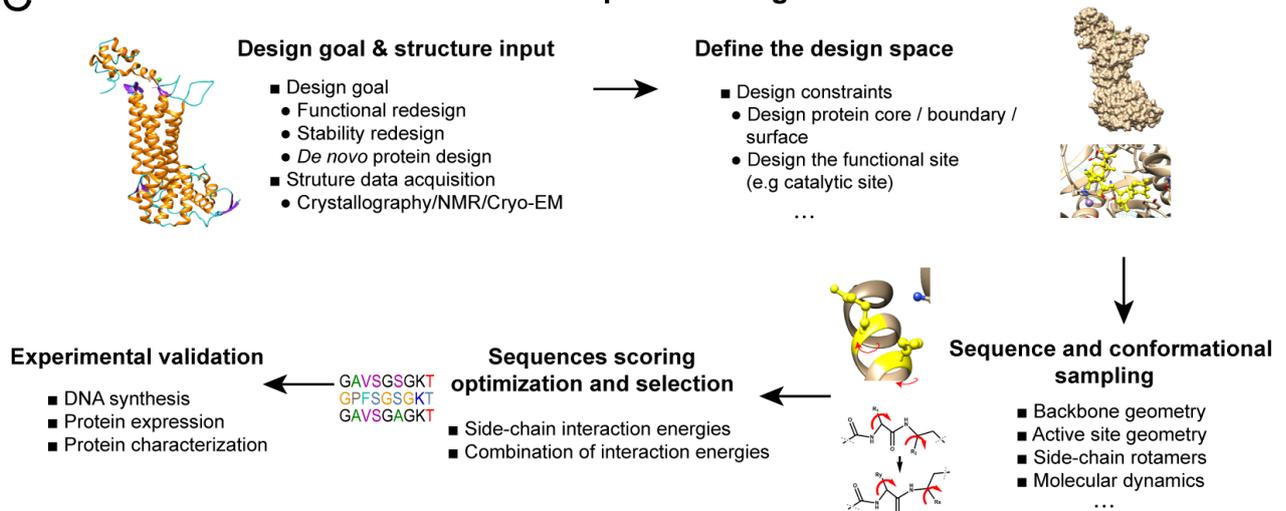


### B Directed evolution



### C

#### *In silico* protein design



**Fig. 8.1. Schémas de stratégies d'ingénierie des protéines (Wang *et al.*, 2018).**

(A) Mutagenèse dirigée. Les sites cibles peuvent être sélectionnés par comparaison de séquences ou analyse structurale.

(B) Evolution dirigée. Des banques de mutants sont créées avec une mutagenèse aléatoire ou semi-aléatoire telle qu'une réaction en chaîne de la polymérase sujette aux erreurs, un brassage de l'ADN "DNA shuffling" ou RECODE, et sont sélectionnées par croissance cellulaire et / ou par survie ou criblées via un criblage à haut / moyen débit.

(C) Conception de protéines *in silico*. Les objectifs de conception pourraient être de redéfinir la fonctionnalité ou la stabilité d'une protéine connue ou de concevoir une protéine de novo. L'espace de conception est défini en fonction des objectifs (par exemple, redéfinir la protéine pour augmenter la stabilité ou modifier la spécificité du substrat). Les conformations et séquences souhaitées sont calculées en fonction des contraintes définies.

En conclusion de ce chapitre, l'ingénierie des protéines est devenue un outil important pour surmonter les limites des enzymes naturelles en tant que biocatalyseurs. L'ingénierie des protéines utilise deux stratégies principales pour optimiser les propriétés de l'enzyme en vue de l'application souhaitée: conception rationnelle ou évolution dirigée. La conception rationnelle des protéines nécessite des informations structurelles et mécanistes sur l'enzyme d'intérêt (ou un bon modèle d'homologie). En utilisant ces données et la modélisation moléculaire, on peut généralement identifier quelques acides aminés intéressants pour modifier ou induire les propriétés souhaitées. Ensuite, les variants choisis sont créés en laboratoire et leurs fonctionnalités sont testées expérimentalement. D'autre part, l'évolution dirigée implique soit une mutagenèse aléatoire du gène cible codant pour le catalyseur (par exemple, une PCR sujette à erreur), soit la recombinaison d'un ensemble de gènes apparentés (par exemple, par un brassage d'ADN "DNA shuffling"). Au moyen de ces procédés, une librairie de mutants de la protéine est générée et ensuite criblée pour trouver les meilleurs mutants présentant la propriété souhaitée. Les variants présentant des résultats prometteurs peuvent être soumis à d'autres cycles d'évolution. Bien que les deux approches aient montré leur utilité, certaines limites doivent être prises en compte. Par exemple, si la conception rationnelle de protéines nécessite une compréhension profonde et précise de la protéine pour établir des prévisions fiables, l'évolution dirigée ne nécessite pas de données structurelles. De plus, l'effet de mutations simultanées est souvent extrêmement difficile à prédire par des méthodes rationnelles. Heureusement, de nos jours, les séquences, les structures et les données biochimiques des protéines sont de plus en plus disponibles, associées à des progrès considérables dans les technologies informatiques. D'un autre côté, lorsque vous utilisez une évolution dirigée, d'énormes bibliothèques de mutants sont généralement créées (> 10<sup>4</sup>–10<sup>7</sup> variants). Par conséquent, un système de criblage ou de sélection à haut débit doit être disponible. De plus, des biais dans les méthodes expérimentales et la dégénérescence du code

génétique limitent la conception de la librairie. Une autre alternative pour surmonter les contraintes de chaque méthode consiste à utiliser une conception semi-rationnelle, une combinaison de ces deux stratégies. Ce concept fusionne des informations mécanistiques et structurales, ainsi que des algorithmes prédictifs de calcul pour présélectionner des sites cibles prometteurs. L'accent mis sur des régions spécifiques de l'enzyme se traduit par une réduction considérable de la taille des librairies avec des taux d'accès encore plus élevés, ce qui est particulièrement avantageux.

Par ailleurs, la conception de protéines *de novo* est devenue un outil prometteur pour la génération de biocatalyseurs sur mesure. Actuellement, des efforts importants sont déployés pour élargir l'ensemble des réactions catalysées, y compris des réactions non biologiques telles que la transformation rétroaldol à plusieurs étapes (formation de liaisons carbone-carbone en chimie organique.). Généralement, ces enzymes conçues n'atteignent pas les activités affichées par les enzymes naturelles, mais elles peuvent être considérablement améliorées par l'évolution dirigée.

## Références

1. Boutureira, O. and G.a.J. Bernardes, *Advances in chemical protein modification*. Chemical Reviews, 2015. 115(5): p. 2174-2195.
2. Lane, M.D. and B. Seelig, *Advances in the directed evolution of proteins*. Current opinion in chemical biology, 2014. 22: p. 129-136.
3. Guilloton, M. and B. Quintard, *Biochimie* 2003: Dunod.
4. Voet, D., J.G. Voet, and L. Domenjoud, *Biochimie* 2016: De Boeck.
5. Botham, K.M., et al., *Biochimie de Harper* 2017: De Boeck Supérieur.
6. Kamoun, P., et al., *Biochimie et biologie moléculaire* 2003: Flammarion médecine-sciences.
7. Müller-Esterl, W., *Biochimie et biologie moléculaire: cours* 2007: Dunod.
8. Deléage, G. and M. Gouy, *Bioinformatique - 2e édition: Cours et applications* 2015: Dunod.
9. Karp, G., J. Isawa, and W. Marshall, *Biologie cellulaire et moléculaire* 2018: De Boeck supérieur.
10. Hingorani, K.S. and L.M. Gierasch, *Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge*. Curr Opin Struct Biol, 2014. 24: p. 81-90.
11. Dodevski, I., G.C. Markou, and C.A. Sarkar, *Conceptual and methodological advances in cell-free directed evolution*. Current opinion in structural biology, 2015. 33: p. 1-7.
12. Hartl, F.U. and M. Hayer-Hartl, *Converging concepts of protein folding in vitro and in vivo*. Nat Struct Mol Biol, 2009. 16(6): p. 574-81.
13. Dill, K.A., K.M. Fiebig, and H.S. Chan, *Cooperativity in protein-folding kinetics*. Proceedings of the National Academy of Sciences, 1993. 90(5): p. 1942-1946.

14. Sachsenhauser, V. and J.C. Bardwell, *Directed evolution to improve protein folding in vivo*. *Curr Opin Struct Biol*, 2018. 48: p. 117-123.
15. Mendonca, L., *Dynamique conformationnelle des protéines étudiée par dichroïsme circulaire résolu en temps*, 2013, Palaiseau, Ecole polytechnique.
16. Zorn, K., et al., *Engineering and application of enzymes for lipid modification, an update*. *Progress in lipid research*, 2016. 63: p. 153-164.
17. Sirisha, V.L. and A. Jain, *Enzyme Immobilization: An Overview on Methods, Support Material, and Applications of Immobilized Enzymes*. *Adv Food Nutr Res*, 2016. 79: p. 179-211.
18. Lafaye, C., *Etudes Biochimique et Structurale de DsbA1, DsbA2 et DsbA3 : les trois homologues à l'oxydoréductase de Thiol-disulfure DsbA chez Neisseria meningitidis.*, 2009, Université Joseph-Fourier - Greenberg I.
19. Tzul, F.O., D. Vasilchuk, and G.I. Makhatadze, *Evidence for the principle of minimal frustration in the evolution of protein folding landscapes*. *Proc Natl Acad Sci U S A*, 2017. 114(9): p. E1627-E1632.
20. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels*. *Nature structural biology*, 1997. 4(1): p. 10.
21. Yoo, Y.J., et al., *Fundamentals of Enzyme Engineering 2017*: Springer Netherlands.
22. Wang, Y., J. Chen, and Z. Kang, *In Silico Protein Design Promotes the Rapid Evolution of Industrial Enzymes*, 2018, ACS Publications.
23. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *In vivo aspects of protein folding and quality control*. *Science*, 2016. 353(6294): p. aac4354.
24. Du, X., et al., *Insights into protein–ligand interactions: mechanisms, models, and methods*. *International journal of molecular sciences*, 2016. 17(2): p. 144.

25. Benhabilès, N., A. Thomas, and B. Brasseur, *Les mécanismes de repliement des protéines solubles*. Biotechnologie, Agronomie, Société et Environnement, 2000. 4(2): p. 71-81.
26. Adam, P., et al., *Metallothioneins in prion-and amyloid-related diseases*. Journal of Alzheimer's Disease, 2016. 51(3): p. 637-656.
27. Guilloton, M., B. Quintard, and P.F. Gallet, *Mini manuel de biochimie: cours + exos + QCM-QROC 2013*: Dunod.
28. Goldenzweig, A. and S.J. Fleishman, *Principles of Protein Stability and Their Application in Computational Design*. Annu Rev Biochem, 2018. 87: p. 105-129.
29. Wu, Y.W. and R.S. Goody, *Probing protein function by chemical modification*. Journal of Peptide Science, 2010. 16(10): p. 514-523.
30. Lutz, S. and U.T. Bornscheuer, *Protein Engineering Handbook 2012*: Wiley.
31. Poluri, K.M. and K. Gulati, *Protein Engineering Techniques: Gateways to Synthetic Protein Universe 2016*: Springer Singapore.
32. Bornscheuer, U.T. and M. Hohne, *Protein Engineering: Methods and Protocols 2018*: HUMANA Press.
33. Liu, S.-Q., et al., *Protein folding, binding and energy landscape: A synthesis*, in *Protein engineering 2012*, InTech.
34. Hill, A.F., et al., *Protein Folding, Misfolding, and Disease: Methods and Protocols 2011*: Humana Press.
35. Dwevedi, A., *Protein folding: Examining the challenges from synthesis to folded form 2014*: Springer.
36. Hartl, F.U., *Protein Misfolding Diseases*. Annu Rev Biochem, 2017. 86: p. 21-26.
37. Bushmarina, N., et al., *Repliement des protéines: exemple de l'alpha-lactalbumine*. in *Journal de Physique IV (Proceedings)*. 2005. EDP sciences.

38. Spicer, C.D. and B.G. Davis, *Selective chemical protein modification*. Nature communications, 2014. 5: p. 4740.
39. Krall, N., et al., *Site-selective protein-modification chemistry for basic biology and drug development*. Nature Chemistry, 2015. 8: p. 103.
40. Arrault, A., *Stratégies de docking-scoring assistées par analyse de données. Application au criblage virtuel des cibles thérapeutiques COX-2 et PPAR gamma*, 2007, Université d'Orléans.
41. Tobi, D. and I. Bahar, *Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state*. Proceedings of the National Academy of Sciences, 2005. 102(52): p. 18908-18913.
42. Finkelstein, A.V., et al., *There and back again: Two views on the protein folding puzzle*. Phys Life Rev, 2017. 21: p. 56-71.
43. Weinman, S. and P. Méhul, *Toute la biochimie 2004*: Dunod.
44. Svendsen, A., *Understanding Enzymes: Function, Design, Engineering, and Analysis* 2016: Pan Stanford Publishing.
45. <https://www.malvernpanalytical.com/fr/products/technology/microcalorimetry/differential-scanning-calorimetry>.