

**REPUBLIQUE ALGERI ENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE**



**UNIVERSITE ABOU BAKR BELKAID-TLEMCEM  
FACULTE DES SCIENCES  
DEPARTEMENT DE CHIMIE**



**MEMOIRE DE MAGISTER**

**En chimie physique**

**OPTION : CHIMIE THEORIQUE ET MODELISATION MOLECULAIRE**

**INFLUENCE DU pH SUR LE SPECTRE D'ABSORPTION DES PROTEINES  
FLUORESCENTES : CAS DE LA CERULEAN**

**PRESENTE PAR :**

**Mr. ADIL MERABET**

Soutenu le :30 /06 / 2012 devant la commission d'examen :

<b>Pr. Saïd Ghalem</b>	Président	UAB Tlemcen
<b>Pr. Dalila Hammoutene</b>	Rapporteur	USTHB Alger
<b>Pr. Latifa Negadi</b>	Rapporteur	UAB Tlemcen
<b>Dr. Sihem Azizi</b>	Encadreur	UAB Tlemcen

**Année Universitaire 2011 -2012**

## **Remerciements**

J'exprime ma gratitude à Mme S.AZIZI, mon encadreur qui n'a pas hésité à me faire confiance. Je tiens à la remercier pour sa patience à mon égard durant les étapes de ce travail, pour ses conseils ainsi que son soutien permanent. Je lui suis très reconnaissant d'avoir partagé avec moi sa grande expérience.

Je remercie Mr S.GHALEM, de m'avoir fait l'honneur de présider le jury d'examen.

Je remercie également Mme. D. Hammoutene et Mme L.NEGADI pour l'honneur qu'elles m'ont fait de bien vouloir examiner mon travail.

Je remercie tous ceux qui ont participé, de près ou de loin, à mon modeste travail, en particulier : ma grande famille ainsi que ma petite famille, mes amis : Karim et Aymen.

J'exprime ma reconnaissance à tous les enseignants qui ont contribué à ma formation tout au long de mon cursus.

# Sommaire

Introduction Générale .....	5
Chapitre I: Généralités sur les protéines .....	7
I.1 Définition des protéines .....	7
I.1.1 Fonctions biologiques des protéines .....	9
I.1.2 Structure des protéines .....	10
I.1.3 Classement des protéines .....	16
I.1.4 Propriétés des protéines .....	18
I.2 Protéines fluorescentes .....	20
I.2.1 Historique .....	20
I.2.2 Structure des protéines fluorescentes .....	22
I.2.3 Pourquoi le marqueur GFP est-il révolutionnaire ? .....	25
I.2.4 La famille de la GFP .....	26
Chapitre II: Etude de la cerulean .....	28
II.1 paramètres de la fluorescence .....	29
II.1.1 Rendement quantique, temps de vie et constante de vitesse .....	29
II.1.2 FRET (fluorescence resonance energy transfer) .....	31
II.1.3 Calcul du déclin de l'intensité de fluorescence .....	31
II.2 Présentation et structure de la Cerulean .....	33
II.3. Mutation H148D - de la ECFP à la cerulean .....	35
II.4 réseau de liaisons H au voisinage du chromophore .....	37
II.5 Influence du pH sur le spectre d'absorption .....	39
Chapitre III: Résultats et interprétations .....	41
III.1 dynamique moléculaire de la Cerulean :Etat de protonation de la protéine .....	42
III.2 Calcul du spectre d'absorption par simulation de dynamique moléculaire .....	43
III.3. spectre d'absorption de la cerulean .....	45
Conclusion générale .....	48
Références .....	49
Annexes .....	51

# Introduction Générale

La fluorescence est une propriété qui a fait de la protéine un outil révolutionnaire dans la recherche scientifique (biologique, imagerie cellulaire, etc...).

L'idée d'utiliser des molécules fluorescentes est ancienne. Réémettant une lumière d'une certaine couleur lorsqu'ils sont éclairés, ces colorants originaux ont donné naissance à la microscopie par fluorescence, ensemble de techniques faisant appel à des microscopes spéciaux. Mais ces molécules lumineuses ont longtemps été toxiques, limitant leur usage.

Au début des années 1960, un biologiste marin qui était aussi chimiste, le japonais Osamu Shimomura, s'est intéressé à une méduse du Pacifique, *Aequorea victoria*, capable, comme d'autres organismes marins, d'émettre de la lumière par fluorescence. Cette dernière utilise deux molécules, la première, baptisée aequorine, émettant une lumière bleue (395 nm étant l'idéal) qui excite la seconde, la Protéine Fluorescente Verte (Green Fluorescent Protein : GFP), laquelle réémet une lumière verte (508 nm).

Le prix nobel de chimie de 2008 a été attribué à Osamu Shimomura pour sa découverte et à deux américains Roger Tsien et Martin Chalfie qui ont mis au point l'utilisation de la GFP comme marqueur.

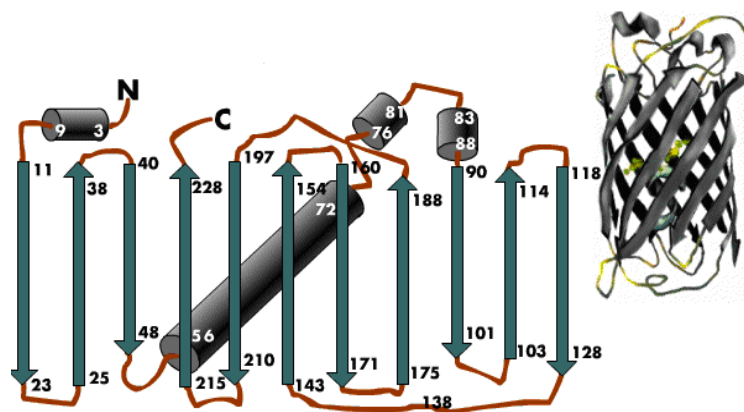


Figure 1. Structure tridimensionnelle de la GFP

La GFP est une petite protéine de 238 acides aminés sous forme d'un tonneau, au centre duquel se trouve fixée, la partie fluorescente de la protéine appelée chromophore, qui est issue d'une réaction de cyclisation entre trois acides aminés (ou résidus) consécutifs : la serine, la tyrosine et la glycine respectivement en position 65, 66, et 67 (Ser65, Tyr66, Gly67).

Les protéines fluorescentes, homologues de la GFP sont devenues ces dernières années des outils incontournables en imagerie cellulaire. Elles offrent la possibilité d'explorer un nombre toujours croissant de processus biologiques en cellules ou tissus vivants. Leur succès repose sur le fait qu'elles peuvent être exprimées dans tous les organismes vivants et toutes les subdivisions du milieu cellulaire (noyau, mitochondrie,...)

Cependant, leur utilisation est à l'heure actuelle encore limitée par les propriétés des protéines elles-mêmes. Les expérimentateurs qui utilisent ces protéines fluorescentes ou qui développent de nouvelles méthodes d'imagerie utilisant ces protéines ont un réel besoin de connaître précisément leurs caractéristiques photophysiques.

Les approches numériques peuvent apporter une compréhension microscopique des processus mis en jeu, complémentaire aux études expérimentales et contribuer à l'interprétation des propriétés photophysiques de ces protéines.

Ce mémoire s'inscrit dans l'étude des propriétés photo-physiques de la Cerulean: un des mutants de la GFP, qui est devenu un outil incontournable dans l'imagerie cellulaire vu sa capacité à s'introduire dans les systèmes vivants et la large gamme de fluorescence qu'elle couvre.

La Cerulean est obtenue en faisant une série de substitution des résidus de la GFP dans la principale est la mutation Y66W : la substitution de tyrosine 66 par le tryptophane, qui conduit à un déplacement du spectre d'absorption vers les 435 nm (bleu).

Pour comprendre le phénomène de la fluorescence il est crucial de l'associer aux propriétés physicochimiques des protéines : la répartition dans l'espace des résidus et la structure tridimensionnelle sont des facteurs très importants, c'est ce que nous présentons dans le premier chapitre.

Les propriétés photophysiques de la Cerulean ; le calcul de son spectre d'absorption et l'interprétation de son comportement lors de la variation du pH, en tenant compte de la structure de la protéine, feront l'objet de notre deuxième chapitre.

Dans le troisième chapitre nous calculons le spectre d'absorption de la Cerulean en fonction du pH que nous comparerons aux résultats théoriques et expérimentaux.

Nous terminons ce mémoire par une conclusion.

Enfin, Les méthodes de calculs utilisées seront quant à elles présentées en annexes.

# **Chapitre I**

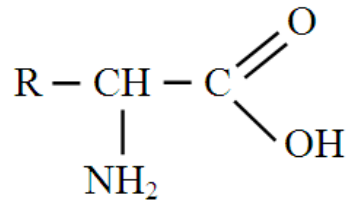
## **Généralités sur les protéines**

## **I.1 Définition des protéines :**

Les protéines ont été mises en évidence en 1838 par le chimiste hollandais Gerhard Mulder (1802-1880). Le terme protéine vient du grec prôtos qui signifie premier, essentiel. Ceci fait probablement référence au fait qu'elles sont indispensables à la vie et qu'elles constituent souvent la part majoritaire du poids sec des organismes (plus de 60% du poids sec de la cellule) [1].

En Biochimie, une protéine se définit simplement comme une molécule complexe (macromolécule) composée d'une ou de plusieurs chaînes d'acides aminés, eux-mêmes liés par des liaisons peptidiques. La liaison peptidique se fait entre le groupement acide (COOH) d'un acide aminé et le groupement amine (NH<sub>2</sub>) de l'autre [2].

Un acide aminé est un composé bifonctionnel comportant un groupe acide carboxylique – COOH un groupe amine –NH<sub>2</sub> portés par le même atome de carbone C<sub>α</sub> (c'est un carbone chiral : il contient quatre groupes différents) qui porte aussi un atome d'hydrogène et un radical -R. De tels composés protidiques sont appelés acides α-aminés. Ils ont comme formule générale [3] :



Les acides aminés se différencient les uns des autres par leur radical. On peut donc théoriquement combiner une infinité d'acides aminés. Cependant, on constate que chez l'homme, comme chez de nombreuses espèces, seuls vingt acides aminés différents constituent les protéines (voir tableau 1).

Le radical peut contenir un groupe aliphatique (-H, -CH<sub>3</sub>...), une fonction supplémentaire (-OH : hydroxyle ou -SH : sulfhydrile), une deuxième fonction acide (-COOH) ou une deuxième fonction amine (-NH<sub>2</sub>), il peut contenir aussi un noyau aromatique (le cas de la phénylalanine et la tyrosine) ou un hétérocycle (le cas du tryptophane et l'histidine) [4].

Ces vingt acides aminés sont représentés dans le tableau 1 [3] :

Nom complet de l'acide aminé	Code à trois lettres	Code à une lettre
Alanine	Ala	A
Arginine	Arg	R
Acide aspartique (aspartate)	Asp	D
Asparagine	Asn	N
Cysteine	Cys	C
Acide glutamique (glutamate)	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Tableau 1 : Codes internationaux des acides aminés

Généralement, on part d'un minimum de 100 acides aminés (c'est très peu) pour parler d'une protéine. Concernant des ensembles plus petits d'acides aminés on parlera naturellement de peptides ou de polypeptides [2].



### **I.1.1 Fonctions biologiques des protéines :**

Les fonctions biologiques des protéines sont très diverses au sein de la cellule et de l'organisme. On distingue généralement les fonctions cellulaires qui définissent le rôle de la protéine dans la cellule ou l'organisme, et les fonctions biochimiques, définissant l'activité des protéines au niveau moléculaire.

A/ Les fonctions cellulaires peuvent être regroupées en divers groupes, on peut citer :

- Les protéines de structures : telles que le collagène et la kératine, qui permettent à la cellule de maintenir son organisation dans l'espace. Ces protéines peuvent former des tubes ou des filaments qui s'assemblent pour former des fibres très résistantes. Elles forment des structures aussi diverses que les tendons et les ligaments, l'armature interne des cellules, les toiles d'araignées, les cornes, les poils, les fibres des caillots sanguins...etc [5].
- Les protéines de transport : la fonction spécifique de ces protéines est de transporter une substance spécifique d'un endroit vers un autre (exemple : L'hémoglobine transporte l'oxygène des poumons vers les cellules). Des protéines membranaires spécifiques effectuent un type de transport très différent, ces protéines fixent des molécules d'un métabolite (molécules qui interviennent ou se forment au cours d'une biosynthèse) qui se trouve d'un côté de la membrane puis les libèrent de l'autre côté, elles permettent par exemple la pénétration des substances nutritives essentielles telles que le glucose ou les acides aminés dans la cellule [6].
- Les protéines motrices : kinésines et dynéine. Elles constituent une autre classe de protéines qui interviennent dans les mouvements. Ces protéines assurent le transport des organites (éléments constitutifs d'une cellule vivante ayant des fonctions bien précises ; exemple : noyau, mitochondrie...qui rentrent dans la synthèse des protéines) et des vésicules vers différents compartiments de la cellule en se déplaçant sur le microtubule [7].

- Les protéines de signalisation : elles captent les signaux extérieurs, et gèrent leur communication dans la cellule ou l'organisme. Ces récepteurs sont capables de reconnaître une molécule dans l'environnement cellulaire et de prévenir la cellule. Les récepteurs hormonaux font partie de ce type de protéines [8].

B/ Les fonctions biochimiques sont nettement plus nombreuses, et sont généralement équivalentes à la notion d'activité enzymatique. Les enzymes sont des protéines douées d'activité catalytique spécifique (biocatalyseurs) : chaque enzyme ne catalyse qu'un seul type de réactif et n'agit que sur un seul substrat. Elles accélèrent les réactions chimiques nécessaires à la vie avec une spécificité qui élimine la formation de sous-produit et on les retrouve intactes à la fin de la réaction. Les réactions pouvant être catalysées par les enzymes s'effectuent dans des conditions souvent qualifiées de douces, c'est-à-dire à la température de l'organisme qui les abrite (37°C pour l'organisme humain) et à un pH peu éloigné de la neutralité (aux alentours de 7) cependant l'enzyme ne peut pas catalyser une réaction thermodynamiquement impossible (énergie libre positive). Exemple d'enzymes et leurs rôles :

- La trypsine coupe de façon spécifique les liaisons peptidiques du côté carboxylique de la cystéine et l'arginine, la chymotrypsine coupe les liaisons après phénylalanine, tryptophane ou tyrosine (du côté carboxylique) alors que la pepsine coupe les liaisons du côté de l'azote terminal des acides aromatiques d'un groupement aminé.
- La saccharase catalyse l'hydrolyse du saccharose dans le tube digestif de l'insecte en quelques secondes alors que par voie chimique cette réaction prend des mois voire des années [9].

Il est à noter qu'une protéine peut avoir plusieurs fonctions cellulaires, et porter plusieurs fonctions biochimiques.

### **I.1.2 Structure des protéines :**

On distingue 4 types de structures de complexité croissante pour caractériser une protéine :

### - Structure primaire :

La liaison peptidique forte (covalente) en est responsable, elle présente des atomes adjacents dans un même plan d'où sa coplanarité.

La liaison du carbone carbonyle avec l'azote dans la liaison peptidique (1,33, non indiquée dans la figure 2) est plus courte que la liaison simple C-N mais plus longue qu'une liaison double C=N classique [10].

Ce caractère partiel de double liaison est responsable de la restriction de la rotation autour de cette liaison C-N.

Il existe donc deux conformations possibles pour la liaison: cis ou trans. C'est la conformation trans qui est observée car favorisée énergétiquement, parce que la forme cis est plus encombrée.

Les mouvements de rotation restent possibles entre le carbone  $\alpha$  et l'azote du groupe amine et entre le carbone  $\alpha$  et le carbone du groupe carboxylique.

On comprend donc aisément que le squelette axial est non spécifique d'une protéine mais que c'est une structure commune. Ce qui confère l'identité d'une protéine est l'ensemble des chaînes latérales des acides aminés qui la composent. Cet ensemble constitue la signature de la protéine [4,11].

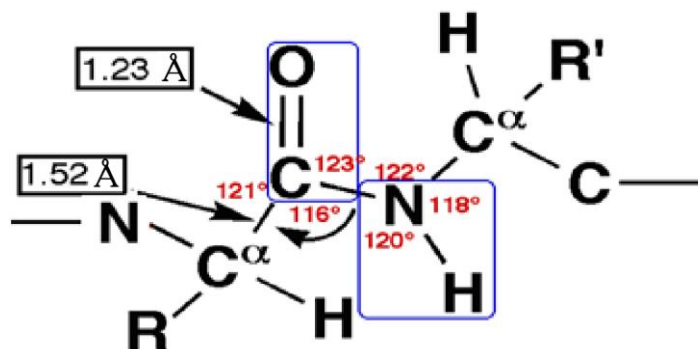


Figure 2 : structure primaire et propriétés d'une liaison peptidique [10]

### - Structure secondaire

Certaines parties de la chaîne d'acides aminés adoptent une structure régulière appelée structure secondaire. On reconnaît deux grands types de structure secondaire voir figure 3

a- L'hélice alpha

Dans la structure dite en hélice alpha, la chaîne d'acides aminés prend la forme d'un tire-bouchon. Les différentes spires sont stabilisées par des liaisons hydrogène.

b- Le feuillet bêta

Dans un feuillet bêta, il se forme des liaisons hydrogène entre certains segments de la chaîne disposés parallèlement les uns par rapport aux autres. L'ensemble forme comme une membrane plissée.

Une protéine est donc faite d'hélices alpha et de feuillets bêta reliés par des segments qui n'ont pas de structure secondaire particulière [6].

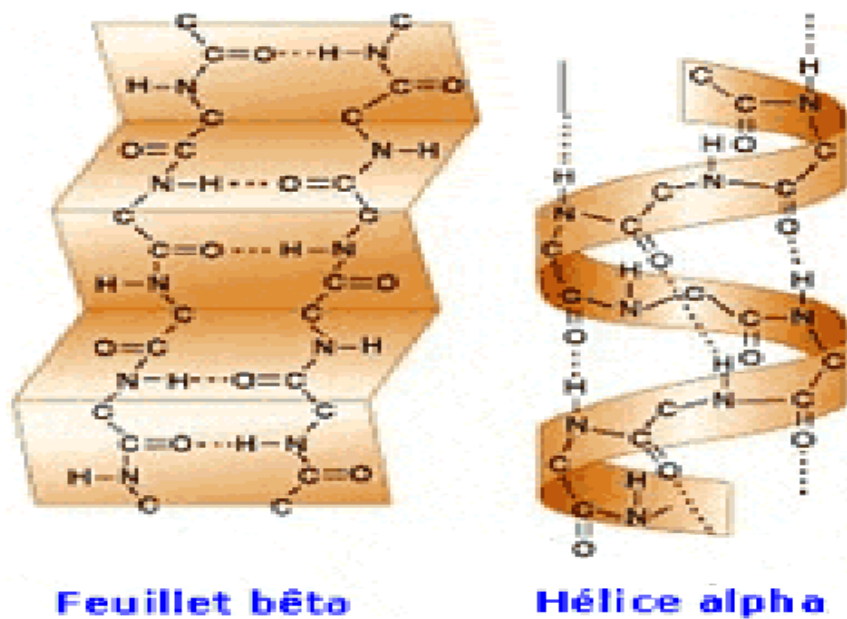


Figure 3 : Structure secondaire d'une protéine [1]

Le type de structure secondaire constituant la molécule est en liaison avec la forme globulaire ou fibreuse de celle-ci [12].

- protéines globulaires :

Les protéines globulaires ont une forme compacte. Les résidus hydrophiles se retrouvent souvent à la surface; les résidus hydrophobes le plus souvent à l'intérieur. Les ponts disulfures peuvent se former plus facilement à l'intérieur des protéines

globulaires, parce qu'ils nécessitent des conditions oxydantes et que le cytoplasme (milieu extérieur de la protéine) est généralement réducteur [13].

Ces protéines possèdent des cavités et des protubérances sur leur surface (appelé hème pour l'hémoglobine par exemple, cet hème contient un atome de fer qui peut se lier avec la molécule de dioxygène). D'autres biomolécules peuvent s'insérer dans les cavités et se fixer aux protubérances par un mécanisme semblable à celui d'une clé et d'une serrure. Par leurs motifs de surface, les protéines globulaires interagissent de façon spécifique avec d'autres molécules, cette interaction est appelée stéréospécificité.

Exemples : enzymes : catalyseurs de réactions biologiques

Anticorps : reconnaissance moléculaire

Éléments de transport et d'entreposage

Hormones : messagers chimiques [14], exemple : insuline

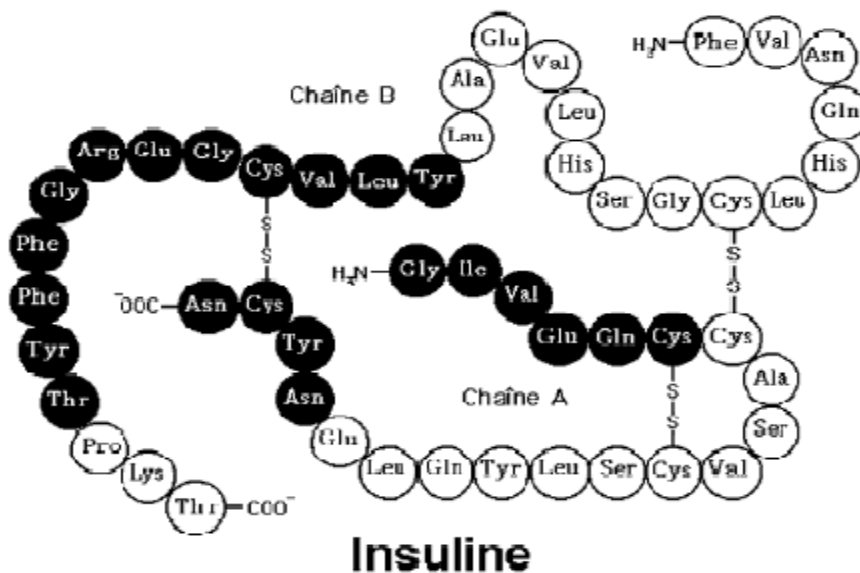


Figure 4 : Structure d'une protéine globulaire : insuline [15]

- protéines fibreuses :

Puisque toutes les protéines sont à la base une longue chaîne polypeptidique, on peut les représenter comme des spaghettis. Alors qu'une protéine globulaire est une boule de spaghettis refroidis et figée, une protéine fibreuse s'étire en longueur [14]. Exemple : le collagène.

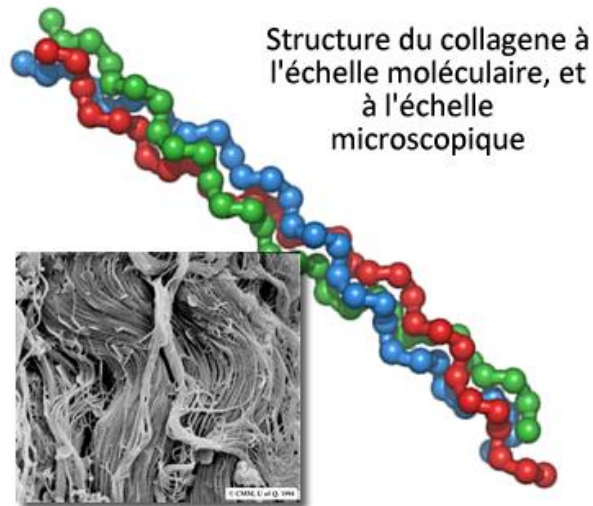


Figure 5 : structure du collagène [16]

### - Structure tertiaire

Elle ne s'applique obligatoirement qu'aux protéines globulaires. Cette structure tridimensionnelle très compacte, enroulée comporte des structures secondaires et des segments sans structure secondaire.

On retrouve donc également les interactions ioniques, les interactions hydrophobes (fortes au centre de la protéine), les liaisons hydrogène stabilisant les repliements, les forces de Van der Waals et les ponts disulfures. Ces derniers se créent après achèvement du repliement tridimensionnel de la protéine et ont un rôle prépondérant dans le maintien de la structure tertiaire. La structure tertiaire d'une protéine globulaire est sa conformation tridimensionnelle biologiquement active, toute modification de sa structure interrompt son activité. La structure tridimensionnelle, la stéréospécificité et le rôle biologique d'une protéine sont nécessairement liés. C'est ce qui définit la notion de site actif : la conformation spatiale de la protéine rapproche des acides aminés normalement éloignés et forme le site actif, qui est souvent logé dans la crevasse [6,17].

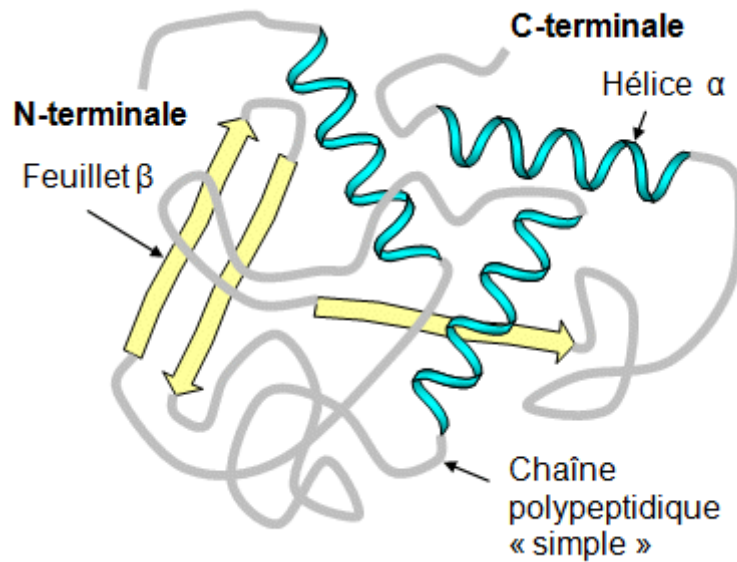


Figure 6 : Structure tertiaire d'une protéine [17]

### -Structure quaternaire

La structure quaternaire correspond à l'association spécifique de plusieurs chaînes peptidiques en une unité d'ordre supérieur seule capable d'assurer complètement les fonctions biologiques telles que les catalyseurs, les protéines de transport...etc. Une telle protéine dite oligomère (hémoglobine, enzyme comme la thréonine désaminase ...) est constituée par l'association de plusieurs chaînes peptidiques ou protomères (deux chaînes alpha et deux chaînes beta pour l'hémoglobine humaine normale). L'assemblage réalisé par des liaisons secondaires se fait de façon spécifique et selon une certaine symétrie. Il conditionne l'activité de la protéine [12].

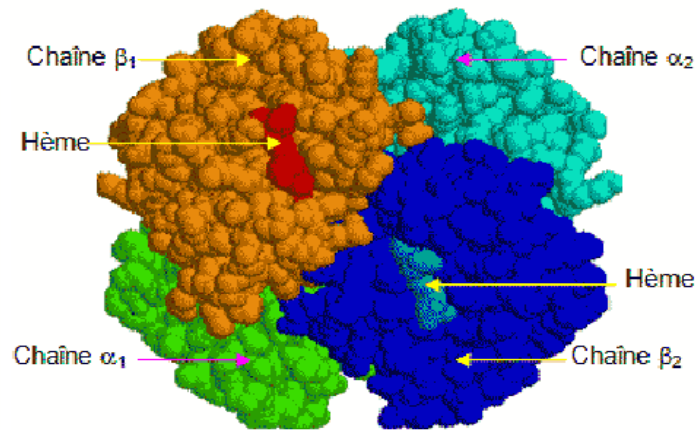


Figure 7 : les 4 chaînes (globines) formant la molécule d'hémoglobine [17]

### **I.1.3 Classement des protéines :**

Les protéines sont tellement nombreuses et leur rôle dans l'organisme est tellement varié qu'on les classe en différentes catégories.

A/ Suivant leur composition chimique, on classera les protéines en deux groupes: les holoprotéines et les hétéroprotéines.

- **Les holoprotéines :** comme le suggère le préfixe, sont les protéines composées exclusivement d'acides aminés. Elles sont divisées en protamines, histones, gluténines, albumines, prolamines, globulines et scléroprotéines ;

- Protamines : ce sont des protéines à faible poids moléculaire qui peuvent être considérées comme des polypeptides, elles sont présentes dans les noyaux des cellules des spermes de poisson (laitance), solubles dans l'eau, incoagulables par la chaleur, leurs pHi (Par définition le pH iso-ionique correspond au pH de la solution quand le nombre des protons  $H^+$  libérés est égal au nombre des  $H^+$  captés) est de l'ordre de 12, riches en acides aminés basiques (arginine en particulier) et pauvres en acides aminés aromatiques et acides. Exemple : Salmine (laitance du saumon)
- Histones : sont de masse moléculaire plus élevée et moins basiques que les protamines (pHi = 10), elles sont solubles dans l'eau et coagulables à la chaleur,



elles sont présentes dans les globules rouges des oiseaux, les leucocytes et dans les noyaux des cellules animales.

- Prolamines : ce sont des protéines végétales complexes, elles sont solubles dans l'éthanol (à 70-80%), riches en prolines et acide glutamique et pauvres en lysine et glycines. Exemple : Zéine de maïs.

- **Les hétéroprotéines** : sont composées d'une partie protéique et d'une partie non protéique appelée groupement prosthétique, ce dernier est important car il détermine les différentes classes des hétéroprotéines selon leur nature :

- Un acide orthophosphorique pour les phosphoprotéines. Ce groupe est relativement riche en acide phosphorique. Exemple : la caséine (dans le lait et le jaune d'œuf)
- Un pigment pour les chromoprotéines, dont le groupement prosthétique lui confère une couleur donnée. Exemple : l'hémoglobine.
- Un acide nucléique pour les nucléoprotéines, présentes dans les cellules, en particulier dans le noyau. Elles participent à la synthèse des protéines de l'organisme. Exemple : ribonucléoprotéine (associée à l'ARNm)
- Un glucose pour les glycoprotéines, comme la mucine qui joue le rôle de protecteur dans la paroi du tube digestif.
- Un lipide pour les lipoprotéines, ce groupe joue des fonctions vitales à l'échelle cellulaire. Exemple : les lipoprotéines membranaires qui assurent les échanges entre la cellule et le milieu extérieur.
- Du fer, du zinc, du calcium, du molybdène ou du cuivre pour les métalloprotéines. Exemple : la céruloplasmine qui renferme du cuivre [12].

B/ Suivant le rôle biologique des protéines dans l'organisme, on détermine deux grandes classes de protéines, les fibreuses et les globulaires.

- **Les protéines fibreuses** : font partie des tissus et des éléments structurant du corps comme les muscles, la peau, les os, les composants cellulaires ou leurs membranes. Leur rôle est d'assurer une fonction structurelle par la construction de structures simples et répétées (comme le collagène et la kératine).

- **Les protéines globulaires** : ont un rôle important à plusieurs niveaux du métabolisme, elles remplissent d'autres fonctions. Dans ce groupe, on compte les enzymes dont le rôle est primordial au maintien de la vie de l'organisme. On compte aussi les hormones dans le groupe des protéines globulaires (comme l'insuline), on parle généralement d'hormones peptidiques. Enfin, les protéines motrices sont constituées de la myosine des fibres musculaires, les protéines de transport (hémoglobine...), les protéines de l'immunité (immunoglobuline du petit lait...) et les protéines de stockage comme l'ovalbumine de l'œuf [18].

La classification des protéines peut se faire aussi selon leurs propriétés.

#### **I.1.4 Propriétés des protéines** :

Les principales propriétés des protéines sont :

##### **I.1.4.1 propriétés physiques** :

Les protéines sont des substances solides qui se décomposent bien avant la fusion. Elles se présentent souvent sous forme de poudre amorphe, on peut aussi les obtenir sous forme cristallisée, toutefois leur cristallisation n'est pas un critère de pureté.

- ***Solubilité:***

On considère la solubilité aqueuse puisque les protéines sont généralement insolubles dans les solvants organiques (sauf dans l'alcool).

La solubilité des protéines dans l'eau est variable, elle dépend du milieu : le milieu est-il acide ? alcalin ? effet des sels, des métaux, effet de la température, effet du solvant organique.

Les facteurs de la solubilité sont conditionnés par la proportion des groupements polaires et leur répartition.

- ***Propriétés des solutions protéiques:***

Les protéines solubles forment avec l'eau des solutions macro moléculaires. Ces solutions présentent des propriétés particulières:

a/ la viscosité: Les solutions protéiques sont visqueuses, leur viscosité dépend de la forme, la taille et la concentration des molécules.

*b/ Propriétés optiques* : Les solutions protéiques ne sont pas limpides, elles absorbent et diffusent de la lumière (effet Tyndall), l'axe d'orientation du faisceau lumineux dépend de l'activité optique de la protéine [19]. Quand on projette un faisceau lumineux très fin sur une solution protéique, une partie de la lumière est diffusée dans toutes les directions. Le rapport de l'intensité de la lumière diffusée sur celle de la lumière incidente est d'autant plus grand que les particules dissoutes sont plus nombreuses et plus volumineuses et que la direction d'observation est plus proche de celle de la lumière incidente. On peut donc calculer le poids moléculaire d'une protéine en solution [20].

#### **I.1.4.2 Propriétés chimiques:**

Les propriétés chimiques des protéines sont celles des chaînes latérales des acides aminés, elles ont comme propriété principale l'amphotérie : toutes les protéines possèdent un nombre important de groupements ionisables, (groupement amide, acide carboxylique, thiol, phénol). Selon le pH de la solution par rapport à leur pK, ces groupements sont plus ou moins ionisés et la protéine se comportera soit comme un acide ou une base, soit comme un anion (excès de charges négatives) ou comme un cation.

#### **I.1.4.3 Propriétés biologiques :**

Presque toutes les protéines, de même que les polypeptides de masse moléculaire suffisante présentent un pouvoir antigénique. Introduite dans un organisme étranger, elles déterminent la production d'anticorps cellulaires ou circulants. Certaines sont douées de propriétés toxiques (exotoxines bactériennes), enzymatiques, hormonales ou anticorps [19].

Le tableau ci dessous présente les différentes protéines (enzyme, anticorps, antigène, récepteur hormonal et les protéines d'un système de signalisation) avec leurs propriétés :

- type d'activité : la fonction biologique que la protéine exerce
- ligand : la molécule avec laquelle la protéine agit
- nature et devenir du complexe : le produit résultant de la réaction protéine-ligand [21]

Tableau 2 : Propriétés biologiques des protéines

Type de molécule	Enzyme	Anticorps ou antigène	Récepteur hormonal	Protéine d'un système de signalisation
<i>Type d'activité</i>	Catalyseur	Réaction immunitaire	Régulation de l'activité cellulaire : Etape de fixation de l'hormone      Etape ultérieure d'une chaîne de signalisation	
<i>Ligand</i>	Substrat	Antigène ou anticorps	Hormone	Elément d'un système de signalisation (second messenger. Ca <sup>+2</sup> , protéines,...)
<i>Nature et devenir du complexe</i>	Complexe enzyme substrat ; modification de structure d'où transformation du substrat en produit	Complexe Antigène-anticorps	Complexe hormone récepteur, ou protéine régulatrice ligand ; modification de structure qui entraîne une modification des propriétés (catalyse ou possibilité d'interactions avec d'autres molécules d'une chaîne de signalisation).	

## **I.2 Protéines fluorescentes :**

### **I.2.1 Historique :**

Les protéines fluorescentes, homologues de la *Green Fluorescent Protein* (GFP) sont devenues ces dix dernières années des outils incontournables en imagerie cellulaire.

En 1960, **Osamu Shimomura** vient d'achever sa thèse de doctorat en chimie organique au Japon. Il rejoint alors le laboratoire de Frank Johnson à Princeton (USA) afin d'étudier la bioluminescence de la méduse *Aequorea victoria*, phénomène naturel qu'il avait décrit cinq ans plus tôt.



Figure 8 : La méduse *Aequorea victoria* [22]

En 1962, il met en évidence l'existence de deux protéines fluorescentes : l'aéqorine et la GFP. La première émet une lumière bleutée et la seconde une lumière verte. Il faudra attendre les années 1969 et 1971 pour que Shimomura et ses collègues révèlent le mécanisme entre l'aéqorine et la GFP qui est à l'origine de cette bioluminescence verte. Ces deux protéines agissent de façon complémentaire, l'aéqorine absorbe une longueur d'onde excitatrice et émet une lumière plus faible (bleue), cette lumière excite à son tour la GFP qui, en dernier émet une lumière verte. En fluorescence, un tel phénomène est connu sous le terme de FRET (Fluorescence Résonance Energy Transfer).

D'autres chercheurs s'intéressent à la biochimie de la GFP modifiée [21]. Cette modification, d'abord énigmatique, va être élucidée au cours des années suivantes. Lorsque **Martin Chalfie** entend parler pour la première fois de la GFP, au début des années 1990, il se montre très vite enthousiaste : ce neurobiologiste a alors l'idée de fusionner la GFP avec d'autres protéines, et d'obtenir ainsi des protéines chimères fluorescentes capables de servir de marqueur cellulaire *in vivo*. En 1992, il obtient de Prasher un clone du gène *GFP* et réussit à obtenir avec succès son expression en l'injectant dans la bactérie *Escherichia coli*. Puis, il reproduit l'expérience sur son modèle d'étude : le ver nématode *Caenorhabditis elegans*.

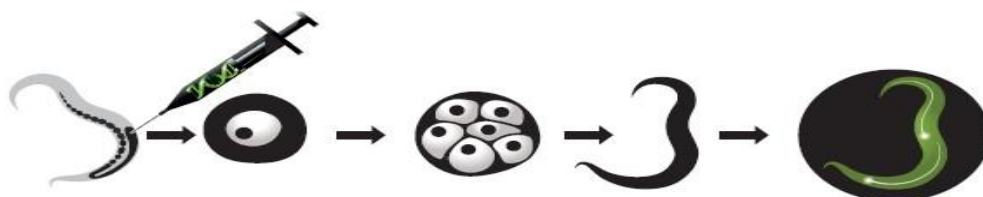


Figure 9 : l'injection de la GFP sur le ver *C. elegans* [23]

Sa réussite n'a été publiée qu'en 1994 dans la revue *Science* et a ouvert la voie à une méthode très prometteuse d'exploration de l'expression des gènes : l'engouement est au rendez-vous, et dès la même année, plusieurs équipes se lancent à leur tour dans l'aventure !

**Roger Tsien** fait partie de ces biologistes ; avec ses collègues, ils remarquent que le gène *gfp* ne fluoresce pas en absence d'oxygène (chez *E. coli* cultivée en anaérobiose) malgré sa bonne expression. Ils déterminent alors le rôle de l'oxygène dans la formation du chromophore. Poursuivant leurs travaux, les chercheurs élucident les mécanismes chimiques impliqués dans la maturation du chromophore. Par la suite, Tsien contribue à l'élaboration de différentes formes mutantes de la GFP, modifiant sa structure biochimique afin d'émettre une fluorescence à différentes longueurs d'ondes excitatrices, ou dans une large gamme de couleurs [20].

### **I.2.2 Structure des protéines fluorescentes**

La GFP dans sa structure primaire est une chaîne de 238 acides aminés pesant environ 27.000 unités de masse atomique (27 KDaltons ; 1 Dalton = 1 unité de masse atomique), sa structure secondaire montre une série d'hélices et de feuillets plissés stabilisés par des liaisons hydrogène. Quant à sa structure tertiaire, elle représente un tonneau de 42 Å de longueur et 24 Å de diamètre, il est formé de 11 feuillets  $\beta$  (voir figure 10).

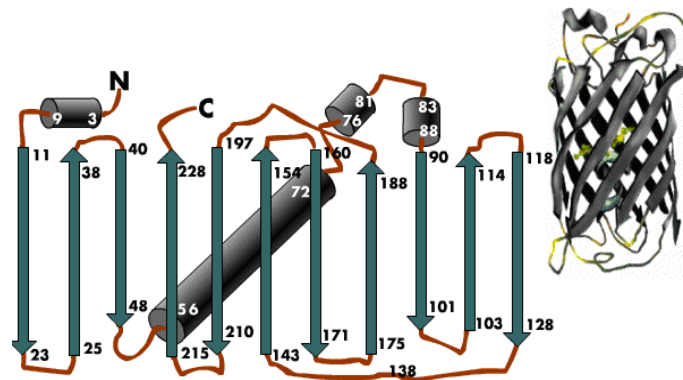
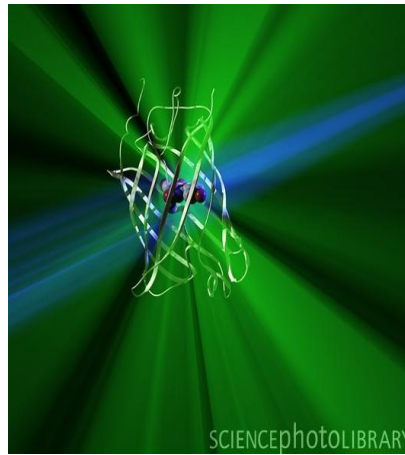


Figure 10 : Structure Tridimensionnelle de la GFP [24]

Au centre de ce tonneau se trouve, fixée sur une hélice  $\alpha$ , la partie fluorescente de la protéine appelée chromophore (figure 11.a), un groupement chimique qui absorbe la lumière bleue et UV et qui fluoresce en vert, il est issu d'une réaction de cyclisation entre trois acides aminés (ou résidus) consécutifs : la serine, la tyrosine et la glycine respectivement en position 65, 66, et 67 (Ser65, Tyr66, Gly67) (figure 11.b).

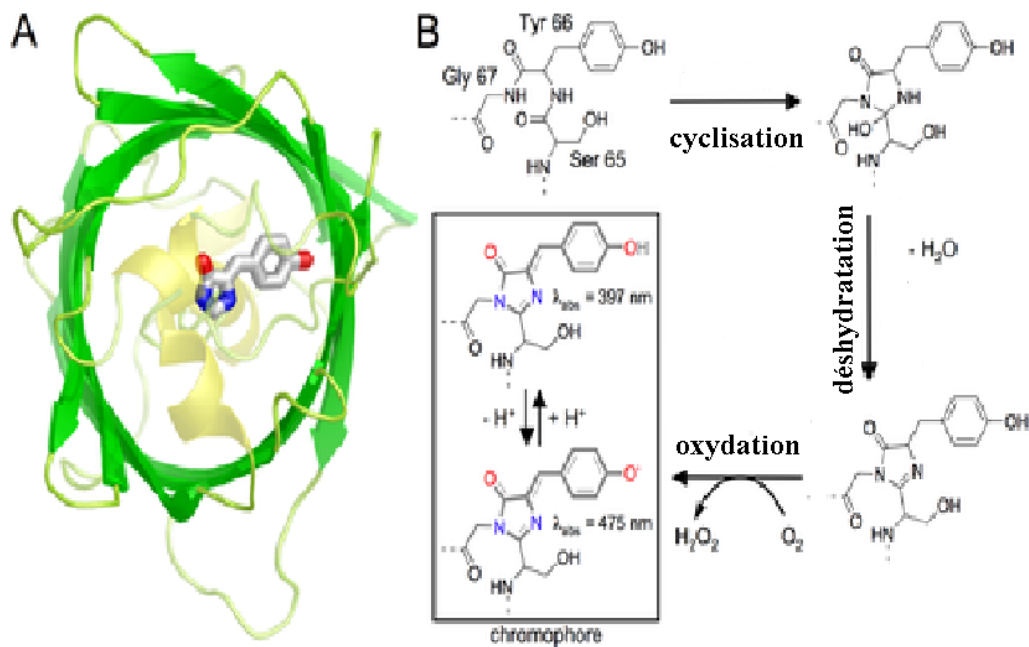


Figure 11 : A. vue d'en haut de la GFP

B. synthèse intramoléculaire du chromophore de la GFP [25]

La GFP non modifiée, dite sauvage (wild type : wGFP) a deux maxima d'excitation. Le premier se trouve à une longueur d'onde de 395 nm (lumière UV), le second à 475 nm (lumière bleue). La longueur d'onde d'émission maximale est de 504 nm.

La structure tridimensionnelle de la GFP protège le chromophore, ce qui la rend stable [21]. En outre, la GFP présente d'autres propriétés biochimiques :

- la tolérance à des agents physiques :
  - des températures jusqu'à 65°C ;
  - un pH de 5,5 à 12,2 ;
- la tolérance à des agents chimiques :
  - traitement avec une concentration de 8 moles/litre d'urée : dénaturant, il facilite la pénétration de l'eau dans l'intérieur protéique hydrophobe ;
  - 1% SDS (Sodium Dodecyl Sulfate) : détergent anionique hautement chargé dénature la chaîne protéique ;
- elle résiste à un traitement prolongé avec des protéases (enzyme hydrolysant la liaison peptidique d'une protéine)
- elle forme des dimères
- la forme monomérique n'est pas fluorescente
- la fluorescence ne nécessite ni substrat ni cofacteur.



- dans le chromophore le remplacement de la Ser65 par la Thr65 conduit au mutant GFP65T excitable à 490 nm et 6 fois plus fluorescent [26].

### **I.2.3 Pourquoi le marqueur GFP est-il révolutionnaire ?**

Quand Antoine Van Leeuwenhoek utilise le microscope au 17<sup>ème</sup> siècle, un nouveau monde s'ouvre. Les scientifiques ont soudainement pris conscience d'organismes dont ils n'imaginaient même pas l'existence auparavant comme les bactéries ou les cellules.

Au sein de la cellule, les processus biochimiques sont en général régulés par les protéines, dont les variétés fonctionnelles sont multiples. En couplant la protéine GFP -facilement traçable de par sa fluorescence verte- à une de ces protéines cellulaires, les chercheurs peuvent à présent suivre son comportement, ses mouvements et ses interactions avec le milieu de la cellule. Elle devient alors "le microscope" des biochimistes, biologistes et autres chercheurs dans le domaine médical. Ainsi dans les applications directes, la GFP permet d'illuminer la croissance des tumeurs cancéreuses, le développement de la maladie d'Alzheimer ou l'évolution de bactéries pathogènes [24].

Le fait que la fluorescence de la GFP ne nécessite pas de cofacteur et ne modifie pas le fonctionnement normal ou la localisation de la protéine fusionnée; elle est particulièrement utile en biologie cellulaire, pour la fabrication de protéines de fusion, elle fournit un marqueur pour localiser la protéine recombinante dans la cellule. Cela facilite l'étude du trafic intracellulaire des protéines et même le transport des protéines entre les cellules. Parmi les premiers exemples de cette application, l'utilisation de la GFP pour suivre le mouvement des particules ribonucleoproteiques chez *Drosophila* par Wang et al en 1994 [25].

Les protéines de la famille de la GFP peuvent également être utilisées comme indicateur actif. Ces sondes émettent alors un signal différent en fonction du milieu et de l'environnement dans lequel on se trouve. Ce type de sensibilité est cependant limité à cause de la rigidité de ces protéines et du fait que le chromophore soit confiné à l'intérieur de la protéine [26].

Loukianov et son groupe [27] ont constaté que les protéines fluorescentes peuvent agir en tant que donneurs d'électrons lorsqu'elles sont excitées. Selon leur publication dans la revue « Nature Chemical Biology », le transfert d'électrons du chromophore vert excité (GFP) à un accepteur d'électrons passe par une étape intermédiaire courte. Si aucun accepteur n'est disponible en permanence, la GFP perd sa fluorescence. Cependant, si elle réagit avec un accepteur d'électrons la protéine fluoresce en rouge.

### **I.2.4 La famille de la GFP**

Les recherches sur la protéine fluorescente verte (GFP) ont permis l'identification de plus de 150 gènes distincts qui, avec leurs mutants, couvrent toute la gamme d'excitation de 380 à 600 nm et la gamme d'émission de 440 à 650 nm.

Elles diffèrent généralement de la GFP sauvage par, au plus, une dizaine de mutations.

(La mutation : consiste à substituer un acide aminé par un autre avec un pK différent et par conséquent il n'aura pas la même conformation, les mêmes liaisons hydrogène, pas les mêmes interactions avec le chromophore... ce qui pourrait fortement modifier le spectre d'absorption).

Il existe deux types de mutations, d'une part celles qui affectent le chromophore, soit la mutation de la tyrosine 66 par un autre acide aminé possédant un cycle aromatique (histidine, phénylalanine ou tryptophane), ou la mutation de la serine 65 en thréonine, glycine, glutamine ou méthionine, et d'autre part les mutations des résidus qui sont proches du chromophore dans la structure tridimensionnelle : par exemple, la mutation de la thréonine 203 par une tyrosine, (figure 12.A)

La figure 12.B illustre une structure typique d'un peptide, les feuillets  $\beta$  sont représentés par des cylindres verts avec une flèche pointant vers la partie carbone terminal et les hélices  $\alpha$  sont représentées par des barils bleus. Les acides aminés mutants sont représentés en ellipses regroupées dans des séries de différentes couleurs. Chaque série représente un déplacement du spectre vers la couleur correspondante : BFP (bleu), CFP (cyan), GFP (vert), YFP (jaune), et Saphir (violet). Les ellipses grises représentent les mutations communes aux autres séries. Il est à noter que près de 75 % des mutations sont situées dans l'hélice centrale et les feuillets  $\beta$  7, 8 et 10.

Malgré la diversité spectrale, toutes ces protéines, représentant la famille de la GFP, possèdent la même structure, les mêmes caractéristiques biochimiques et photophysiques.

Les protéines fluorescentes peuvent être divisées en plusieurs groupes selon leur couleur de fluorescence dans la gamme visuelle:

- **Blue, BFP (< 460 nm)**
- **Cyan, CFP (460-500 nm)**
- **Green, GFP (~500-520 nm)**
- **Yellow, YFP (~520-550 nm)**
- **Orange, OFP (~550-570 nm)**
- **Red, RFP (~570-620 nm)**

- Far red, FRFP (> 620 nm) [28]

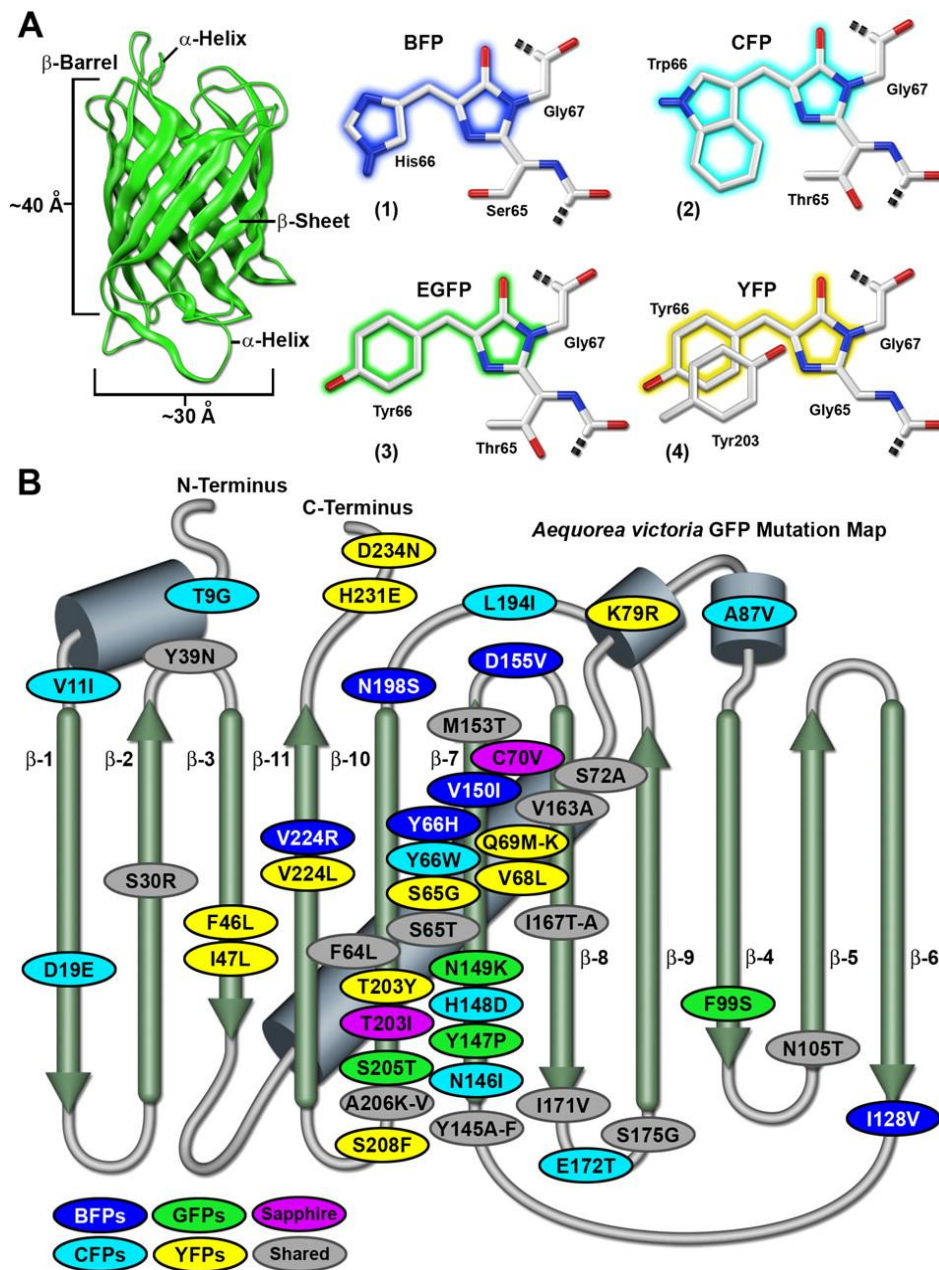


Figure 12 : Les modifications des mutations sur la famille de GFP [29]

# **Chapitre II**

## **Etude de la Cerulean**

## **II.1 Paramètres de la fluorescence :**

*Le but de cette partie est de donner le vocabulaire et les grandeurs associées à l'étude des processus photo-physiques de molécules en phase condensée. L'ensemble des processus photo-physiques se divisent en deux familles : les processus radiatifs dans lesquels le système absorbe ou émet un photon et les processus non radiatifs dans lesquels aucun échange de photon n'a lieu*

### **II.1.1 Rendement quantique, temps de vie et constante de vitesse :**

A chacun des processus photo-physiques est associée une constante de vitesse qui donne la fréquence avec laquelle se produisent ces processus. Ces constantes de vitesses ne sont pas des observables directement mesurables expérimentalement. Nous allons présenter les observables couramment employées et leur lien avec ces constantes de vitesse.

Après le passage de l'impulsion laser, une population de molécules est portée à l'état excité et atteint rapidement l'état S1 (premier état excité de la couche S). L'évolution de la population de S1 est le plus souvent donnée par la relation suivante :

$$\frac{dN(t)}{dt} = -k_r N(t) - k_{nr} N(t) \quad (\text{II.1})$$

où  $N(t)$  est la population de l'état excité,  $k_r$  est la constante de vitesse de la fluorescence et  $k_{nr}$  est la constante de vitesse non radiative associée aux processus non radiatifs éventuels. Dans le cas où les constantes de vitesse sont indépendantes du temps, cette équation a une solution très simple :

$$N(t) = N(0) \exp\left(-\frac{t}{\tau_s}\right) \quad (\text{II.2})$$

où  $\tau_s = (k_r + k_{nr})^{-1}$  est homogène à un temps.  $\tau_s$  définit la durée de vie de l'état excité et quantifie le temps que le système reste à l'état excité et pendant lequel on observe la fluorescence. Plus le nombre de processus non radiatifs est élevé plus ce temps sera court. Le temps de vie de l'état excité peut être facilement obtenu à partir des expériences de

fluorescence résolues en temps. Celles-ci mesurent le déclin de l'intensité de fluorescence en fonction du temps à partir d'un temps  $t=0$  correspondant à l'absorption des photons. L'intensité de fluorescence,  $I_f(t)$ , correspond au nombre de photons émis à l'instant  $t$ . Elle est proportionnelle à la population  $N(t)$ .

$$I_f(t) = k_r N(t) = k_r N(0) \exp\left(-\frac{t}{\tau_s}\right) \quad (II.3)$$

Dans ce cas simple, la fonction  $I_f(t)$  est une exponentielle décroissante. On peut alors évaluer  $\tau_s$  à partir de la pente du logarithme de la courbe  $I_f(t)$ . Expérimentalement, la quantité mesurée est proportionnelle à  $I_f(t)$  et dépend des conditions opératoires. Cependant cela ne gêne en rien la détermination de  $\tau_s$ . Le temps de vie de l'état excité  $\tau_s$  ne doit pas être confondu avec le temps de vie de fluorescence  $\tau_f$ . Il correspond au temps de vie de l'état électronique excité si le seul mécanisme de retour à l'état fondamental était la fluorescence et est défini par  $\tau_f = 1/k_r$ . Ce dernier n'est pas mesurable directement expérimentalement car la fluorescence ne peut pas être séparée des processus non radiatifs. On peut l'obtenir de façon indirecte grâce au rendement quantique de fluorescence  $\phi_f$ . Il est mesuré expérimentalement et est défini comme le rapport du nombre de photons émis sur le nombre de photons absorbés. Plus généralement l'efficacité des réactions photochimiques ou des processus photophysiques est mesurée ou exprimée en terme de rendement quantique défini comme le rapport du nombre de fois que le processus s'est produit par photon absorbé. L'intensité de fluorescence étant proportionnelle au nombre de photons émis, le rendement quantique de fluorescence est donné par l'intégrale de l'intensité de fluorescence normalisée à  $t=0$ .

$$\phi_f = \frac{1}{N(0)} k_r \int_0^{\infty} N(t) dt \quad (II.4)$$

Dans le cas simple présenté ici, il s'exprime plus simplement de la façon suivante

$$\phi_f = \frac{k_r}{k_r - k_{nr}} = \frac{\tau_s}{\tau_f} \quad (II.5)$$

Connaissant le rendement quantique de fluorescence et le temps de vie de l'état excité, on peut donc calculer le temps de vie de fluorescence. Le rendement quantifié de fluorescence donne une bonne indication de l'efficacité de la fluorescence du système car il quantifie le nombre de photons émis par photon absorbé. La « brillance » d'un système est définie comme le produit du coefficient d'extinction molaire  $\epsilon(\lambda)$  et du rendement quantique de fluorescence  $\phi_f$ . Elle peut également être utilisée pour quantifier l'efficacité de la fluorescence.

### **II.1.2 FRET (fluorescence resonance energy transfer) :**

Le FRET est une technique qui permet de détecter la proximité de deux molécules fluorescentes. La première, (dite le "donneur"), est excitée par une longueur d'onde  $\lambda_1$  (dans l'UV, par exemple). Le donneur va libérer des photons de longueur d'onde  $\lambda_2$  (dans le bleu, par exemple), qu'on peut détecter. Il peut aussi transférer par résonance une partie de son énergie d'excitation à la seconde molécule fluorescente (le "receveur") qui libèrera aussi des photons, mais à une longueur d'onde qui lui est propre,  $\lambda_3$  (dans le jaune, par exemple). Puisqu'il y a transfert d'énergie entre le donneur et le receveur (et que rien ne se perd, rien ne se crée), plus le transfert d'énergie est efficace et plus le receveur émet de photons par rapport au donneur.

Le transfert d'énergie du donneur au receveur, souligné par la détection de l'émission de lumière  $\lambda_3$ , indique que le donneur et le receveur ont été très proches l'un de l'autre car l'efficacité du FRET est proportionnelle à l'inverse de la puissance 6 de la distance entre les deux molécules.

Les deux molécules doivent être séparées par 10-100 Å ; le spectre d'absorption de l'accepteur doit chevaucher le spectre d'émission du donneur; et les orientations des dipôles du donneur et de l'accepteur doivent être à peu près parallèles. La distance à laquelle le transfert est efficace à 50% est appelée distance de Förster et varie selon les agents fluorescents utilisés [30].

### **II.1.3 Calcul du déclin de l'intensité de fluorescence :**

Le déclin de l'intensité de fluorescence, correspond au nombre de photons émis en fonction du temps.

Il est directement relié au déclin de la population de chromophore à l'état électronique excité et donc à la population  $N(t)$  de l'état excité :

$$k_r N(t) = k_r N(0) \exp\left(-\frac{t}{\tau_s}\right) \quad (II.6)$$

Pour calculer le déclin de la fluorescence, il faut résoudre l'équation différentielle :

$$\frac{dN(t)}{dt} = -k_r N(t) - k_{nr} N(t) \quad (II.7)$$

où  $k_r$  est la constante de vitesse de la fluorescence et  $k_{nr}$  est la constante de vitesse non radiative associée aux processus non radiatifs éventuels. Dans le cas où les constantes de vitesse sont indépendantes du temps, cette équation a une solution très simple :

$$N(t) = N(0) \exp\left(-\frac{t}{\tau_s}\right) \quad (II.8)$$

où  $\tau_s = (k_r + k_{nr})^{-1}$  est homogène à un temps.  $\tau_s$  définit la durée de vie de l'état excité et quantifie le temps que le système reste à l'état excité et pendant lequel on observe la fluorescence.

C'est une équation du premier ordre avec un second membre dépendant du temps. La solution de cette équation est :

$$N(t) = N(0) \exp(-k_r t) \left[ 1 - \sum_n \lambda^n \int_0^t p_n(t') dt' \right] \quad (II.9)$$

Où les fonctions  $p_n(t)$  sont les distributions des temps de  $n^{\text{ème}}$  passage à la géométrie  $TW_\varphi$  qui correspond à une conformation "twistée" où les deux cycles aromatiques du chromophore (cycle d'indole et cycle issu de la formation du chromophore) sont perpendiculaires l'un par rapport à l'autre.

En utilisant des méthodes multiconfigurationnelles de haute précision, Martin et al [31], Toniolo et al [32], Altoe et al [33], Olsen et al [34] ont recherché les chemins d'énergie minimum du chromophore anionique excité dans le vide. Ils ont mis en évidence deux géométries d'énergie plus basse que l'état FS (FS est la géométrie dans laquelle le chromophore fluoresce, elle correspond au premier état excité qui présente un faible transfert de charge par rapport à l'état fondamental), qui correspondent à des conformations



"twistées" où les deux cycles aromatiques sont perpendiculaires l'un par rapport à l'autre. L'une de ces géométries correspond à  $\varphi = \pm 90^\circ$  et  $\tau = 0^\circ$ . Ils l'ont notée par la suite  $TW_\varphi$ . L'autre correspond à  $\tau = \pm 90^\circ$  et  $\varphi = 0^\circ$ , ils l'ont appelée  $TW_\tau$ . Ces deux géométries sont présentées sur la figure 13. D'après les résultats de Martin et al,  $TW_\varphi$  se situe à environ 7 kcal.mol<sup>-1</sup> en dessous de la géométrie FS et  $TW_\tau$  se situe à 9 kcal.mol<sup>-1</sup> en dessous de la géométrie FS.

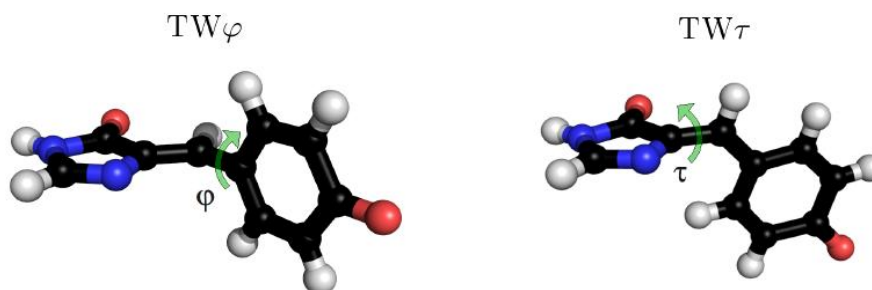


Figure 13 : Aperçu des géométries  $TW_\varphi$  et  $TW_\tau$ . [35]

Où  $\tau$  et  $\varphi$  sont des angles de torsion entre les deux cycles.

## II.2 Présentation et structure de la Cerulean :

L'utilisation des protéines fluorescentes est, actuellement, d'une grande importance dans la recherche scientifique, et pour répondre aux besoins des chercheurs, de nombreux mutants différents de la GFP ont été conçus, notamment les mutants de couleur telle que la protéine fluorescente cyan (ECFP : Enhanced Cyan Fluorescent Protein et la Cerulean).

La Cerulean et le CFP font partie d'un groupe de mutants de la GFP dont le cycle phénol du chromophore est substitué par un cycle indole, ce qui correspond à la mutation d'une tyrosine par un tryptophane (Tyr66Trp ou Y66W). Cette mutation conduit à un décalage des spectres d'absorption et d'émission vers le bleu [35].

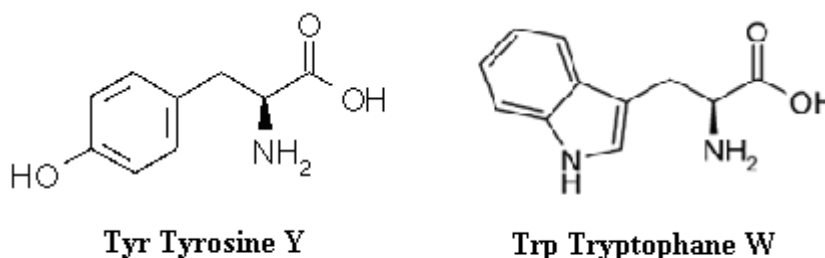


Figure 14 : représentation de la Tyrosine et de la Tryptophane

La ECFP est obtenue à partir de la GFP sauvage en faisant la mutation Y66W (substitution de la tyrosine 66 par la tryptophane) et cinq autres mutations : S65T (la sérine 65 par la thréonine), F64L (la phénylalanine 64 par la leucine), N146I (l'asparagine 146 par l'isoleucine), M153T (la méthionine 153 par la thréonine), V163A (la valine 163 par l'alanine); ces cinq mutations sont nécessaires pour restaurer la luminosité de ce chromophore modifié en raison de l'encombrement du groupe indole et permettent entre autres d'augmenter le rendement quantique de la fluorescence [36].

La structure PDB (Protein Data Bank : répertoire qui archive les structures des protéines obtenues par cristallographie aux rayons X) de la ECFP a deux conformations « A » et « B », son spectre d'absorption est caractérisé par 2 maxima à 430 nm et 450 nm, il est peu dépendant du pH.

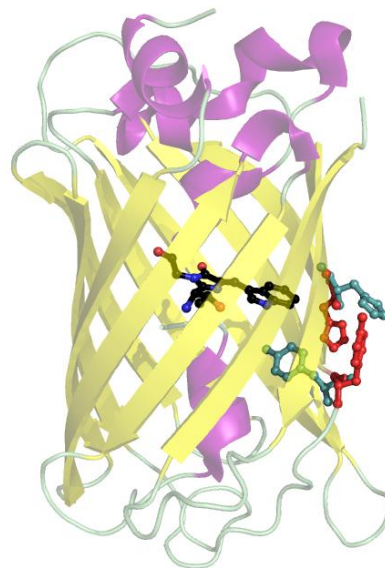


Figure 15 : Structure de la ECFP. Tyr145 et His148 sont représentés en bleu pour la conformation A (code PDB 1OXD) et en rouge pour la conformation B (code PDB 1OXE). [35]

Cependant une étude par dynamique moléculaire [35] a écarté l'hypothèse que ces deux maxima correspondent aux deux conformations de la protéine. Ils proviennent de la résonance du noyau indole du tryptophane Trp66. Ceci a été justifié par l'étude du spectre d'absorption de l'indole qui contient aussi deux pics à 280 et 290 nm [31,39].

Cette hypothèse est en bon accord avec le spectre d'absorption du miellat (un mutant rouge de la GFP) où le chromophore est à base de tryptophane (Trp) avec un environnement de résidu très différent par rapport à celui de la Cerulean.

### **II.3. Mutation H148D - de la ECFP à la Cerulean :**

L'intérêt suscité par la ECFP pour son utilisation dans les expériences de FRET a poussé les biologistes à chercher des mutants de la ECFP avec des déclin de fluorescence plus simples mais dont les spectres d'absorption et d'émission sont situés dans la même gamme de longueur d'onde. Considérant que la complexité du déclin de fluorescence de la ECFP est due à la présence de deux conformations A et B, Rizzo et al. [40] ont proposé de substituer les acides aminés Histidine 148 et Tyrosine145 pour supprimer une des 2 conformations. Il s'agit des mutations H148D : substitution de l'histidine (hydrophobe) par l'acide aspartique (hydrophile) et Y145A : substitution de la tyrosine (hydrophile) par l'alanine (hydrophobe), [39], et la mutation S72A : substitution de la serine par l'alanine. Rizzo et al [40] ont émis l'hypothèse que la protéine s'est stabilisée dans la conformation où la chaîne latérale de l'acide aspartique est orientée vers le solvant et celle de l'alanine vers l'intérieur et obtenir ainsi le résultat voulu : un déclin de fluorescence plus simple et dans la même gamme du spectre d'absorption et d'émission de la ECFP.

Il s'est avéré [41] après des simulations de dynamique moléculaire que l'acide aspartique ASP148 s'est mis à l'intérieur de la protéine et s'est stabilisé par une liaison hydrogène avec le chromophore entre les deux atomes d'oxygène de la fonction carboxylique de l'acide aspartique et le groupement N-H de l'indole du chromophore [42]. Cette structure correspond à la conformation cis du chromophore à pH 5.

Une telle position de la chaîne latérale de ASP148 a été observée dans la GFP-S65T/H148D (soit la protéine de la GFP dont on substitue la sérine par la tryptophane)[40].

De plus, un des atomes d'oxygène de la fonction acide carboxylique de Asp148 est à une distance de 2.87 Å de l'atome d'oxygène du carbonyle de Asp148. Cette distance étant typique d'une liaison hydrogène et l'atome d'oxygène du carboxyle étant dépourvu de proton, Malo *et al* [41] en déduisent qu'il y a une liaison hydrogène intra-acide aminé entre ces deux atomes d'oxygène et que l'atome d'oxygène de la fonction acide carboxylique doit alors nécessairement porter un proton.

Dans la ECFP, le résidu 145 interagit avec le résidu 148 ce qui le pousse loin du chromophore permettant ainsi la torsion de l'indole par rapport au reste du chromophore, alors que dans la Cerulean le résidu 146 est en forte interaction avec le résidu 148 en le

maintenant proche du chromophore, avec une énergie fortement répulsive entre Asp148 et Trp 66 ce qui permet au chromophore d'adopter une conformation plane. Cette planéité corrige la fluorescence de la Cerulean par rapport à la ECFP, elle est 2,5 fois plus lumineuse que la ECFP [43], et simplifie le déclin en mono exponentiel. Ce qui explique l'influence de l'hétérogénéité des résidus 143-150 entre la ECFP et la Cerulean sur le chromophore et sa fluorescence.

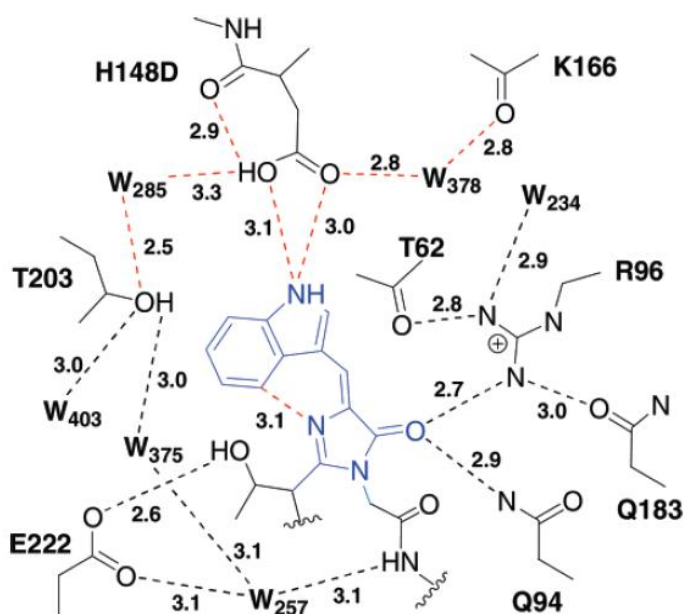


Figure 16 : Représentation schématique du chromophore de la Cerulean et de son environnement dans la structure cristallographique. Les tirets représentent les liaisons hydrogène identifiées à partir de la structure RX (les distances entre atomes lourds sont données en Å). Les lignes rouges correspondent aux liaisons hydrogène supposées jouer un rôle dans la stabilisation du chromophore en conformation cis [41].

A notre connaissance, aucune structure cristalline de la Cerulean à pH neutre ou basique n'a été publiée à ce jour.

Des études théoriques ont montré que la chaîne latérale de ASP148 se trouve dans le solvant et est stabilisée par une liaison hydrogène avec SER205 [37,39]

Ce changement peut permettre au chromophore de passer de la conformation cis à la conformation trans ce qui est énergiquement favorable vu l'espace créé par le déplacement de la chaîne latérale de ASP148 permettant ainsi au chromophore de changer de conformation.

## II.4 Réseau de liaisons H au voisinage du chromophore :

Ce réseau décrit la façon dont le chromophore est maintenu au sein de la protéine. Il est très important pour la compréhension de la relation entre la structure et les propriétés photophysiques des protéines, d'une part parce que la présence des acides aminés polaires proches du chromophore induit un champ électrique au voisinage du chromophore susceptible de modifier ses transitions électroniques, et d'autre part, parce qu'ils sont les chemins les plus probables des transferts de protons et d'électrons.

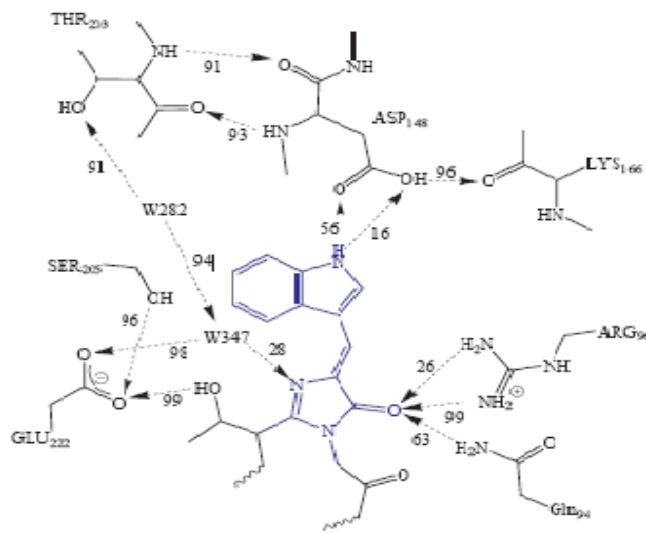


Figure 17 : Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à pH 5. Les flèches indiquent la présence d'une liaison hydrogène et sont orientées du donneur vers l'accepteur d'hydrogène. Les nombres associés à chaque liaison hydrogène indiquent en pourcentage le temps pendant lequel la liaison hydrogène existe. Les critères utilisés pour déterminer l'existence d'une liaison hydrogène sont de 3.2Å pour la distance et 40° pour l'angle [37].

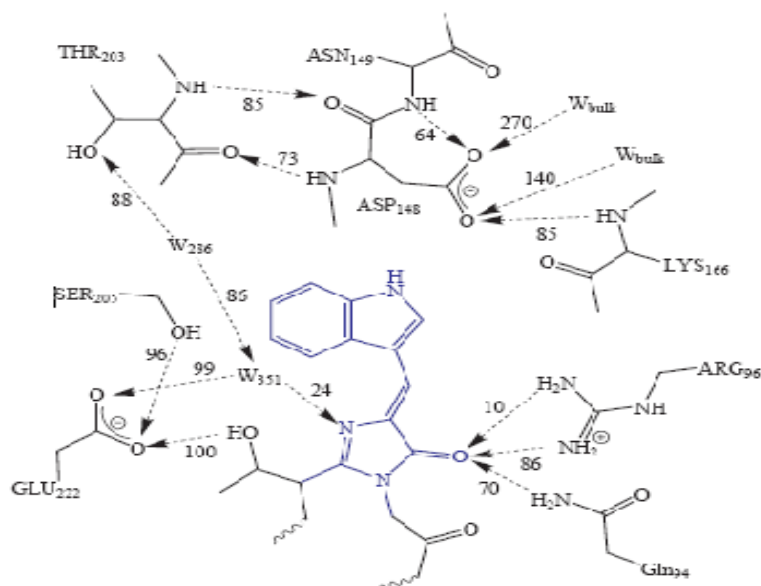


Figure 18 : Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à pH 8 avec le chromophore en conformation cis. Les flèches indiquent la présence d'une liaison hydrogène et sont orientées du donneur vers l'accepteur d'hydrogène. Les nombres associés à chaque liaison hydrogène indiquent, en pourcentage, le temps pendant lequel la liaison hydrogène existe. Les critères utilisés pour déterminer l'existence d'une liaison hydrogène sont de 3.2 Å pour la distance et 40° pour l'angle [37].

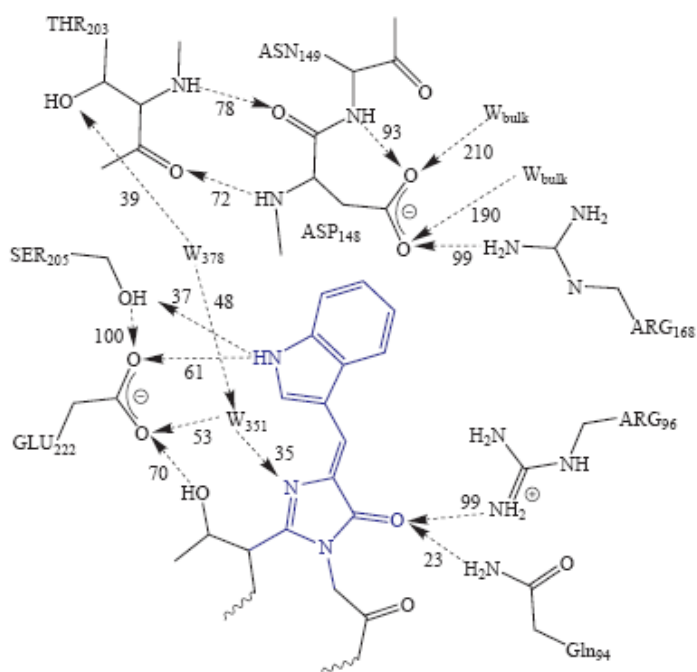
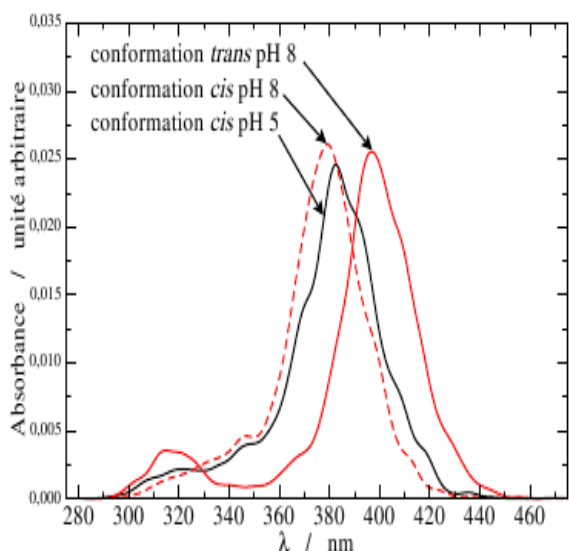


Figure 19 : Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à pH 8 avec le chromophore en conformation trans. Les flèches indiquent la présence d'une liaison hydrogène et sont orientées du donneur vers l'accepteur d'hydrogène. Les nombres associés à chaque liaison hydrogène indiquent, en pourcentage, le temps pendant lequel la liaison hydrogène existe. Les critères utilisés pour déterminer l'existence d'une liaison hydrogène sont de 3.2 Å pour la distance et 40° pour l'angle [37].

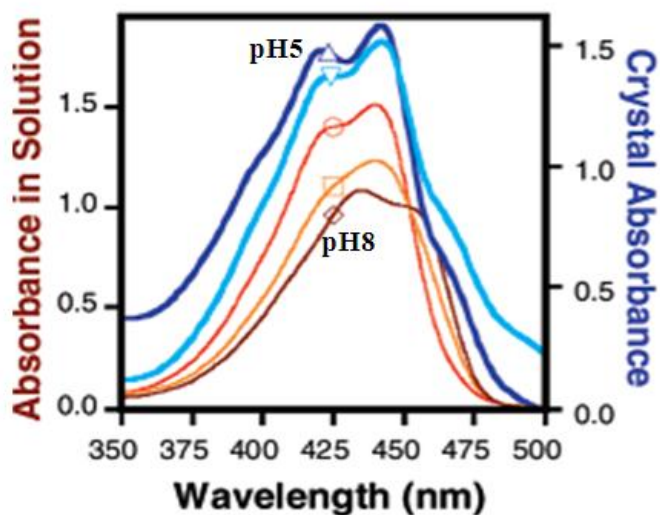
## II.5 Influence du pH sur le spectre d'absorption :

Une étude par cristallographie [39] a montré que le spectre d'absorption est stable de pH 8 à 10 et qu'il se déplace vers le bleu vers les courtes longueurs d'ondes pour un pH = 5.



La variation du spectre d'absorption de la Cerulean calculée théoriquement

[35]



La variation du spectre d'absorption de la Cerulean obtenue expérimentalement

[40]

Figure 20 : Variation du spectre d'absorption de la Cerulean en fonction du pH allant de 5 à 8

La Figure 20 montre la dépendance du spectre d'absorption au pH lors d'un saut de ce dernier de 8 à 5, le spectre se déplace du rouge vers le bleu.

Une observation qui a duré 150 minutes pendant laquelle les auteurs ont pris des titrages dans des intervalles de temps différents  $T_0$  correspond à pH 8 (rouge) et  $T_{150\text{min}}$  correspond à pH 5 (bleu).

Les courbes représentent l'absorption en fonction des longueurs d'ondes à différents pH compris entre 8 et 5. On remarque que chaque spectre contient deux maxima, à pH 8 ils sont à 435 nm et 454 nm (rouge) et à pH 5 soit un décalage de 11 nm et 15 nm respectivement, ils seront donc à 424 nm et 439 nm (bleu).

En prenant comme base la comparaison entre le spectre d'absorption expérimental de la Cerulean (obtenu par l'analyse aux rayons X à  $2.0\text{\AA}$  de la structure cristalline) et le spectre théorique (calculé par la méthode TD-DFT dans les trois situations pH5-cis, pH8-cis et

pH8-trans) et lors du déplacement du pH de 8 à 5 nous constatons que lorsqu'on garde la même structure du chromophore (cis) et modifie la distribution des protons dans la protéine la valeur de  $\lambda_{\max}$  se décale de 4 nm. Par contre, le décalage est de 15 nm lorsqu'on prend en considération l'isomérisation du chromophore ce qui est conforme aux résultats expérimentaux, de plus le maximum théorique à pH8-trans est plus proche à l'expérimental (405 nm et 434 nm) que celui du pH8-cis (380 nm et 434 nm), en effet ces deux remarques favorisent l'hypothèse que le chromophore change de conformation (trans-cis) lors de l'abaissement du pH.

Les spectres d'absorption expérimentaux de la Cerulean présentent un double maximum. Ce double maximum est généralement attribué à une structure vibrationnelle du spectre d'absorption qui n'est pas reproduite dans les calculs théoriques.

On remarque aussi un décalage important entre les maxima théoriques et expérimentaux, il est de 382 nm à pH5 théorique alors que ce maxima est de 424 nm, ce large décalage pourrait être dû à la méthode théorique utilisée en l'occurrence la TD-DFT.



# **Chapitre III**

## **Résultats et interprétation**

Dans ce chapitre, nous avons utilisé les résultats des simulations de dynamique moléculaire de la Cerulean effectuées par G. Vallverdu au cours de sa thèse au LCP [37]. Malheureusement, des 800 simulations qu'il a effectuées, nous n'avons pu récupérer qu'une vingtaine.

On rappellera simplement ici que les simulations de la dynamique moléculaire sont basées sur l'utilisation d'un champ de force qui contient l'expression analytique de l'énergie potentielle du système ainsi que tous les paramètres nécessaires à son utilisation. Dans le programme AMBER utilisé pour les simulations; le champ de force couramment employé est ff99 (voir annexe).

L'étude concernant la Cerulean met en jeu deux conformations du chromophore à l'état électrique fondamental ; nommées *cis* et *trans* (Voir figure 21)

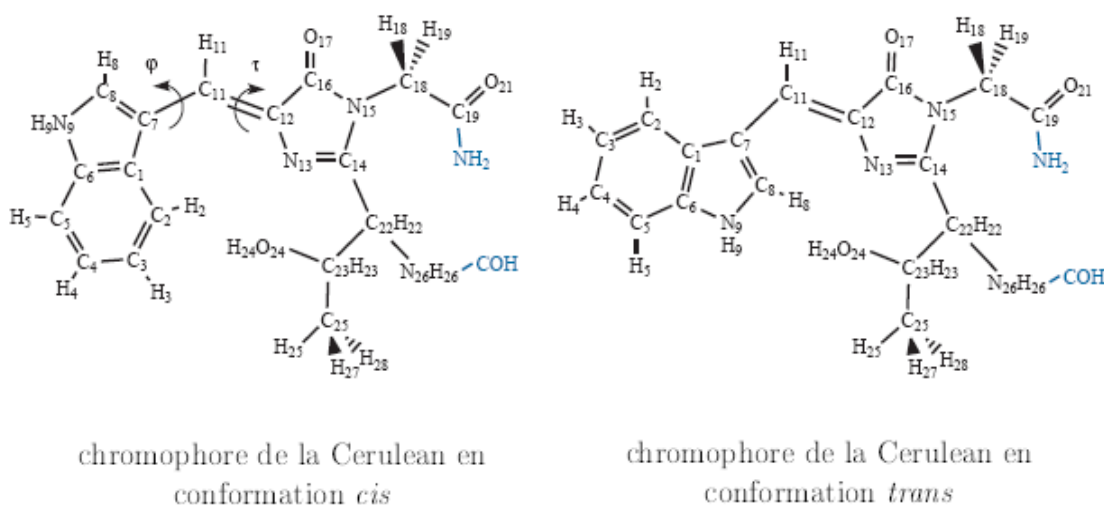


Figure 21 : représentation et numérotation des atomes du chromophore de la Cerulean en conformation *cis* et *trans* [37]

Un champ de force du chromophore à l'état fondamental pour chacune des deux conformations a été déterminé par G.Vallverdu [37] (voir annexe).

### **III.1 dynamique moléculaire de la Cerulean : Etat de protonation de la protéine :**

Les acides aminés dont la chaîne latérale porte une fonction acido-basique peuvent être sous leur forme acide (protonée) ou sous leur forme basique (non protonée) selon le pH. Mais dans les protéines, l'état de protonation individuel de chaque acide aminé dépend de celui des autres acides aminés acido-basiques.

L'ensemble des états de protonation individuels définit un micro état de protonation de la protéine ou une distribution de protons.

L'état de protonation de la protéine à un pH donné correspond alors à un ensemble de distributions de protons.

Le rôle du pH a été modélisé en utilisant le micro-état de protonation le plus probable à pH5 ou à pH8. Dans la cerulean, 77 acides aminés présentent une fonction acido-basique.

Les distributions des protons et leur probabilité ont été obtenues à l'aide du serveur de calcul  $H^{++}$  [44,45]. Le serveur résout l'équation de Poisson-boltzmann linéarisée et calcule pour l'ensemble des acides aminés acido-basiques les valeurs de  $PK_{1/2}$ .

A chaque micro-état de protonation est associée une énergie libre et pour un pH donné, plusieurs micro-états de protonation peuvent avoir des valeurs d'énergies libres proches et donc être équiprobables. La distribution des protons dans la protéine à un pH donné n'est donc généralement pas définie par un micro-état de protonation mais par la moyenne des micro-états de protonation thermiquement accessibles.

A partir des structures cristallisées et des micro-états de protonation choisis pour représenter l'état de protonation de la protéine à pH5 et pH8, des simulations de dynamique moléculaire ont été réalisées à pH5 en conformation cis et à pH8 en conformation cis et trans.

Dans ces simulations, la structure initiale a été complétée par des atomes d'hydrogène. Ceci est réalisé de façon systématique par le programme AMBER pour les atomes d'hydrogène non labiles en comparant la structure issue de la protéine Data Bank aux structures et géométries contenues dans les bibliothèques de chaque acide aminé.

### **III.2 Calcul du spectre d'absorption par simulation de dynamique moléculaire :**

Les simulations de dynamique moléculaire utilisent la mécanique classique, et pour calculer le spectre d'absorption de la cerulean, des calculs quantiques des longueurs d'onde d'absorption sont nécessaires. Ils nécessitent d'utiliser une molécule modèle dont les propriétés photophysiques soient aussi proches que possible de celles du chromophore tel qu'il est dans la protéine. Dans notre cas, il suffit juste de définir les liaisons qui déterminent les bords du chromophore.

Un spectre d'absorption donne la probabilité qu'une molécule absorbe un photon en fonction de sa longueur d'onde. Cette probabilité peut être calculée par la mécanique quantique, elle est proportionnelle au module du moment de transition au carré. En

supposant que la transition électronique est très rapide devant le mouvement des noyaux, le moment de transition se sépare en deux termes : un terme purement électronique, appelé moment de transition électronique, et le facteur de Franck-Condon qui correspond au recouvrement entre la fonction d'onde vibrationnelle de l'état électronique fondamental et les fonctions d'onde vibrationnelles de l'état électronique excité. Pour calculer un spectre d'absorption, on doit donc évaluer ces deux termes afin de calculer une probabilité d'absorption.

En simulation de dynamique moléculaire, on suppose que le mouvement des noyaux est classique. On ne dispose donc pas des fonctions d'onde vibrationnelles ni de l'état fondamental (dans lequel est faite la dynamique) ni de l'état excité et les facteurs de Franck-Condon ne sont donc pas calculables. On suppose alors que le spectre d'absorption est la somme des probabilités de transitions électroniques calculées sur l'ensemble des configurations du système, multipliée par la densité de probabilités de ces configurations. Pour chacune des configurations, on peut calculer les énergies de transition électronique vers les premiers états électroniques excités. On obtient ainsi les longueurs d'onde d'absorption pour chaque configuration du système. La probabilité que le système absorbe un photon à la longueur d'onde  $\lambda_c$  est alors donnée par la probabilité d'obtenir une configuration dans laquelle la longueur d'onde d'absorption est égale à  $\lambda_c$  multipliée par la probabilité d'absorption électronique. Cette dernière est en fait proportionnelle au module du moment de transition électronique au carré qui peut être évalué à partir de la force d'oscillateur de la transition.

Pour calculer le spectre d'absorption nous procédons donc de la façon suivante :

Par une dynamique moléculaire on obtient un échantillon de configurations représentatif de la région Franck-Condon.

Pour chacune de ces configurations et par la méthode TD-DFT avec la fonctionnelle (B3LYB), on calcule l'énergie de transition vers les premiers états excités du système et les forces d'oscillateur associées à ces transitions. Le spectre d'absorption est ensuite obtenu en construisant l'histogramme des longueurs d'ondes associées aux énergies de transitions pondérées par les forces d'oscillateurs.

$$S(\lambda) = \sum_{c=1}^{N_{conf}} \sum_{i=1}^{N_{etat}} f_{c,i} \delta(\lambda - \lambda_c) \quad (III.1)$$

Afin de lisser le spectre d'absorption obtenu, on peut utiliser des fonctions gaussiennes à la place de fonctions de Dirac. Le spectre est alors obtenu par la relation :

$$S(\lambda) = \sum_{c=1}^{N_{conf}} \sum_{i=1}^{N_{etat}} f_{c,i} \exp\left(-\frac{(\lambda-\lambda_{c,i})^2}{2\sigma^2}\right) \quad (III.2)$$

où  $\sigma$  est la largeur des gaussiennes, que nous avons prise égale à 3 nm. Cette largeur représente la largeur naturelle d'une raie d'absorption. Le nombre  $N_{conf}$  de configurations utilisées pour calculer le spectre d'absorption est égal à 20. Le nombre  $N_{etat}$  d'états électroniques excités à prendre en compte correspond au nombre de transitions électroniques dont les longueurs d'onde sont dans un même intervalle de valeurs. Dans notre cas, le calcul des quatre premiers états excités s'est révélé suffisant. Afin de prendre en compte l'effet de la protéine le calcul des énergies de transition du chromophore a été fait dans le champ des charges ponctuelles des atomes de la protéine (et de la boîte d'eau). Les valeurs des charges sont celles du champ de force AMBER et leurs positions sont celles de la configuration concernée.

Cette méthode de calcul du spectre d'absorption présente plusieurs limites. D'une part, du fait que l'on suppose que le mouvement des noyaux des atomes est classique, on ne peut pas calculer les facteurs de Franck-Condon et on n'aura donc pas accès à la structure vibrationnelle du spectre d'absorption. D'autre part, cette méthode est limitée par le choix d'un modèle restreint pour le chromophore dans les calculs quantiques.

### **III.3. spectre d'absorption de la Cerulean :**

Le décalage du spectre d'absorption sous l'effet d'un changement de pH, observé expérimentalement (figure 20), peut avoir deux origines [41] :

- La modification de l'environnement du chromophore dû à la variation du pH (changement de l'état de protonation et éventuellement de la conformation des acides aminés voisins)
- Et/ou l'isomérisation du chromophore

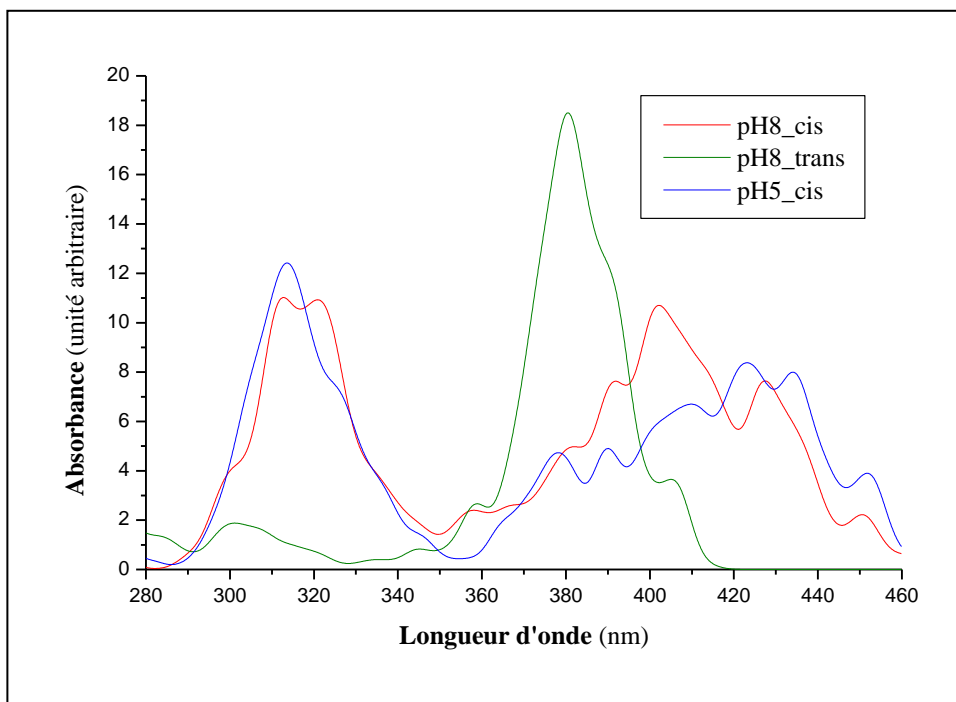


Figure 22 : spectre d'absorption de la Cerulean calculé dans les cas pH8 trans, pH8 cis et pH5 cis.

Nous avons calculé le spectre d'absorption de la Cerulean dans les trois situations, pH5 cis, pH8 cis et pH8 trans. Les spectres associés aux trois situations ont été calculés avec la méthode décrite dans le paragraphe précédent en utilisant une vingtaine de configurations dans chaque cas. Les trois spectres sont représentés dans la figure 22

Le principal résultat concerne la variation de  $\lambda_{\max}$  du maximum du spectre. Nous ne retrouvons pas le spectre obtenu par G.Vallverdu dans sa thèse [37], de plus notre spectre n'est pas lissé, ceci est dû principalement au nombre des configurations  $N_{\text{conf}}$  insuffisant, en effet pour réduire l'incertitude sur le spectre d'absorption calculé, il est nécessaire d'augmenter la taille de l'échantillon de configurations utilisées. Bien sur cela augmente le temps de calcul mais on considère qu'il faut au moins 860 configurations pour donner un résultat acceptable.

En accord avec les résultats expérimentaux et théoriques, les spectres d'absorption correspondant à la conformation cis à pH5 et pH8 sont décalés vers le bleu alors que celui qui correspond à la conformation trans à pH8 est décalé vers le rouge.

Cependant le décalage de nos résultats avec les résultats de G.Vallverdu est très grand.

Il faut noter que le décalage observé entre les valeurs expérimentales et théorique a été constaté à plusieurs reprises dans plusieurs études employant la méthode TD-DFT (ce qui est notre cas) pour calculer le spectre d'absorption [46,47].

On remarque aussi que le décalage de  $\lambda_{\max}$  est très faible lorsqu'on abaisse le pH de 8 à 5 pour le chromophore en conformation cis, alors qu'il devient important pour un même pH lorsqu'il y a isomérisation du chromophore.

L'effet du changement du champ électrostatique sur le spectre d'absorption du chromophore est faible, lorsque l'isomérisation du chromophore induit un décalage du spectre plus grand.

Nos résultats confortent l'hypothèse de l'isomérisation du chromophore comme origine de la dépendance du spectre d'absorption au pH.

Le rôle limité des effets électrostatiques associés au changement de pH a été constaté en comparant les deux micro-états de protonation les plus probables à pH5 et à pH8.

## Conclusion générale

Dans ce mémoire, après avoir présenté les protéines d'une manière générale, nous nous sommes intéressés aux protéines fluorescentes et plus spécifiquement à celle de la famille de la Green Fluorescent Protein GFP. La protéine qui fait l'objet de cette étude est un mutant de la GFP : la Cerulean. Son utilisation en imagerie médicale n'est plus à prouver et il est donc important de caractériser certains aspects de ses propriétés photophysiques.

Pour cela nous avons présenté de façon succincte, les principes et les grandeurs associés à l'étude des processus photophysiques en phase condensée.

De la même manière, nous avons décrit les méthodes de simulation utilisées pour caractériser certains aspects des propriétés photophysiques de la Cerulean.

Nous avons calculé le spectre d'absorption de la Cerulean à pH5 en conformation cis et à pH8 en conformation cis et trans. Nous avons démontré que l'isomérisation du chromophore induit un décalage du spectre très semblable à celui observé expérimentalement lors du passage de pH8 à pH5.

Nous avons montré aussi le rôle limité des effets électrostatiques associés au changement de pH en comparant les deux micro-états de protonation les plus probables à pH5 et pH8.



## Référence :

- [1] Acides aminés et protéines, Mohammed EL ATYQY, Ingénieur I.A.A. 2007.
- [2] Structure et fonction des protéines - Page 3 Gregory A. Petsko, Dagmar Ringe, Chrystelle Sanlaville 2008.
- [3] Matière et énergie dans les systèmes: manuel de chimie-biochimie Pierre Goudet, Philippe-José Yindoula 2008.
- [4] Chimie organique; Jonathan Clayden, Stuart Warren, Nick Greeves 2002.
- [5] Les protéines ; Gilles Bourbonnais, Cécep de Sainte-Foy 2001.
- [6] Biochimie; Reginald H. Garrett, Charles M. Grisham, B. Lubochinsky 2000.
- [7] Christophe Nicolas Magnan, Cécile Capponi, François Denis "A Protocol to Detect Local Affinities Involved in Proteins Distant Interactions", pp. 252-257, 2007.
- [8] structure et fonction des protéines ; Gregory A. Petsko, Dagmar Ringe, Chrystelle Sanlaville ; 2008.
- [9] Biochimie structurale et métabolique ; Christian Moussard 2006.
- [10] chimie pharmaceutique ; Graham L. Patrick ; 2002
- [11] Structures des protéines: pathologie et approches thérapeutiques ; Max Perutz ; 1996.
- [12] Biochimie structurale ; Claude Audigié, François Zonszain 1991.
- [13] Banki et al. Nat Methods 2: 659-62 et Ge et al. (2005) J Am Chem Soc 127: 11228-9. 2005
- [14] A. Manz, N. Pamme & D. Iossifidis, Bioanalytical Chemistry, 2004.
- [15] les analogues de l'insuline ; Bernard Vialettes, Denis Raccah, D Raccah ; 2006
- [16] biochimie ; Isabelle Claverie, Mireille Panet, Sylvie Barbeau ; 2008
- [17] toute la biochimie ; Serge Weinman, Pierre Méhul ; 2004.
- [18] Éléments de biologie cellulaire - Page 84 Daniel Robert, Brigitte Vian 2004.
- [19] Cody, W.C., Prasher, D.C. et al. Biochemistry 32 1212-1218, 1993.
- [20] Heim, R. et al., Proc. Natl. Acad. Sci. USA 91 12501-12504, 1994.
- [21] Green Fluorescent Protein A molecular tag that can be inserted into genes to make animals and plants glow green. Par Timothy King and Paul May University of Bristol. 2010.
- [22] Observatoire Océanologique de Villefranche sur Mer ; David Luquet 2005
- [23] Green Fluorescent Protein as a Marker for Gene Expression ; Martin Chalfie, Yuan Tu, Ghia Euskirchen, William W. Ward, Douglas C. Prasher ; 1994
- [24] <http://ead.univangers.fr/~jaspard/Page2/TexteTD/5TDBioCell1/2TDMethodEtudCell/3Figures/3Fluorescence/1Fluorescence.htm>.
- [25] fluorescent protéines ; Robert E. Campbell ; 2008
- [26] Les Protéines Fluorescentes ; Tounsia Aït-Slimane, UMRS 938, CDR Saint-Antoine CHU Saint-Antoine
- [27] [http://www.conncoll.edu/ccacad/zimmer/GFP-ww/Natural\\_Function.html](http://www.conncoll.edu/ccacad/zimmer/GFP-ww/Natural_Function.html)
- [28] Day RN, Davidson MW. The fluorescent protein palette: tools for cellular imaging. Chem Soc Rev. 38, 2887-921 2009.
- [29] <http://jcs.biologists.org/content/vol120/issue24/images/large/JCS005801F1.jpeg>
- [30] biochimie des protéines, Benoit Leblanc, université Sherbrooke, 2004
- [31] M.E. Martin, F. Negri, and M. Olivucci. Origin Nature, and Fate of the Fluorescent State of the Green Fluorescent Protein Chromophore at the CASPT2//CASSCF Resolution. J. Am. Chem. Soc., 126 :5452#5464, 2004.
- [32] A. Toniolo, S. Olsen, L. Manohar, and T.J. Martinez. Conical intersection dynamics in solution : the chromophore of green fluorescent protein. Faraday Discuss., 127 :149#163, 2004.

- [33] P. Altoe, F. Bernardi, M. Garavelli, G. Orlandi, and F. Negri. Solvent effects on the vibrational activity and photodynamics of the green fluorescent protein chromophore : A quantum-chemical study. *J. Am. Chem. Soc.*, 127(11) :3952-3963, 2005.
- [34] S. Olsen and S.C. Smith. Bond selection in the photoisomerization reaction of anionic green fluorescent protein and kindling fluorescent protein chromophore models. *J. Am. Chem. Soc.*, 130(27) :8677-8689, 2008.
- [35]- Roger Y. Tsien *Annual Reviews Biochem. Rev Biochem.* 1998.
- [36]- Roger Y. Tsien *THE GREEN FLUORESCENT PROTEIN Annual Review of Biochemistry Vol. 67: 509-544* (Volume publication date July 1998).
- [37]- Thèse de l'étude Théorique de Processus Photophysiques dans des Protéines fluorescentes par Germain VALLVERDU page 50, Juillet 2009.
- [39]- Mickael Lelimosin, Marjolaine Noirclerc-Savoie, "Intrinsic Dynamics in ECFP and Cerulean Control Fluorescence Quantum Yield "Biochemistry, 2009, 48 (42), pp 10038-10046.
- [40]- Rizzo et al. (2004) *Nat. Biotechnol.* 22, 445
- [41]-Gabrielle D. Malo, Lauren J. Pouwels, Meitian Wang, X-ray Structure of Cerulean GFP : A Tryptophan-Based Chromophore Useful for Fluorescence Lifetime Imaging. *Biochemistry*, Vol. 46, N° 35, 2007.
- [42]- Leiderman, P.; Genosar, L.; Huppert, D.; Shu, X.; Remington, S. J.; Solntsev, K. M.; Tolbert, L. M. "Ultrafast Excited-State Dynamics in the Green Fluorescent Protein Variant S65T/H148D. 3. Short- and Long-Time Dynamics of the Excited-State Proton Transfer." *Biochemistry* 2007, 46, 12026-12036.
- [43]- Dr. David Piston Department of Molecular Physiology and Biophysics Variants of Cyan Fluorescent Protein with Improved Fluorescent Properties
- [44]- Natacha Gillet Variation des états de protonation dans deux protéines fluorescentes en fonction du pH, 2008.
- [45]- Gordon JC, Myers JB, Folta T, Shoja V, Heath LS and Onufriev A., "H<sup>++</sup>: a server for estimating pK<sub>a</sub>s and adding missing hydrogens to macromolecules", *Nucleic Acids Res.* Jul 1;33:W368-71. (2005).
- [46]- D. Sinnecker, P. Voigt, N. Hellwig, and M. Schaefer. Reversible photobleaching of enhanced green fluorescent proteins. *Biochemistry*, 44(18) :7085-7094, 2005
- [47]- K.R.F. Somers, P. Kruger, S. Bucikiewicz, M. De Maeyer, Y. Engelborghs, and A. Ceulemans. Protein simulations : The absorption spectrum of barnase point mutants. *Protein Science*, 13(7) :1823-1831, 2004.
- [48]- W.D. Cornell, P. Cieplak, C.I. Bayly, L.R. Gould, K.M. Merz, D.M. Fergusson, D.C. Spell-meyer, T. Fox, J.W. Caldwell, and P.A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19) :5179- 5197, 1995.
- [49]- J. Wang, P. Cieplak, and P.A. Kollman. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.*, 21 :1049-1074. 2000.
- [50]-M.P. Allen and D.J. Tiddesley. *Computer simulation of liquids*. Oxford University Press, 1987.
- [51]-J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of cartesian equations of motion of a system with constraints-molecular-dynamics of n-alkanes. *J. Comput. Phys.*, 23(3) :327-341, 1977.
- [52]-P.P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3) :253-287, 1921.

## Annexe A

### Simulation moléculaire

Les simulations moléculaires regroupent un ensemble de méthodes numériques permettant de décrire le comportement physique ou chimique d'un ensemble de  $N$  particules. Nous considérerons le cas des simulations moléculaires classiques dans lesquelles les noyaux des atomes sont traités par la mécanique classique et le déplacement des électrons n'est pas pris en compte.

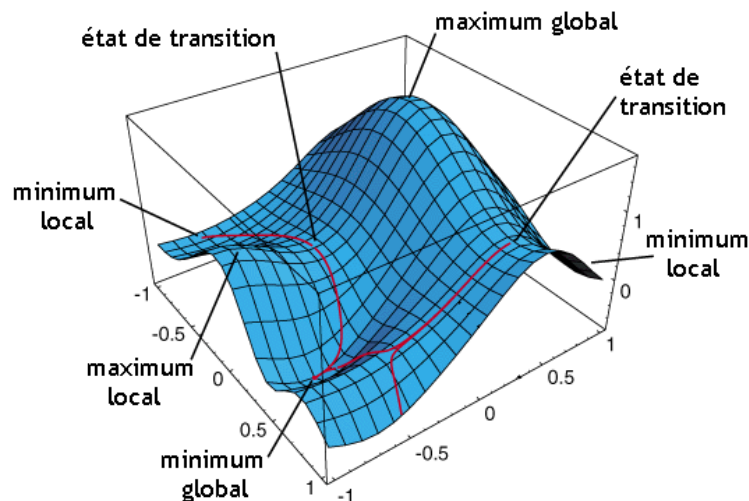
#### **A.1 Modélisation de la surface d'énergie potentielle**

En dynamique moléculaire classique les résultats (propriétés dynamiques aussi bien que structurales) sont déterminés par les forces qui s'exercent sur chaque atome. Cette partie présente la façon dont les interactions intra-moléculaires et inter-moléculaires sont modélisées pour des systèmes biologiques.

##### **A.1.1 Surface d'énergie potentielle**

L'énergie potentielle d'un système de  $N$  atomes en interaction est une fonction des coordonnées  $q_i$  de chaque atome. Les valeurs de cette fonction définissent donc une hypersurface de dimension  $3N$  si le système n'est pas isolé, elle est notée  $U(qm)$  et appelée surface d'énergie potentielle. L'image que l'on se donne le plus souvent d'une surface d'énergie potentielle est celle d'un relief montagneux (figure A.1). Elle comporte en particulier un ensemble de minima où les dérivées premières sont nulles et les dérivées secondes positives. Ces minima représentent des géométries stationnaires où le système peut rester un certain temps. Ils sont séparés par des barrières appelées "états de transition" et dont la hauteur est appelée "énergie d'activation" (voir le schéma en dimension 2 sur la figure A.1).

Dans un système complexe où le nombre d'atomes et donc de coordonnées est grand, comme dans les systèmes biologiques, le nombre de points stationnaires est grand et la surface d'énergie potentielle est complexe. La recherche du minimum global, des points stationnaires et des états de transition est alors très difficile à cause du nombre de coordonnées à faire varier.



**Figure A.1.** Schéma d'une surface d'énergie potentielle simple en deux dimensions. Sur la figure sont indiqués les points spécifiques de la surface d'énergie potentielle. Ils correspondent aux points pour lesquels les dérivées de la surface d'énergie potentielle en fonction des deux coordonnées sont nulles.

### **A.1.2 Définition d'un champ de force**

Un champ de force définit l'expression analytique de la surface d'énergie potentielle et l'ensemble des paramètres nécessaires pour la calculer. Il existe différents types de champs de force, chacun étant adapté à un type de système et à la précision avec laquelle il est décrit. Plus le nombre d'atomes représentés par une seule particule est faible, plus la description du système est précise. La précision la plus fine correspond au modèle dit "tout atome" (traduction de l'expression anglaise *all atom force field*) dans lequel tous les atomes sont représentés, chacun étant considéré comme une particule ponctuelle chargée, de masse correspondant à la masse atomique de l'atome. Le champ de force de ce type de représentation est décrit dans ce qui suit.

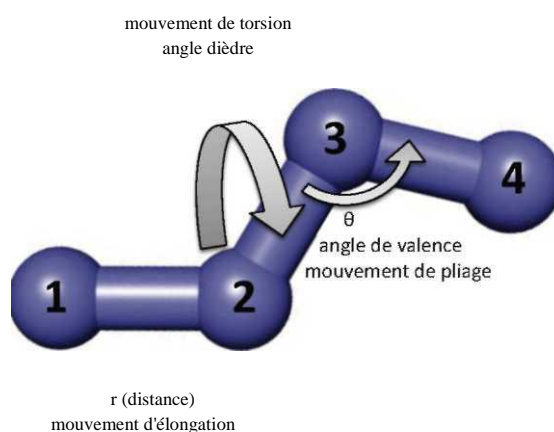
Les paramètres des champs de force sont généralement associés à la notion de type d'atome. Un type d'atome dépend de l'élément chimique de l'atome, de son hybridation (c'est à dire le nombre de liaisons qu'il engage) et de la fonction chimique dans laquelle il se trouve. A chaque type d'atome est associé un jeu de paramètres et l'ensemble des paramètres nécessaires pour calculer l'énergie potentielle d'un système est alors défini à partir de sa topologie.

### A.1.3 Modèle "tout atome" et coordonnées internes

Dans un modèle tout atome, une configuration du système est donnée par la position dans l'espace de l'ensemble des atomes. Il existe deux types principaux de coordonnées : les coordonnées cartésiennes et les coordonnées internes.

Dans un système de coordonnées cartésiennes, la position des atomes est repérée par trois coordonnées ( $x$ ,  $y$ ,  $z$ ) calculées par rapport à un repère fixe.

Dans un système de coordonnées internes, un atome est repéré par rapport aux autres atomes du système. On utilise alors une distance, un angle de valence et un angle dièdre (figure A.2).



**Figure A.2.** Définition des coordonnées internes d'une chaîne d'atomes. L'élongation est décrite par la distance  $r$  entre les atomes 1 et 2. Le pliage est décrit par l'angle de valence  $\theta$  entre les atomes 2, 3 et 4. La torsion est mesurée par un angle dièdre  $\phi$ , défini comme l'angle entre les plans passant par les atomes 1, 2 et 3 d'une part et 2, 3 et 4 d'autre part.

Les coordonnées internes sont très pratiques pour étudier les déformations d'une molécule ou imposer des contraintes géométriques. En effet, celles-ci peuvent être représentées par des fonctions simples des coordonnées internes alors qu'elles sont plus difficiles à modéliser avec les coordonnées cartésiennes car un plus grand nombre de coordonnées varient simultanément.

Cependant, dans les programmes de dynamique moléculaire, les algorithmes utilisés pour propager les trajectoires utilisent le plus souvent les coordonnées cartésiennes car elles sont plus pratiques pour le calcul de l'énergie cinétique qui fait intervenir des dérivées secondes.

### **A.1.4 Champ de force ff99**

Dans ce travail, le champ de force que nous avons utilisé pour les simulations de dynamique moléculaire est un champ de force tout atome intégré dans la suite de programme AMBER et nommé ff99 [48,49]. L'expression analytique de l'énergie potentielle,  $W/799$ , associée à ce champ de force est la suivante :

$$U_{ff99} = \sum_{\text{liaisons}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dièdres}} \sum_n \frac{V_n}{k_n} [1 + \cos(n\phi - \phi_0)] + \sum_{i < j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j} \left( \frac{q_i q_j}{\epsilon r_{ij}} \right) \quad (\text{A.1})$$

Les différents termes de l'énergie potentielle représentent les différentes interactions modélisées par ce champ de force. On remarquera, entre autre, qu'il suppose que les coordonnées internes ne sont pas couplées.

### **A.2 Simulation de dynamique moléculaire de protéines**

Dans cette partie, le principe d'une simulation de dynamique moléculaire sera présenté. L'utilisation de cette méthode pour étudier des systèmes de la taille d'une protéine contenant plusieurs milliers d'atomes dans des conditions de température et de pression similaires à celles des expériences nécessite d'utiliser des techniques spécifiques. Les plus importantes d'entre elles seront brièvement décrites.

#### **A.2.1 Principe d'une dynamique moléculaire**

La dynamique moléculaire consiste à intégrer les équations du mouvement de la mécanique classique afin de déterminer la trajectoire des atomes du système. On peut alors suivre l'évolution dans le temps du système. L'équation fondamentale de la dynamique provient de la troisième loi de Newton

$$m_i \times \frac{d^2 \vec{q}_i}{dt^2} = \vec{F}_i$$

où  $\vec{q}_i$  représente les coordonnées de l'atome  $i$  de masse  $m_i$  et  $\vec{F}_i$  est la somme des forces qui s'appliquent sur l'atome  $i$ . Cette somme des forces est calculée à partir de la surface d'énergie potentielle,  $U(q^{3N})$ , et donc de l'expression analytique du champ de force selon :

$$\vec{F}_i = -\overrightarrow{\text{grad}}_i U(q^{3N})$$

Les trajectoires sont obtenues par intégration numérique des équations du mouvement par différence finie. Etant donné un pas de temps  $\delta t$  et les positions et les vitesses de l'atome  $i$  au temps  $t$  on détermine les positions et les vitesses au temps  $t + \delta t$  et la trajectoire est ensuite obtenue en itérant ce processus.

Il existe un grand nombre d'algorithmes visant à satisfaire au mieux le compromis entre une précision maximale (utilisation de  $\delta t$  petit) et un temps de calcul limité (utilisation de grandes valeurs de  $\delta t$ ). Dans la suite de programme AMBER, avec lequel les simulations de dynamique moléculaire ont été faites, l'algorithme utilisé est celui dit "Leap-Frog" [50]. Comme tous les autres, il est itératif et nécessite donc une initialisation. Les coordonnées initiales sont issues de la structure RX de la protéine. Les vitesses initiales de chaque atome sont attribuées aléatoirement suivant une distribution des vitesses de Maxwell-Boltzmann.

Le choix du pas d'intégration  $\delta t$  est un paramètre important d'une simulation de dynamique moléculaire. On considère le plus souvent que  $\delta t$  doit être plus faible d'au moins un ordre de grandeur que le temps caractéristique du mouvement le plus rapide du système étudié. Dans un système biologique, le mouvement le plus rapide correspond aux élongations des liaisons X—H dont le temps caractéristique est de l'ordre de 10 fs. Le pas de temps doit alors être au maximum de 1 fs. Afin de pouvoir augmenter ce pas de temps, on suppose qu'on peut considérer que la longueur des liaisons X—H est fixe. Dans la suite de programmes AMBER, cette contrainte est imposée à l'aide de la procédure SHAKE [51] qui permet de contraindre toutes les longueurs de liaisons faisant intervenir un atome d'hydrogène. Dans ce travail, nous avons utilisé cet algorithme avec un pas de temps de 2 fs.

Dans tous les algorithmes, l'étape qui nécessite le plus de temps est le calcul des forces s'exerçant sur chaque atome et plus particulièrement les forces résultant des interactions entre atomes non liés. En effet, le calcul de ces interactions nécessite de calculer la distance entre toutes les paires d'atomes du système. Le temps nécessaire à ce calcul varie comme  $N^2$ , où  $N$  est le nombre d'atome. Pour limiter le nombre d'interactions à calculer et donc réduire le temps de calcul de chaque itération on définit un rayon de coupure,  $r_c$ , au delà duquel on néglige les interactions entre atomes non liés. Les valeurs de  $r_c$  couramment utilisés sont de l'ordre de 8 Å à 10 Å. Cette approximation est acceptable pour les interactions de van der Waals qui diminuent rapidement avec la distance (en  $r^{-6}$ ). Pour les interactions électrostatiques qui varient en  $r^{-1}$  l'interaction entre deux atomes n'est

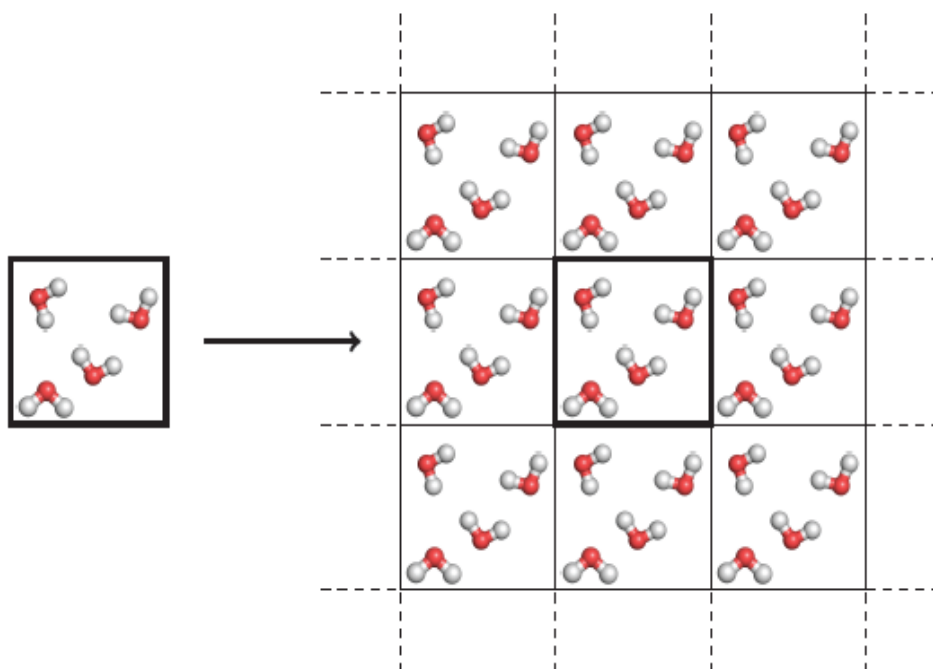
pas nulle même à grande distance et l'utilisation d'un rayon de coupure n'est pas satisfaisante. Une autre approche sera présentée au paragraphe A.2.3.

### **A.2.2 Conditions aux limites périodiques**

Dans le milieu cellulaire, les protéines sont en phase aqueuse. Pour décrire la présence des molécules d'eau on peut utiliser une représentation implicite ou explicite. Dans nos simulations, nous avons toujours utilisé une représentation explicite du solvant. Cette représentation modélise de façon plus réaliste les interactions entre la surface de la protéine et le solvant qui sont quelques fois cruciales pour obtenir une bonne description de la structure d'une protéine.

Le système dont on calcule la trajectoire contient alors la protéine, entouré de molécules d'eau enfermées dans un volume  $V$ . Nous avons utilisé un volume  $V$  de forme parallélépipédique, il est appelé "boîte de simulation". Cependant, cette boîte ne pouvant être de dimension infinie, une partie des molécules d'eau crée une interface avec le vide qui entoure le volume  $V$ . Si ce volume est trop faible, les effets de bords dus à la stabilisation des molécules à la surface de la boîte deviennent importants et perturbent fortement la dynamique du système. Pour contourner ce problème, on utilise des conditions aux limites périodiques qui sont une caractéristique importante des simulations en phase condensée. La boîte de simulation est répliquée à l'infini dans toutes les directions de l'espace et ses images périodiques entourent la boîte de simulation primitive (figure A.3).





**Figure A.3.** Conditions périodiques appliquées à un système en deux dimensions. La cellule correspondant à la boîte de simulation est encadrée, les autres cellules sont des images de la boîte de simulation primitive obtenues par translation. La flèche représente le déplacement d'une molécule qui sort de la boîte de simulation.

Au cours du mouvement, les coordonnées des atomes de la boîte de simulation primitive sont copiées dans toutes ses images. De plus, si une particule sort de la boîte de simulation, les conditions aux limites consistent à la faire entrer par la face opposée de la boîte (figure B.3). De cette manière le nombre d'atome dans chaque boîte est constant tout au long de la simulation. Les forces sont calculées pour tous les atomes de la cellule centrale en incluant les interactions avec les atomes des cellules voisines. Ces conditions périodiques donnent au système un caractère périodique artificiel. Il faut donc s'assurer que le soluté (la protéine dans notre cas) ne soit pas en interaction avec l'une de ses images. Pour ce faire, on utilise une boîte de simulation dont la plus petite dimension est au minimum égale à deux fois le rayon de coupure utilisé pour limiter le calcul des interactions entre atome non liés.

### **A.2.3 Calcul de l'énergie électrostatique par la méthode d'Ewald**

L'interaction électrostatique entre deux charges effectives est représentée par un terme de Coulomb (équation A.1). En phase condensée, le système comporte un ensemble de charges positives ou négatives réparties dans l'espace. Pour calculer l'énergie d'interaction entre les charges, on fait la somme des interactions de paires de charges.

$$E_{elec} = \frac{1}{4\pi\epsilon_0} \sum_i^N \sum_{i<j}^N \frac{q_i q_j}{r_{ij}} \quad (\text{A.2})$$

où N est le nombre d'atomes du système,  $q_i$  est la charge de l'atome i et  $r_{ij}$  est la distance entre les atomes i et j. En pratique, un système a toujours une taille finie et l'énergie électrostatique est donc un nombre fini. Dans nos simulations, le fait que le système est en phase condensée est modélisé par l'utilisation des conditions périodiques. Or, comme on calcule les interactions entre les atomes de toutes les images de la boîte de simulation, le nombre d'interaction à calculer serait infini. Pour contourner ce problème nous avons vu qu'on peut utiliser un rayon de coupure qui permet de négliger les interactions à longue distance qui ne contribuent que faiblement à l'énergie totale. Cette approche ne peut pas être utilisée dans le cas des interactions électrostatique car ce sont des interactions à long portée (elles décroissent en  $1/r_{ij}$ ) et la somme de l'équation A.2 a une convergence qui dépend de la façon dont on fait tendre le système vers l'infini.

La méthode d'Ewald [52,50] consiste à utiliser un artifice mathématique permettant de calculer l'énergie électrostatique d'un système infini en utilisant les conditions périodiques. Pour cela, on réexprime la densité de charge de la boîte de simulation (qui correspond à N charges ponctuelles) comme la somme de deux densités de charge (équation A.3).

$$\rho(R) = \left( \sum_{i=1}^N q_i \delta(R - R_i) - g_{q_i}(R) \right) + \sum_{i=1}^N g_{q_i}(R) \quad (\text{A.3})$$

où R représente une position dans l'espace, représente la position de la charge  $q_i$  et les fonctions  $g_{q_i}(R)$  représentent des densités de charges gaussiennes centrées en et dont la charge totale vaut  $q_i$ . Dans le premier terme de l'équation A.3, chaque charge ponctuelle de la protéine est écrantée par des densités de charges gaussiennes de qui sont ensuite réintroduites pour conserver la densité de charge réelle.

L'utilisation de la densité de charge de l'équation A.3 pour calculer l'énergie d'interaction électrostatique est un avantage car il remplace la somme de l'équation A.2 par la somme de deux termes plus faciles à calculer : le premier parce qu'il s'agit d'une somme qui converge rapidement et qui sera calculée en utilisant le même rayon de coupure que celui utilisé pour calculer les interactions de van der Waals ; le second parce qu'il est

possible de profiter de la périodicité du système pour l'évaluer dans l'espace réciproque en utilisant une transformée de Fourier.

Dans la suite de programme AMBER, les simulations utilisant les conditions périodiques utilisent la méthode d'Ewald pour calculer les interactions électrostatiques. Elle est implémentée en utilisant l'algorithme Particle Mesh Ewald [52,53]. Celui-ci est identique à la méthode d'Ewald pour le calcul du premier terme de l'équation A.3. Le second terme, dans l'espace réciproque, est calculé en approximant la densité de charges de la boîte de simulation par des charges réparties sur une grille fixe qui quadrille le volume de la boîte de simulation. Cette méthode permet une évaluation plus rapide de l'énergie électrostatique. Les programmeurs de AMBER ont particulièrement travaillé à l'optimisation de cette méthode, ce qui a abouti au module PMEMD (Particle Mesh Ewald Molecular Dynamics) qui est une version parallèle plus efficace du programme de dynamique moléculaire de AMBER.

#### **A.2.4 Ensembles thermodynamiques**

L'ensemble des forces mises en jeu dans le champ de force ne dépend que des positions des atomes. Donc l'énergie totale est conservée tout au long d'une simulation de dynamique moléculaire si on considère que le système est isolé. D'autre, part, l'utilisation des conditions périodiques aux limites permet de maintenir constant le nombre de particules  $N$  par unité de volume. Une dynamique moléculaire standard échantillonne donc des configurations dans l'ensemble micro-canonique, à énergie, volume et nombre de particules constant ( $N, V, E$ ). En pratique, des fluctuations ou une faible dérive de l'énergie sont observées à cause de l'accumulation d'imprécisions numériques ou encore à cause des approximations utilisées pour calculer les interactions entre atomes non liés.

Il est également possible de mettre en œuvre des algorithmes permettant de faire des simulations de dynamique moléculaire dans l'ensemble canonique ( $N, V, T$ ), à température volume et nombre de particule constant ou dans l'ensemble isobare-isotherme ( $N, P, T$ ) qui est celui que nous avons le plus utilisé au cours de cette thèse. Les algorithmes permettant de contrôler la température sont appelé thermostats et ceux qui permettent de contrôler la pression sont appelés barostats. Ils sont basés sur le couplage entre la dynamique des coordonnées du système et un réservoir de température ou un réservoir de pression.

## **Thermostats**

La température  $T$  du système est calculée à partir des vitesses  $V_i$  des atomes en coordonnées cartésiennes. Dans la limite classique où la densité d'état est élevée, le théorème de l'équipartition de l'énergie nous donne :

$$T = \frac{1}{3Nk_B} \sum_{i=1}^N m_i v_i^2 \quad (\text{A.4})$$

où  $N$  est le nombre de particules et  $m_i$  la masse de l'atome  $i$ .

Dans AMBER, plusieurs algorithmes de contrôle de la température sont disponibles [54,55,56,57]. Ils consistent à prendre en compte les échanges d'énergie cinétique entre le système et un thermostat à une température  $T_0$ .

L'algorithme de Berendsen [54] est le plus simple. Il impose que les fluctuations de la température  $T$  du système suivent la relation suivante :

$$\frac{dT}{dt} = \frac{T_0 - T(t)}{\tau_T}$$

où  $\tau_T$  représente le temps de relaxation de la température aussi appelé taux de couplage et  $T_0$  est la température d'un thermostat. Cette expression correspond à une contrainte harmonique qui impose à la température du système  $T(t)$  de fluctuer autour de la température  $T_q$  d'un thermostat. Les contraintes imposées sur la température du système induisent des contraintes sur les vitesses des atomes par l'intermédiaire de l'équation A.4. A chaque pas de temps on ajuste les vitesses  $V_i$  des atomes en les multipliant par un facteur correctif  $\lambda$  unique pour l'ensemble des atomes. Ce facteur est calculé en posant que la variation d'énergie cinétique est proportionnelle à la variation de la température du système. Il est donné par la relation :

$$\lambda = \sqrt{1 + \frac{\delta t}{p} \left( \frac{T_0}{T} - 1 \right)}$$

où  $p$  est l'unique paramètre du modèle qui dépend de  $\tau_T$ .

Nous avons utilisé ici un algorithme plus simple qui consiste à remplacer l'équation de Newton utilisée pour propager la trajectoire du système par une équation de Langevin. Dans cette dernière, une force de frottement et une force aléatoire sont ajoutées aux forces

déjà existantes et permettent de maintenir le système en équilibre avec un thermostat. Les paramètres déterminant ces deux forces sont directement codés dans la suite AMBER [57].

### **Barostat**

La pression qui s'exerce sur le système est calculée par la méthode du viriel. Dans le cas d'un système de  $N$  particules, si l'on ne considère que des interactions de paire additives, la pression est donnée par [58] :

$$P = \frac{2}{3V} \left( \sum_{i=1}^N \frac{1}{2} m_i v_i^2 - \sum_{i=1}^N \overline{q_i \cdot \overline{F_i}} \right)$$

où  $q_i$  est le vecteur position de l'atome  $i$  et  $\overline{F_i}$  représente la somme vectorielle des forces qui s'exerce sur l'atome  $i$ .

Dans AMBER l'algorithme de contrôle de la pression est un algorithme de Berendsen, qui est traité de manière similaire à celui de contrôle de la température. On contraint la pression du système à suivre la relation suivante

$$\frac{dP}{dt} = \frac{P_0 - P(t)}{\tau_p}$$

où  $P_0$  est la pression d'un barostat en équilibre avec le système et  $\tau_p$  est le temps caractéristique de relaxation de la pression. Dans ce cas la contrainte imposée sur la pression impose une contrainte sur le volume de la boîte de simulation. A chaque pas de temps il est multiplié par un facteur  $p$ , qui est donné par la relation :

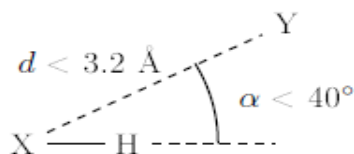
$$\mu = 1 - \beta(P_0 - P) \frac{\delta t}{\tau_p}$$

où  $\tau_p$  est l'unique paramètre de ce modèle et  $\beta$  est la compressibilité isotherme de l'eau.

### **A.2.5 Détermination du réseau de liaisons hydrogène**

Les liaisons hydrogène sont très importantes dans l'étude de la structure des protéines. Elles sont notamment à l'origine de la stabilisation des structures secondaires. Une liaison hydrogène se forme entre un atome d'hydrogène d'une liaison polaire et un atome qui porte un doublet non liant. Ces liaisons sont directionnelles et se forment grâce au recouvrement entre l'orbitale 1s d'un hydrogène et le nuage électronique du doublet non liant. Les simulations moléculaires classiques ne permettent pas de prendre en compte l'existence d'une liaison hydrogène car le nuage électronique est représenté de manière

implicite par l'intermédiaire de la charge effective. Pour caractériser l'existence de liaisons hydrogène un critère géométrique est utilisé, impliquant une distance et un angle (figure A.4).



**Figure A.4** Le critère utilisé pour savoir si une liaison hydrogène existe comporte la distance X-Y et l'angle  $\angle HXY$ . Les valeurs les plus souvent utilisées sont telles que  $d$  soit inférieur à 3.2 Å et que  $\alpha$  soit inférieur à 40°.

Les critères les plus souvent utilisés sont un angle de 40° et une distance de 3.2 Å. Ce critère permet de calculer au cours d'une simulation de dynamique moléculaire la probabilité pour qu'une liaison hydrogène existe.

## **A.2.6 Généralités sur la minimisation et la mise à l'équilibre du système**

Avant de démarrer une dynamique moléculaire il est nécessaire d'avoir une conformation raisonnable du système pour les conditions physico-chimiques dans lesquelles on souhaite faire la dynamique moléculaire. Pour un système biologique, les conditions habituelles sont une température ambiante (300K), une pression de 1 bar et un pH neutre (ou proche). Voici la méthode que nous avons employée dans ce travail.

### **A.2.6.1 Obtention des coordonnées de départ**

Les données structurales d'une protéine sont extraites des expériences de diffraction des rayons X de cristaux de protéines qui permettent d'obtenir la densité électronique du système. En combinant ces résultats expérimentaux et les bases de données sur les structures de protéines et les acides aminés on peut en déduire les coordonnées des atomes de la protéine qui sont reliés aux positions des maxima de la densité. La résolution de ces structures est définie comme la plus petite distance identifiable entre deux maximums de la densité et elle est de l'ordre 1 Å à 2 Å.

La limite inférieure de la résolution expérimentale est de l'ordre de grandeur de la longueur des liaisons X—H. En effet, les atomes d'hydrogène ont une faible densité électronique, ils diffractent donc faiblement les rayons X et les maxima de la densité électronique associés à ces atomes sont difficilement détectables et les positions des atomes d'hydrogène sont inconnues. De la même manière, dans les zones flexibles des structures de protéines, le nombre de conformations possibles est très important, ce qui rend les maxima de la densité électronique difficilement identifiables et les positions des atomes de certains acides aminés peuvent être indéterminées.

La première étape pour préparer le système avant une dynamique moléculaire consiste donc à rajouter les atomes manquants. D'autre part, une protéine étant un polyacide, il faut également choisir l'état de protonation des acides aminés acido-basiques. Cette étape a été détaillée au paragraphe 1.5. Enfin, une quantité suffisante de molécules d'eau est ajoutée autour de la protéine pour représenter le solvant. La position de ces molécules d'eau est obtenue à partir d'une configuration issue d'une simulation d'une boîte d'eau à température ambiante et à faible densité. Des contre-ions ( $\text{Na}^+$  ou  $\text{Cl}^-$ ) sont également ajoutés pour que la charge totale de la boîte de simulation soit nulle.

#### **B.2.6.2 Minimisation de l'énergie du système**

La minimisation de l'énergie du système va permettre de relaxer les fortes répulsions dues à l'ajout des atomes manquants et d'obtenir une meilleure description de la surface entre la protéine et le solvant. En effet, l'ajout des atomes à partir d'une géométrie prédéterminée introduit un choix arbitraire notamment pour les conformères des chaînes aliphatiques ou encore la position des groupes méthyle. Cela conduit souvent à une conformation aberrante de la protéine dans laquelle certaines distances inter-atomiques sont très courtes et impliquent de fortes répulsions, ce qui ne permet pas d'utiliser cette conformation comme point de départ d'une dynamique. D'autre part, la configuration des molécules d'eau autour de la protéine n'est pas adaptée à la présence de la protéine. Or l'interface entre la protéine et le solvant est cruciale puisqu'elle contrôle le repliement de la protéine et la maintient de sa structure.

Nous avons procédé en trois étapes. Dans un premier temps, seules les positions des atomes d'hydrogène sont optimisées car la grande majorité des atomes ajoutés sont des atomes d'hydrogène et la plupart des répulsions peuvent être facilement supprimées par

une simple rotation des méthyles ou des molécules d'eau. Ensuite, tous les atomes de la protéine sont fixés et on laisse se relaxer toutes les molécules d'eau. Cette étape est une première optimisation de la surface entre la protéine et le solvant. Cette optimisation est facilitée par la faible densité des molécules d'eau ajoutées autour de la protéine. Pour finir, l'énergie du système est optimisée en laissant libre les molécules d'eau et les contre ions et en affaiblissant les contraintes sur les atomes de la protéine (on utilise une constante de force de l'ordre de 5 - 10 kcal • mol). Cette contrainte permet aux chaînes latérales de se relaxer tout en garantissant une faible déviation par rapport à la structure cristalline. Cette dernière étape contribue faiblement à la baisse d'énergie potentielle. Ceci est dû au fait que le système est en phase condensée et que les déplacements possibles des chaînes latérales sont limités.

### **A.2.6.3 Mise à l'équilibre du système**

Les coordonnées obtenues à l'issue de la minimisation ne sont pas encore adaptées pour démarrer une dynamique moléculaire à température et pression constante. Il faut tout d'abord chauffer le système pour atteindre la température souhaitée et ajuster le volume de la boîte de simulation de sorte que la pression du système atteigne la valeur voulue. On procède généralement en deux étapes :

- Le système est équilibré en température, à volume constant.
- La pression du système est équilibrée en imposant une pression de 1 atm.

La configuration de départ étant issue d'une structure cristalline généralement obtenue à faible température, il est préférable de commencer la dynamique à faible température (100 K) afin d'éviter des collisions dues à des vitesses initiales trop importantes. Lors de l'augmentation de la température, le volume est maintenu constant. De plus, pour éviter un dépliement de la protéine, des contraintes faibles sont imposées sur les atomes de la protéine (5 - 10 kcal • mol). La température est augmentée progressivement de 100 K à 300 K puis maintenue à 300 K pour équilibrer le système.

La dernière étape consiste à équilibrer la pression en faisant une dynamique moléculaire dans l'ensemble (N,P,T) dans laquelle on fixe la pression du système à 1 atm. L'augmentation de la pression a pour effet de diminuer le volume et de faire augmenter la densité. Elle augmente, rapidement pendant les 100 premières picosecondes de la



simulation de 0.8 à 1. La relaxation des contraintes utilisées pour maintenir la structure de la protéine dans les étapes précédentes n'est généralement pas bien documentée. Au cours de ce travail, nous avons utilisé plusieurs protocoles différents variant en fonction de la situation dans laquelle se trouve le système au moment où les contraintes sont relaxées.

Les configurations obtenues correspondent à des configurations d'équilibre du système à la température et la pression voulue. La simulation peut alors être poursuivie jusqu'à obtenir la convergence des grandeurs que l'on souhaite calculer.

## Annexe B

### Calculs de chimie quantique

La chimie quantique correspond à un niveau de description du système dans lequel les électrons sont traités de manière explicite. Elle permet de calculer des propriétés électroniques du système comme le moment dipolaire ou encore les énergies de transition et les forces d'oscillateur. Dans cette thèse les calculs de chimie quantique ont été utilisés avec deux objectifs : le calcul des paramètres du champ de force du chromophore et le calcul des énergies de transitions électroniques et des forces d'oscillateur associées.

Les calculs de chimie quantique effectués au cours de cette thèse se place dans l'approximation de Born-Oppenheimer, c'est à dire que les noyaux sont considérés comme immobiles. Toutes les grandeurs calculées sont donc des fonctions des positions des noyaux atomiques. Les calculs de chimie quantique consistent en une résolution approchée de l'équation de Shrodinger et se différencient par le niveau des approximations utilisées et le type de formalisme employé. Dans ce qui suit seront présentées les méthodes utilisées au cours de cette thèse.

#### **B.1. Fonction d'onde et densité électronique**

La fonction d'onde électronique du système, est une fonction des coordonnées de l'ensemble des électrons du système. Elle est exprimée sous la forme d'un déterminant ou d'une somme de déterminants (méthode multi-référence) de Slater, qui garantissent que  $\psi_i$  est antisymétrique. Chaque déterminant représente une configuration électronique du

$$\psi_i = \sum_j c_{ij} \phi_j = \sum_j c_{ij} |\chi_1 \chi_2 \dots \chi_n|$$

où  $\chi_k$  est une orbitale moléculaire, OM, du système et  $n$  est le nombre d'électrons du système.

Dans la méthode Hartree-Fock, la fonction d'onde  $\psi_i$  est approchée par un seul déterminant, et les OM sont obtenues par diagonalisation de l'opérateur de Fock [58], les valeurs propres sont appelées énergies des OM. Les OM sont alors séparées en deux groupes : les OM occupées qui sont celles qui interviennent dans le déterminant considéré et les OM vacantes qui n'interviennent pas dans le déterminant.

On appelle HOMO (*Highest Occupied Molecular Orbital*) l'OM occupée d'énergie la plus élevée et LUMO (*Lowest Unoccupied Molecular Orbital*) l'OM vacante de plus basse énergie. Une transition électronique correspond à un transfert d'électron des OM occupées vers des OM vacantes.

La densité électronique du système  $\rho(r)$  est une fonction de trois variables définissant un point de l'espace. Elle est reliée à la fonction d'onde électronique par :

$$\rho(r) = \int \dots \int \psi_i^* \psi_i dr_1 \dots dr_{n-1}$$

Contrairement à la fonction d'onde, la densité électronique est une observable. Elle peut être mesurée expérimentalement par des expériences de diffraction des rayons X. D'autre part, alors que la fonction d'onde dépend de  $3n$  variables (les 3 coordonnées des  $n$  électrons du système), la densité électronique ne dépend que de 3 variables quelque soit la taille du système. En chimie quantique, on distingue les méthodes Hartree-Fock (HF) et post-Hartree-Fock [59], basées sur le calcul de la fonction d'onde du système, et les méthodes de la théorie de la fonctionnelle de la densité, DFT, basées sur le calcul de la densité électronique.

## **B.2 Calcul utilisant la théorie de la fonctionnelle de la densité**

### **Formalisme**

Les calculs de la théorie de la fonctionnelle de la densité, sont basés sur le théorème de Hohenberg et Kohn [60] qui dit que l'énergie de l'état électronique fondamental est complètement déterminée à partir la densité électronique,  $\rho(r)$  du système. L'énergie et les termes qui la compose sont alors écrits comme des fonctionnelles de la densité électronique (équation D.1). Kohn et Sham [61] ont proposé d'écrire l'énergie comme une fonctionnelle de la densité, de la façon suivante :

$$E_{DFT}[\rho] = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (\text{B.1})$$

Où  $T_s[\rho]$  est l'énergie cinétique d'un gaz d'électron sans interaction (dont l'expression analytique est connue),  $E_{ne}[\rho]$  est l'énergie d'interaction entre les noyaux et la densité électronique,  $J[\rho(r)]$  est l'intégrale de Coulomb et  $E_{xc}[\rho]$  est appelée la fonctionnelle d'échange-corrélation. C'est une fonctionnelle universelle (car elle ne dépend

pas de la position des noyaux), qui représente les effets négligés dans les autres termes de l'équation B.1. Elle est donnée par :

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] + J[\rho])$$

Où  $T[\rho]$  est l'énergie cinétique exacte du système et  $E_{ee}[\rho]$  est l'énergie d'interaction électronique exacte. Les trois premiers termes de l'équation .1 sont des expressions analytiques simples de la densité électronique. Pour faire un calcul DFT, on se donne alors une expression approchée de la fonctionnelle d'échange-corrélation.

### **Utilisation de la DFT**

La DFT présente deux avantages par rapport aux méthodes basées sur les équations Hartree-Fock. Tout d'abord, bien que la fonctionnelle d'échange-corrélation soit inconnue, la DFT inclue explicitement dans son formalisme le traitement de la corrélation électronique. Dans les méthodes post-Hartree Fock, le traitement de la corrélation électronique passe par l'utilisation de méthode multi-référence. D'autre part, pour deux calculs de précision comparable, celui utilisant la DFT est plus rapide que celui utilisant une méthode post-Hartree Fock et nécessite généralement moins de mémoire. La DFT a donc permis d'étendre le champ d'application de la chimie quantique à de plus grandes molécules (de l'ordre d'une centaine d'atomes) et de mettre en œuvre des méthode de dynamique quantique *ah initio* efficaces.

Il existe à l'heure actuelle un grand nombre de fonctionnelles d'échange-corrélation. Elles sont généralement classées en fonction de la façon dont elles dépendent de la densité électronique.

Les fonctionnelles les plus simples sont appelées fonctionnelles locale (LDA, *Local Density Approximation*). Elles supposent que la fonctionnelle est une moyenne pondérée par la densité électronique d'une fonction de la position qui ne dépend que de la valeur de la densité électronique à cette position. Une autre approche est d'inclure dans la fonctionnelle un terme qui dépend du gradient de la densité électronique (fonctionnelles GGA, *Generalized Gradient Approximation*). Dans ce type de fonctionnelles on trouve notamment les fonctionnelles PBE et BLYP. Le dernier type de fonctionnelles, appelées meta-GGA, inclue le laplacien de la densité électronique. La fonctionnelle B3LYP est l'une des fonctionnelles les plus utilisées en chimie moléculaire. C'est une fonctionnelle de type hybride construite à partir de fonctionnelles GGA et corrigée en ajoutant une partie de l'énergie d'échange calculée au niveau Hartree-Fock. Au cours de cette thèse, la plupart des calculs DFT ont été effectués avec cette fonctionnelle.

### **B.3 Etats excités et transitions électroniques**

#### **B.3.1 Opérateur d'échelle**

Soit la fonction d'onde de l'état électronique fondamental. Soit  $\psi_i$  la fonction d'onde d'un état électronique excité. L'opérateur d'échelle,  $T$ , est l'opérateur permettant de construire la fonction d'onde  $\psi_j$  à partir de  $\psi_1$

$$\varphi = T \varphi_1 \quad (B.2)$$

Si on suppose que la fonction d'onde de l'état excité peut s'écrire comme une combinaison linéaire de déterminants mono-excités, l'opérateur d'échelle se décompose comme une somme d'opérateurs mono-électroniques,  $\hat{t}_{kl}$ , qui substituent l'OM occupée  $\chi_k$  l'OM virtuelle  $\chi_l$ . L'opérateur d'échelle devient alors :

$$T = \sum_k \sum_l a_{kl} \hat{t}_{kl} \quad (B.3)$$

où la somme sur  $k$  porte sur les OM occupées et la somme sur  $l$  porte sur les OM vacantes. Les coefficients caractérisent le transfert de charge associé à la transition électronique. Si la transition électronique correspond au transfert d'un électron de l'orbitale HOMO vers l'orbitale LUMO, alors tous les coefficients  $a_{kl}$  sont nuls excepté  $a_{m+1}$ . Dans des systèmes comme le chromophore étudié dans cette thèse, le transfert de charge correspondant à la formation d'un état excité est généralement une somme de transferts partiels d'électrons de plusieurs OM occupées vers plusieurs OM vacantes.

#### **B.3.2 La force d'oscillateur**

La force d'oscillateur est une grandeur sans dimension qui quantifie la probabilité d'une transition électronique.

Soit une transition électronique d'un état  $i$  vers un état  $j$  et  $\nu$  la fréquence du photon émis ou absorbé lors de la transition.

Soit un électron de masse  $m$  et de charges  $e$  dans un potentiel harmonique en trois dimensions, de fréquence  $\nu$  égale à celle de la transition électronique définie ci-dessus.

La force d'oscillateur associée à la transition de l'état  $i$  vers l'état  $j$  est donnée par le rapport de la probabilité de cette transition, sur la probabilité de la première transition de l'électron dans le potentiel harmonique défini ci-dessus. Les probabilités de transition étant proportionnelles aux modules au carré des moments de transitions, le rapport des probabilités est égal au rapport des moments de transition :

$$f = \frac{|M_{ij}|^2}{|M_{12}^{harm}|^2}$$

où  $M_{ij}$  est le moment de la transition de l'état  $i$  vers l'état  $j$  et  $M_{12}^{harm}$  est le moment de la première transition de l'électron dans le potentiel harmonique. Les fonctions d'ondes d'un électron dans un potentiel harmonique sont connues, et le calcul du moment de transition  $M_{12}^{harm}$  aboutit à l'expression de la force d'oscillateur suivante :

$$f = \frac{8\pi^2 m \omega}{3e^2 h} |M_{ij}|^2 \quad (\text{B.4})$$

### **B.3.3 Calcul des énergies de transition**

Au cours de cette thèse nous avons utilisée deux méthodes permettant de calculer les énergies de transition : la méthode TDDFT, *Time Dependent Density Functional Theory*, et la méthode CIS, *Configuration Interaction Single*. Ces deux méthodes, implémentées dans Gaussian03, permettent de calculer pour une transition électronique, l'énergie de transition, la force d'oscillateur et les coefficients  $a_{kl}$  de l'opérateur d'échelle associé à l'état excité formé par la transition électronique.

La méthode TDDFT utilise le formalisme de la théorie de la fonctionnelle de la densité. L'interaction entre le nuage électronique et un champ électrique dépendant du temps est traitée par la théorie de la réponse linéaire dépendante du temps. Cette approche permet de calculer directement les coefficients  $c_{iki}$  de l'opérateur d'échelle et d'en déduire les énergies de transition, sans avoir à calculer la densité électronique du système.

La méthode CIS fait partie des méthodes dites post-Hartree Fock. Ces méthodes consistent à écrire la fonction d'onde du système comme une somme de déterminants. Les coefficients qui pondèrent ces déterminants sont calculés par diagonalisation de l'hamiltonien du système. Dans la méthode CIS, on ne considère que les déterminants mono-excités par rapport au déterminant de l'état fondamental (l'état fondamental est

représenté par un seul déterminant issu d'un calcul Hartree-Fock), et les OM (et donc les déterminants) ne sont pas modifiées au cours du calcul, seuls leurs poids respectifs varient. Ce calcul s'appelle une interaction de configuration. La fonction d'onde et l'énergie de l'état excité sont alors déterminées par la diagonalisation de l'hamiltonien dans la base des déterminants mono-excités.

## Résumé

Ce travail s'inscrit dans l'étude de l'influence du pH sur le spectre d'absorption de la Cerulean, un mutant de la GFP : la protéine fluorescente verte. La GFP est constituée de 230 acides aminés sous forme de tonneau, au centre de celui-ci se trouve fixée la partie fluorescente appelée chromophore issu de la cyclisation de trois acides aminés consécutifs (Ser65, Tyr66, Glu67).

L'importance des protéines de la famille de la GFP n'est plus à prouver. Grâce à leurs propriétés de fluorescence, elles permettent d'explorer un nombre croissant de processus biologiques in vivo. Les approches numériques, complémentaires aux études expérimentales, peuvent apporter une compréhension microscopique des processus mis en jeu et contribuer à l'interprétation des propriétés photophysiques de ces protéines.

Pour bien comprendre les comportements de cette protéine, on commence par présenter dans un premier chapitre les protéines: (propriétés physiques, chimiques, structure etc...), dans le deuxième chapitre, on présente la Cerulean et les paramètres de la fluorescence. La Cerulean est obtenue en faisant plusieurs mutations à la GFP dont la principale est Y66W. Enfin dans le dernier chapitre, on calcule le spectre d'absorption de la Cerulean à différents pH puis on le compare aux résultats expérimentaux et théoriques.

Nous avons démontré que l'influence des états de protonation est limitée, les résultats ont permis de confirmer l'hypothèse que le décalage du spectre d'absorption du rouge vers le bleu observé expérimentalement lors du passage du pH de 8 à 5 correspondant bien à une isomérisation du chromophore liée au changement de l'orientation de la chaîne latérale d'un acide aminé proche.

**Mots clés :** protéines fluorescentes, Cerulean, mutation, spectre d'absorption, dynamique moléculaire, TDDFT.

## Abstract

This work is part of the theoretical study of the influence of pH on the absorption spectrum of the Cerulean, who is a mutant of the GFP: Green Fluorescent Protein. It consists of 230 amino acids in the form of a barrel. In the center of this barrel is fixed the fluorescent part called the chromophore and derived from the cyclization of three amino acids (ser65, tyr66, glu67).

The importance of the GFP proteins is already proven. These proteins through their fluorescence properties allow us to explore a growing number of biological in vivo processes. Numerical approaches, combined to the experimental studies, can provide a microscopic understanding of the processes involved and contribute to the interpretation of the photophysical properties of these proteins.

To understand the properties of this protein, we present in the first chapter a background on proteins: (physical and chemical properties, proteins structure ...). In the second chapter, we present the structure of Cerulean and fluorescence parameters will be studied. The Cerulean is obtained by several mutations in the GFP, the main one is the Y66W mutation however there are at least eight other mutations.

Using various computational tools (molecular dynamics simulation and quantum computation, introduced in Chapter Three), it has been calculated the absorption spectrum of the Cerulean as a function of pH and compared with experimental and theoretical results.

After we show that the influence of electrostatic fields of neighboring residues of the chromophore is limited, the results confirmed the hypothesis of displacement of the absorption from red to blue (observed experimentally) when going from pH 8 to 5 corresponding well to an isomerization of the chromophore due the orientation changes of the closer amino acids side chains.

**Keywords:** fluorescent protein, cerulean, transfer, absorption spectra, molecular dynamics, TDDFT.

## ملخص

هذا العمل يدخل في إطار دراسة تأثير درجة الحموضة على طيف الإمتصاص لسيرولين , ناتج عن تحول GFP, بروتين مشع أخضر. GFP يتكون من 230 حمض أميني على شكل برميل, في وسطه يوجد العضو المشع الذي يدعى كروموفور الناتج عن ترابط دائري ل 3 أحماض أمينية متتالية (ulG67, ryT66, reS65). تعتبر البروتينات من عائلة GFP ذات أهمية كبيرة بفضل خصائصها المشعة فهي تسمح بالقيام ببحوث معمقة في مجال الخلية و البيولوجيا. إضافة إلى الدراسات التجريبية يمكن للمنهج النظري أن يؤدي إلى فهم و تحليل دقيق لخواص هذه البروتينات.

لمفهوم جيد لهته البروتينات خصصنا الفصل الأول لمعلومات عامة عن البروتينات (خواص فيزيائية وكيميائية وتركيبها...), في الفصل الثاني نشرح السيرولين و عوامل إشعاعها. يتم التحصل على سيرولين بعد إجراء عدة تحولات على GFP ومن أهمها Y66W . وأخيرا نحسب طيف الإمتصاص في عدة درجات الحموضة ثم نقارنها مع النتائج الأخرى.

لقد بينا أن ليس لحالات برتوناتيون تأثير هام والنتائج قد مكنت من إثبات نظرية أن تغيير لون الطيف من الأحمر إلى الأزرق بتقليل درجة الحموضة من 8 إلى 5 يعود إلى تغيير شكل الكلومور.

الكلمات الأساسية: بروتينات مشعة, سيرولين, تحول, طيف الإمتصاص, ديناميكية جزيئية.