



HAL
open science

Etude théorique de processus photophysiques dans des protéines fluorescentes

Germain Vallverdu

► **To cite this version:**

Germain Vallverdu. Etude théorique de processus photophysiques dans des protéines fluorescentes. Autre. Université Paris Sud - Paris XI, 2009. Français. NNT : . tel-00431879

HAL Id: tel-00431879

<https://theses.hal.science/tel-00431879>

Submitted on 13 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

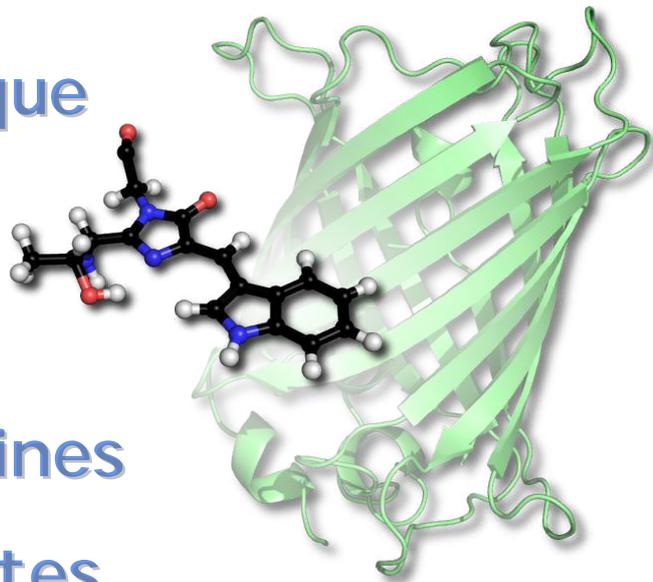
ORSAY

N° d'ordre : 9847

Université Paris Sud 11

Faculté des sciences d'Orsay

**Etude théorique
de Processus
Photophysiques
dans des Protéines
Fluorescentes**



Thèse présentée pour obtenir le grade de Docteur en
Sciences de l'Université Paris-Sud 11

par

Germain Vallverdu

Soutenue le 16 Juillet 2009 devant le jury composé de :

M. David PERAHIA (*Président*)
M. Xavier ASSFELD (*Rapporteur*)
M. Daniel BORGIS (*Rapporteur*)
M. Olivier PARISEL
Mme Fabienne MEROLA
Mme Isabelle DEMACHY (*Directrice de thèse*)

Remerciements

Je remercie Mehran Mostafavi, directeur du Laboratoire de Chimie Physique de m'avoir accueilli au sein de ce laboratoire pour effectuer mon travail de thèse. J'ai pu profiter de conditions de travail très agréables et de la grande diversité des activités de recherches des différents groupes du laboratoire.

Je tiens à remercier David Perahia, Olivier Parisel et Fabienne Mérola d'avoir accepté de juger ce travail et particulièrement Daniel Borgis et Xavier Assfeld d'en être les rapporteurs.

Je souhaite adresser de profonds remerciements à Isabelle Demachy, Jacqueline Ridard et Bernard Lévy pour leur encadrement d'exception. Au cours de ces quelques années j'ai pu profiter de leurs connaissances aussi diverses que nombreuses et j'ai pris beaucoup de plaisir à travailler quotidiennement à leurs côtés.

Je tiens à remercier Abdelfattah Halim et Jean-Marie Teuler, pour leur gentillesse et leur disponibilité pour résoudre tous les tracas d'un théoricien du Laboratoire. Leur aide précieuse m'a permis d'éviter de nombreuses catastrophes.

Je tiens également à remercier Pierre Archirel et probablement Pascal Pernot pour les nombreuses discussions très enrichissantes que j'ai pu avoir avec eux.

Je remercie également les expérimentatrices de l'axe P_3F , Hélène, Marie, Agathe et Isabelle pour les échanges et les discussions que j'ai pu avoir avec elles au sujet de ces fameuses protéines fluorescentes.

Un grand merci à Boris, Brahim, Sylvain, Pierre-Arnaud, Cécile, Marie, Christophe, Laetitia, Luiz, Emilie, Fabien, Yohan et bien d'autres avec qui j'ai pu partager mes années de thèse, mais aussi aux autres membres du Laboratoire de Chimie Physique sans oublier toutes ces tablettes de chocolat, sacrifiées tous les jours au profit de joyeuses pauses déjeuné.

Enfin je souhaite remercier ma famille, mes amis et plus particulièrement Cécile pour son aide et son soutien sans faille.

Merci à tous

Germain 

Sommaire

Introduction générale	7
I Les protéines fluorescentes : Contexte expérimental et méthodes d'étude numérique	9
I.1 Processus photophysiques en phase condensée	10
I.2 Les protéines de la famille de la <i>Green Fluorescent Protein</i>	13
I.3 Champ de force pour un acide aminé non standard	19
I.4 Calcul d'énergie libre – <i>Umbrella Sampling</i>	35
I.5 Distribution de protons dans une protéine	42
I.6 Calcul d'une incertitude par la méthode du <i>bootstrap</i>	45
II Relation entre pH, structure et spectre d'absorption de la Cerulean	49
II.1 Introduction	50
II.2 Dynamique moléculaire de la Cerulean	55
II.3 Spectre d'absorption en fonction du pH	84
II.4 Vers une description plus complète de l'effet du pH	99
II.5 Conclusion du chapitre	103
III Désactivation de la fluorescence de la <i>Green Fluorescent Protein</i>	105
III.1 Introduction	106
III.2 Méthodologie	111
III.3 Résultats de dynamique moléculaire	113
III.4 Calcul du déclin de la fluorescence	134
III.5 Conclusions et perspectives	153
Conclusion générale	155

A	Structure des protéines	157
A.1	Les acides aminés et la structure primaire	157
A.2	Structures secondaires	158
A.3	Structure tertiaire et structure quaternaire	161
B	Simulation moléculaire	163
B.1	Modélisation de la surface d'énergie potentielle	163
B.2	Simulation de dynamique moléculaire de protéines	167
C	Simulation de dynamique Brownienne	177
C.1	Les équations du mouvement	177
C.2	Résolution numérique de l'équation de Langevin	178
C.3	Tirage des nombres aléatoires corrélés	179
D	Calculs de chimie quantique	181
D.1	Fonction d'onde et densité électronique	181
D.2	Calcul utilisant la théorie de la fonctionnelle de la densité	182
D.3	Etats excités et transitions électroniques	183
	Bibliographie	187
	Liste des figures	197
	Liste des tableaux	207
	Résumé - Abstract	211

Introduction générale

Contexte du travail

Cette thèse présente l'étude de deux protéines fluorescentes appartenant à la famille de la *Green Fluorescent Protein*, réalisée au moyen de simulations numériques. Le travail que nous avons réalisé pendant ces trois années a été motivé par deux aspects différents : un aspect expérimental lié à la compréhension des caractéristiques photophysiques des protéines fluorescentes et un aspect théorique lié aux enjeux en rapport avec l'étude de ce type de système au moyen d'outils numériques.

Les protéines fluorescentes, homologues de la *Green Fluorescent Protein* sont devenues ces 10 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 sub-divisions du milieu cellulaire (noyaux, mitochondrie, appareil de Golgi ...). De plus, on dispose à l'heure actuelle d'une large gamme de protéines fluorescentes permettant de les utiliser dans des études très variées. Cependant, leur utilisation est à l'heure actuelle encore limitée par les propriétés des protéines elle mêmes. En effet, pour la plupart d'entre elles, leurs caractéristiques physico-chimiques et biochimiques sont complexes. Or 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 leur caractéristiques photophysiques. Ils pourront alors soit exploiter ces caractéristiques pour développer de nouveaux outils soit les intégrer à des modèles photophysiques permettant une analyse plus fine des données expérimentales. Dans cette optique 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. Cette complémentarité est un axe de recherche du laboratoire et le travail de thèse décrit dans ce manuscrit s'y inscrit pleinement.

Les études par simulation numérique jouent un rôle important dans l'étude de nombreux phénomènes, notamment les phénomènes physiques et chimiques qui nous intéressent. Elles se distinguent des approches analytiques qui nécessitent l'utilisation d'approximations importantes permettant de résoudre de manière analytique les problèmes posés. Les simulations numériques offrent un moyen de compléter et d'alimenter les études expérimentales ou théoriques car elles permettent de mesurer simultanément un grand nombre d'observables microscopiques (distances, positions ...) mais également macroscopiques (spectre d'absorption ou d'émission). Toutes ces données peuvent ensuite être utilisées pour affiner les modèles existants. De nos jours, un grand nombre d'outils numériques

existent avec des caractéristiques et des niveaux de descriptions très divers allant du calcul quantique de la fonction d'onde d'une molécule à la description mésoscopique d'un fluide ou d'un système colloïdal, en passant par les simulations moléculaires classiques à l'échelle atomique. Une description plus fine du système n'est pas nécessairement synonyme de l'obtention de résultats plus précis. Une méthode de simulation donnée est adaptée à un niveau de description et au calcul d'une propriété du système. C'est la combinaison de diverses méthodes de simulations qui permet d'étudier un système dans sa globalité. Nous avons adopté cette approche tout au long de cette thèse pour étudier les protéines fluorescentes et apporter des éléments de réponse dans l'étude des relations entre la structure la dynamique et les propriétés photophysiques de ces protéines. La particularité de cette étude réside donc dans la combinaison de simulations adaptées à diverses échelles de temps et de niveau de description, plus que dans l'utilisation des méthodes elles mêmes.

Présentation du manuscrit

Ce manuscrit est divisé en trois parties suivies de plusieurs annexes. Les caractéristiques générales des protéines de la famille de la *Green Fluorescent Protein* seront présentées, en terme de structure et propriétés photophysiques. Les approches employées pour modéliser des processus photophysiques dans les systèmes complexes comme le sont les protéines seront ensuite décrits.

La deuxième partie est consacrée à l'étude de la Cerulean, un mutant de la *Green Fluorescent Protein*, présentant une fluorescence de couleur cyan qui en fait un bon candidat pour la mise en oeuvre de nouvelles techniques d'imagerie cellulaire. Cependant, les études expérimentales montrent que le spectre d'absorption de la Cerulean présente une forte dépendance au pH, qui est probablement d'origine structurale. L'étude de la structure de cette protéine par simulation de dynamique moléculaire et le calcul de son spectre d'absorption en fonction du pH nous a permis de relier qualitativement et quantitativement la structure, le pH et le spectre d'absorption.

La troisième partie, présente une méthode permettant de calculer l'effet sur le déclin de la fluorescence d'un mécanisme de désactivation non radiatif de l'état excité. Nous avons appliqué cette méthode à l'étude d'une déformation interne du chromophore de la *Green Fluorescent Protein*. Cette étude implique la combinaison de simulations de dynamiques moléculaires biaisés et de dynamiques browniennes qui ont permises d'alimenter un modèle cinétique de l'état excité de la *Green Fluorescent Protein*. De ce modèle, nous avons extrait le rendement quantique de fluorescence et le temps de vie de l'état excité qui sont deux quantités qui intéressent particulièrement les expérimentateurs.

Enfin, les annexes regroupent une brève description de la structure des protéines ainsi que l'ensemble des méthodes de simulations classiques ou quantiques utilisées pour mener à bien ce travail.

L'ensemble des calculs de chimie quantique réalisés ont été fait avec le programme Gaussian03 [1]. Les simulations de dynamique moléculaires ont été réalisées avec la chaîne de programme AMBER [2], avec le champ de force *ff99* [3, 4].

Chapitre I

Les protéines fluorescentes : Contexte expérimental et méthodes d'étude numérique

Sommaire

I.1	Processus photophysiques en phase condensée	10
I.1.1	Diagramme de Perrin-Jablonski	10
I.1.2	Rendement quantique, temps de vie et constante de vitesse	12
I.2	Les protéines de la famille de la <i>Green Fluorescent Protein</i>	13
I.2.1	Un bref historique	13
I.2.2	Structure de la GFP	14
I.2.3	La famille de la GFP	16
I.2.4	Utilisation des protéines fluorescentes	17
I.3	Champ de force pour un acide aminé non standard	19
I.3.1	L'entité chromophore	20
I.3.2	Charges effectives	21
I.3.3	Champs de force pour le chromophore de la Cerulean	29
I.3.4	Champs de force pour le chromophore de la GFP	33
I.4	Calcul d'énergie libre – <i>Umbrella Sampling</i>	35
I.4.1	Considérations générales	36
I.4.2	Principe du <i>umbrella sampling</i>	37
I.4.3	Construction du profil d'énergie libre	39
I.5	Distribution de protons dans une protéine	42
I.5.1	Etat de protonation en phase aqueuse	42
I.5.2	Etat de protonation dans la protéine	43
I.5.3	Le serveur de calcul H++	44
I.6	Calcul d'une incertitude par la méthode du <i>bootstrap</i>	45
I.6.1	La méthode	45
I.6.2	Un exemple d'application de la méthode du <i>bootstrap</i> sur une distribution modèle	47

Dans ce chapitre, après avoir rappelé les principes et les grandeurs associées à l'étude des processus photophysiques en phase condensée, les protéines fluorescentes de la famille de la *Green Fluorescent Protein*, GFP, seront présentées. La suite du chapitre présentera les approches ou les méthodes de simulations développées ou utilisées au cours de cette thèse pour préciser et caractériser certains aspects des propriétés photophysiques de la Cerulean et de la GFP.

I.1 Processus photophysiques en phase condensée

Le but de cette partie est de donner le vocabulaire et les idées générales liées à l'étude des processus photophysiques de molécules en phase condensée [5,6].

I.1.1 Diagramme de Perrin-Jablonski

Le diagramme de Perrin-Jablonski (figure I.1) est commode pour visualiser simplement l'ensemble des phénomènes photophysiques existants dans les molécules en phase condensée. Il représente les états électroniques fondamental et excités d'une molécule (en noir) et la sous-structure vibrationnelle (en gris) de chaque état électronique. Les états électroniques singulets sont notés S_n et les triplets sont notés T_n . Dans la majorité des cas l'état fondamental est un état singulet, il est noté S_0 .

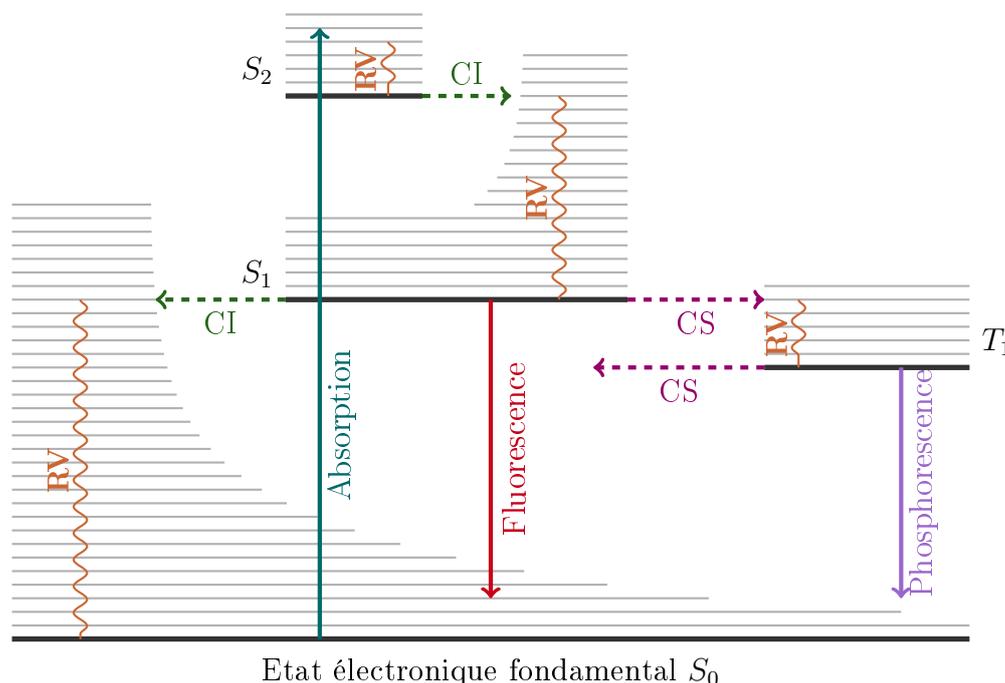


Figure I.1 – Diagramme de Perrin-Jablonski représentant les phénomènes photophysiques possibles dans un état excité. RV signifie relaxation vibrationnelle, CI signifie conversion interne et CS signifie croisement inter-système.

L'ensemble des processus photophysiques se divisent en deux familles. Les processus radiatifs, représentés en trait plein dans lesquels le système absorbe ou émet un photon et les processus non

radiatifs représentés en traits pointillés dans lesquels aucun échange de photons n'a lieu (figure I.1). Ces processus non radiatifs sont souvent nommés "processus de désactivation de la fluorescence", dans le sens où s'ils se produisent l'émission d'un photon n'a pas lieu (en anglais on parle de *quenching*).

La transition $S_0 \rightarrow S_1$ correspond à l'absorption d'un photon. L'efficacité de l'absorption à la longueur d'onde λ est caractérisée par le coefficient d'extinction molaire $\epsilon(\lambda)$. La probabilité que le système absorbe un photon est proportionnelle au module au carré du moment de transition. Pour étudier les transitions électroniques, on se place dans l'approximation de Born-Oppenheimer et on considère que les transitions entre états électroniques ont lieu à noyaux fixes. De plus, l'opérateur moment dipolaire ne dépendant que des coordonnées des électrons, la probabilité que le système absorbe un photon, A_{12} , s'exprime selon

$$A_{12} \propto \left| \langle \Psi_1^{el} | \vec{\mu} | \Psi_2^{el} \rangle \langle \chi_{1,\nu} | \chi_{2,\nu'} \rangle \right|^2 \quad (\text{I.1})$$

où Ψ_1^{el} est la fonction d'onde électronique de l'état fondamental, Ψ_2^{el} celle de l'état électronique excité atteint, les fonctions $\chi_{i,\nu}$ sont les fonctions d'ondes vibrationnelles du niveau ν de l'état électronique i et μ est le moment dipolaire de la molécule. Le dernier terme de cette équation représente le recouvrement entre deux fonctions d'onde vibrationnelles, l'une de l'état électronique fondamental, l'autre de l'état électronique excité. Il est appelé facteur Franck Condon. La région de la surface d'énergie potentielle de l'état fondamental où la fonction d'onde vibrationnelle est non nulle est nommée région Franck-Condon. Dans cette région, il existe le plus souvent une fonction d'onde vibrationnelle de l'état excité qui est non nulle, le facteur Franck Condon est donc non nul et le système a une certaine probabilité d'absorber un photon. La largeur des bandes d'absorption est donnée par la largeur de la région Franck Condon.

Après l'absorption d'un photon, la molécule est portée dans un état S_n électroniquement et vibrationnellement excité. Elle relaxe alors très rapidement (de quelques femtosecondes à quelques picosecondes) son excédent d'énergie vibrationnelle par des collisions avec l'environnement pour se placer dans le niveau vibrationnel le plus bas de S_n . Une succession de conversions internes et de relaxations vibrationnelles conduisent le système dans le niveau vibrationnel le plus bas de S_1 . La conversion interne est un processus non radiatif correspondant à une transition entre deux états singulet de spin sans émission de photon. Au cours de cette dynamique, le système quitte la région Franck-Condon et se stabilise dans la région de la surface d'énergie potentielle correspondant au minimum d'énergie de S_1 . Cette relaxation se produit en quelques picosecondes grâce à la forte densité d'états dans cette gamme d'énergie.

La fluorescence correspond à la transition radiative entre S_1 et un des niveau vibrationnel de l'état fondamental S_0 . Compte tenu de la relaxation vibrationnelle qui suit l'absorption d'un photon, elle a généralement lieu à partir du niveau vibrationnel de S_1 le plus bas. Le spectre de fluorescence est donc décalé vers les grandes longueurs d'onde par rapport au spectre d'absorption. Ce décalage du spectre de fluorescence est connu sous le nom de loi de Stokes. La fluorescence se produit entre 10^{-10} et 10^{-7} s après l'absorption d'un photon. Elle est en compétition avec deux autres processus : la conversion interne entre l'état S_1 et l'état S_0 et un croisement inter-système entre l'état S_1 et l'état

triplet T_1 qui peut être suivi soit de la phosphorescence soit d'un autre croisement inter-système de T_1 vers S_1 . Le croisement inter-système est un processus non radiatif tandis que la phosphorescence est un processus radiatif. Ils sont moins fréquents que la fluorescence car ils concernent des transitions entre états électroniques de multiplicités différentes dont la probabilité n'est significative qu'en présence d'un couplage spin-orbite important. La phosphorescence se différencie de la fluorescence par un temps caractéristique plus long, d'une microseconde à quelques secondes. De plus, l'état T_1 ayant une énergie plus basse que l'état S_1 , le spectre de phosphorescence se situe dans une gamme de longueur d'onde plus grande que celle de la fluorescence. Dans la suite de ce manuscrit, le croisement inter-système et la phosphorescence ne seront pas pris en compte.

I.1.2 Rendement quantique, temps de vie et constante de vitesse

A chacun des processus présents sur le diagramme de Perrin-Jablonski est associé 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 molécule est portée à l'état excité et atteint rapidement l'état S_1 . L'évolution de la population de S_1 en considérant les processus présentés sur le diagramme de Perrin-Jablonski est le plus souvent donnée par la relation suivante

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

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 :

$$N(t) = N(0) \exp(-t/\tau_s)$$

où $\tau_s = (k_r + k_{nr})^{-1}$ est homogène à un temps. τ_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 radiatif est élevé plus ce temps sera court. Le temps de vie de l'état excité peut être facilement obtenue 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 $N(t)$

$$I_f(t) = k_r N(t) = k_r N(0) \exp(-t/\tau_s) \quad (\text{I.2})$$

Dans ce cas simple, la fonction $I_f(t)$ est une exponentielle décroissante. On peut alors évaluer τ_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 τ_s . Le temps de vie de l'état excité τ_s ne doit pas être confondu avec le temps de vie

de fluorescence τ_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, ϕ_f . Il est mesuré expérimentalement et défini comme le rapport du nombre de photons émis sur le nombre de photons absorbés. Plus généralement l'efficacité des réactions photo-chimiques 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 $t=0$.

$$\phi_f = \frac{1}{N(0)} k_r \int_0^\infty N(t) dt \quad (\text{I.3})$$

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

$$\phi_f = \frac{k_r}{k_r + k_{nr}} = \frac{\tau_s}{\tau_f} \quad (\text{I.4})$$

Connaissant le rendement quantique de fluorescence et le temps de vie de l'état excité τ_s , on peut donc calculer le temps de vie de fluorescence τ_f . Le rendement quantique 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 ϕ_f . Elle peut également être utilisée pour quantifier l'efficacité de la fluorescence.

I.2 Les protéines de la famille de la *Green Fluorescent Protein*

Dans cette partie, la structure des protéines de la famille de la GFP et un aperçu général de leurs applications seront présentés, après un bref historique concernant leur découverte.

I.2.1 Un bref historique

La *Green Fluorescent Protein* [7,8,9,10], fut découverte en 1961 par O. Shimomura *et al* [11] comme étant un partenaire de l'aéuorin, une protéine chimio-luminescente de la méduse *Aequoria Victoria*. L'intérêt porté à cette méduse venait du fait que contrairement à d'autres organismes marins de la même famille, la lumière qu'elle émet n'est pas de couleur bleu, caractéristique de l'aéuorin (le spectre d'émission de l'aéuorin se situe aux environ de 470 nm) mais de couleur verte. La GFP a ensuite été isolée et sa fluorescence verte a été caractérisée par son spectre d'émission dont le maximum est situé aux alentours de 508 nm [12]. En comparant ce spectre à celui de l'émission de lumière de tissus vivant de la méduse *aequoria victoria* O. Shimomura identifia que c'est la GFP qui est responsable de l'émission de lumière dans cet organisme. Morise *et al* [13] ont ensuite montré que l'émission de lumière des tissus de la méduse *Aequoria Victoria* est due à un transfert

d'énergie de l'aequorin vers la GFP par un mécanisme de Förster [5, chapitre 9]. Ce décalage du spectre d'émission grâce à la combinaison de la GFP et de l'aequorin a été observée dans d'autres organismes marins [14]. Cependant, l'avantage biologique apporté par cette combinaison n'est pas encore établi.

La principale avancée vient des travaux de Prasher *et al* [15] qui ont réussi à cloner le gène de la GFP et ceux de Chalfie *et al* [16] qui ont exprimé le gène de la GFP dans d'autres organismes et ont observé que la fluorescence était conservée. Chalfie *et al* en ont déduit que le gène de la GFP contient toute l'information nécessaire pour la synthèse de son chromophore et que la présence de cofacteurs ou d'enzymes spécifiques de la méduse *aequoria victoria* ne sont pas nécessaires pour obtenir la fluorescence de la GFP. Ce résultat confère à la GFP un avantage majeur par rapport aux autres protéines fluorescentes et est le point de départ d'un développement considérable de son utilisation en imagerie. Ces dix dernières années, sous l'impulsion de R. Tsien [7] qui a conçu un grand nombre de mutants de la GFP avec des propriétés diverses, elles sont devenues un outil incontournable en imagerie cellulaire *in vivo*. O. Shimomura, M. Chalfie et R. Tsien se sont vu attribuer le prix Nobel de chimie 2008 pour leur travaux sur la découverte et le développement de la GFP.

I.2.2 Structure de la GFP

La GFP a été cristallisée pour la première fois par Morise *et al* [13]. La figure de diffraction aux rayons X de la structure a été obtenue en 1988 par Perrozo *et al* [17]. La séquence primaire de la GFP sauvage de la méduse *aequorea victoria* a été déterminée par Prasher *et al* [15] en 1992 et contient 238 acides aminés. La structure a été résolue une dizaine d'année après l'obtention des résultats des expériences de diffractions des rayons X simultanément par Ormö *et al* [18] (code 1EMA de la *Protein Data Bank* [19]) et par Yang *et al* [20] (code 1GFL de la *Protein Data Bank*)^a La difficulté pour résoudre la structure venait de l'utilisation de sélénométhionines qui ont été utilisées pour faciliter la cristallisation de la protéine. Celles-ci diffusent les rayons X et rendent les figures de diffractions difficilement interprétables. Les deux structures de la GFP obtenues par Ormö *et al* et par Yang *et al* protéines ne sont pas celles de la GFP sauvage. La GFP dont le code est 1GFL porte la mutation Q80R^b et la GFP dont le code est 1EMA porte la mutation S65T.

La structure tridimensionnelle de la GFP a une forme cylindrique due à 11 brins qui forment deux à deux des feuillets β (figure I.2 a). Cette forme particulière donne à la structure le nom de tonneau- β (β -barrel ou β -can) et certains auteurs qualifient la GFP de *light in a can* [8]. Les dimensions de ce cylindre sont d'environ 24 Å de diamètre et 42 Å de hauteur. La majorité de la chaîne primaire forme des structures secondaires ce qui confère à la structure une rigidité importante notamment dans la région centrale. Une hélice α dirigée suivant l'axe du cylindre de la protéine porte environ à mi-hauteur le chromophore responsable de la fluorescence (figure I.2 b et c). Le chromophore n'est

^aPar la suite ces structures obtenues grâce aux expériences de diffractions des rayons X seront appelées structure RX.

^bLa notation X N Y décrit la mutation de l'acide aminé X en position N par l'acide aminé Y. Le code des acides aminés à une lettre est très souvent utilisé dans ce cas. Le nom des acides aminés et les codes à une lettre sont donnés en annexe A.1.

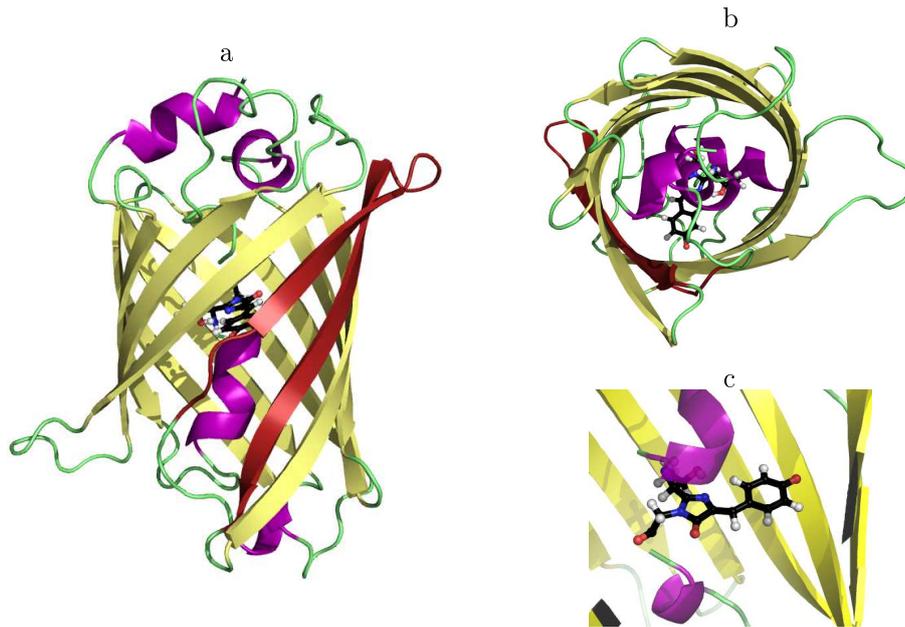


Figure I.2 – a) et b) Structure tridimensionnelle de la GFP vue de face et de dessus. Les feuillets rouges définissent le *cleft* qui est une faiblesse de la structure de la GFP. c) Vue du chromophore de la GFP à l'intérieur de la protéine. Il est lié à l'hélice α centrale par deux liaisons covalentes.

pas un cofacteur, il résulte de la cyclisation de trois acides aminés Ser65, Tyr66 et Gly67. Il avait été identifié par O. Shimomura comme étant une 4-(*p*-hydroxybenzylidène)imidazolidin-5-one [21]. Il se forme à la suite du repliement de la protéine suivant le mécanisme présenté figure I.3 [22]. Le temps de repliement de la protéine et de la formation du chromophore a été mesuré par Reid *et al* dans le mutant S65T et dure environ 100 min [23]. Il ne nécessite aucune intervention d'une enzyme, seule la présence d'oxygène est requise. Le chromophore de la GFP est formé de deux cycles aromatiques, un cycle imidazolinone et un cycle indole, reliés par un groupement CH. L'ensemble forme un système π étendu qui confère au chromophore une fluorescence dans le domaine visible. Les orbitales mises en jeu dans l'absorption et l'émission de lumière sont présentées sur la figure I.4.

La structure cylindrique de la GFP présente deux particularités. Tout d'abord, un nombre important d'acides aminés polaire ou chargés ont leur chaîne latérale à l'intérieur de la protéine et à proximité du chromophore, en particulier Gln69, Arg96, Gln94 His148, Thr203, Ser205 et Glu222 qui forment un réseau dense de liaisons hydrogène autour du chromophore. Ces acides aminés forment également des liaisons hydrogène avec des molécules d'eau cristallographiques présentes à l'intérieur de la protéine. D'autre part, le tonneau de feuillets β présente une irrégularité entre deux brins [24] (en rouge sur la figure I.2 a). Par rapport à la structure idéale d'un feuillet β , il manque un certain nombre de liaisons hydrogène entre des atomes des deux brins, trop éloignés les uns des autres. Dans la suite de ce travail, pour nommer cette irrégularité nous reprendrons le terme anglais *cleft* (crevasse) utilisé dans la référence [24]. Ce *cleft* est situé proche du chromophore et est une voie d'échange possible de molécules d'eau ou de protons entre l'intérieur de la protéine et le solvant.

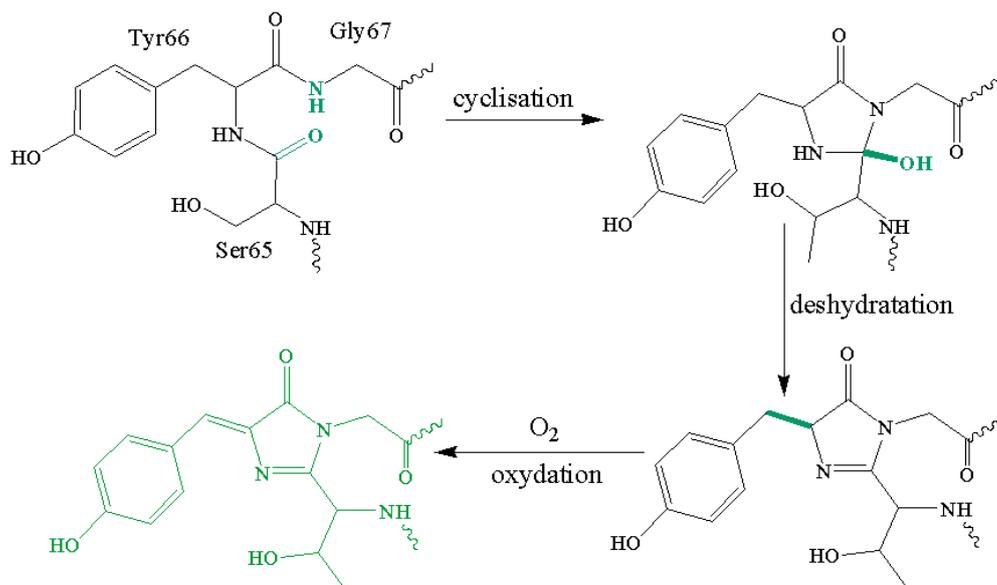


Figure I.3 – Mécanisme proposé par Cubitt *et al* [22] pour la synthèse intra-moléculaire du chromophore de la GFP

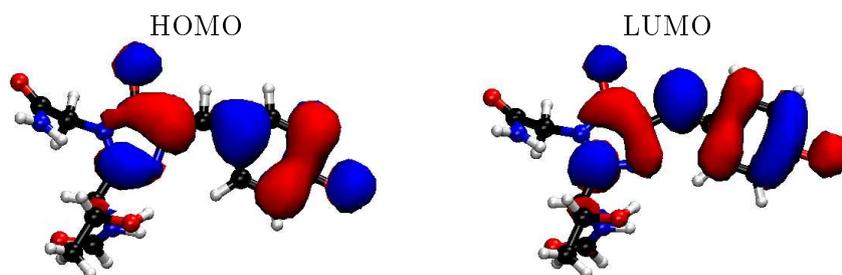


Figure I.4 – Orbitale moléculaire HOMO et LUMO du chromophore de la GFP obtenue par un calcul B3LYP/6-31G*.

I.2.3 La famille de la GFP

La famille de la GFP comportent un grand nombre de protéines. Certaines ont été conçues par les biologistes pour obtenir des mutants performants, fluorescent dans une gamme de couleur différente (figure I.5) [25, 26] (elles diffèrent généralement de la GFP sauvage par au plus une dizaine de mutations). D'autre, les homologues, ont été découvertes dans différents organismes marins. Elles possèdent toutes la structure en tonneau formé de 11 feuillets β décrite plus haut pour la GFP. Le chromophore se forme de la même manière, à partir de la cyclisation de 3 acides aminés, parmi lesquels on retrouve toujours la glycine en position 67. Il comporte toujours deux cycles aromatiques dont l'un est un cycle imidazolinone, reliés par un pont formé d'un atome de carbone.

Dans les mutations qui affectent le chromophore on trouve d'une part la mutation de la tyrosine 66 par un autre acide aminé possédant un cycle aromatique (histidine, phenylalanine ou tryptophane), et d'autre part la mutation de la serine 65 en thréonine, glycine, glutamine ou méthionine [25, 26].

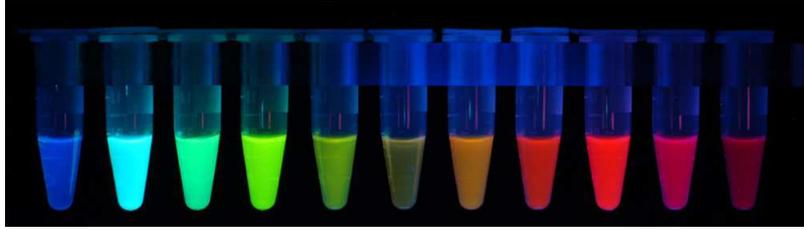


Figure I.5 – Echantillon des protéines fluorescentes actuellement disponibles.

La mutation de Tyr66 modifie fortement le spectre d'absorption et de fluorescence, en les déplaçant vers le bleu. La mutation de la sérine 65 en méthionine résulte en une extension du système π et à un décalage des spectres d'absorption et d'émission vers le rouge. Les mutations ne touchant pas le chromophore ont en général un effet plus limité, par exemple, la mutation de la thréonine 203 par une tyrosine conduit à un décalage vers le jaune, dû à une interaction entre le système π du chromophore et celui de la tyrosine 203.

Parmi les protéines de cette famille les plus utilisées, on trouve, outre la GFP elle-même, la *Cyan Fluorescent Protein*, CFP [27,28], dont la mutation principale est celle de la tyrosine 66 en tryptophane et la *Yellow Fluorescent Protein*, YFP [18,29], dont la mutation principale est celle de la thréonine 203 en tyrosine.

I.2.4 Utilisation des protéines fluorescentes

L'utilisation des protéines de la famille de la GFP qui a eu le plus de succès est la sonde fluorescente. Le gène qui code pour la GFP est ajouté au gène qui code pour une protéine d'intérêt, appelée protéine hôte. L'expression de cette dernière conduira à un ensemble fluorescent correspondant à la protéine hôte à laquelle est accroché une GFP. La fusion idéale consiste à obtenir un ensemble dans lequel ni les propriétés de la protéine fluorescente utilisée ni celles de la protéine hôte ne sont altérées. Grâce à la fluorescence il est alors aisé de localiser et de suivre les mouvements de la protéine hôte dans la cellule ou dans un organisme vivant. Ce type d'expérience a été réalisée avec succès en utilisant des protéines hôtes variées et localiser dans pratiquement tous les organites de la cellule [7].

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. D'autre part, dans la majorité des applications, la sensibilité à l'environnement n'est pas recherchée car elle est souvent mal contrôlée.

Le transfert résonnant d'énergie de fluorescence, FRET [5, Chapitre 9], a été largement utilisé pour concevoir des sondes fluorescentes sensibles à l'environnement. Le FRET a lieu par un mécanisme de Förster et ne peut se produire que si les deux fluorophores (dans notre cas deux GFP) sont suffisamment proches ($<100 \text{ \AA}$) et le spectre d'émission du fluorophore donneur recouvre au moins partiellement le spectre d'absorption du fluorophore accepteur. Dans ces conditions, une fois porté

dans son état excité, le donneur peut transférer son énergie à l'accepteur de manière non radiative. On observe donc la fluorescence de l'accepteur au détriment de celle du donneur. La probabilité que le FRET se produise est proportionnelle à $R_0^6/(R_0^6 + r^6)$ où r est la distance entre le donneur et l'accepteur et R_0 est la distance à laquelle la probabilité que le FRET se produise est de 1/2. R_0 dépend du rendement quantique de fluorescence du donneur, du coefficient d'extinction molaire de l'accepteur, du recouvrement entre le spectre d'émission du donneur et le spectre d'absorption de l'accepteur et de l'orientation entre les fluorophores. La plupart des mutations et des améliorations faites sur les GFP ont pour but d'augmenter la probabilité que le FRET ait lieu.

Le premier type d'indicateur mis au point avec des GFP en mettant à profit le FRET est un indicateur à cations Ca_2^+ [30,31]. Le FRET est également très utilisé dans l'étude des interactions entre protéines. En fusionnant une GFP sur une première protéine hôte et une autre GFP sur une deuxième protéine hôte, il est possible d'identifier par le FRET, les zones de la cellule où les deux protéines hôtes sont proches [32]. Dans ce type d'expérience la CFP est généralement utilisée comme donneur et la YFP est utilisée comme accepteur [33,34]. Pour ce couple donneur/accepteur R_0 vaut entre 49 et 52 Å [7] (la valeur de R_0 mesurée maximale pour un couple de GFP étant de l'ordre de 60 Å) et les maxima des spectres d'émission de ces deux protéines sont suffisamment séparés pour être distingués expérimentalement.

Le principal avantage du FRET est qu'il donne une résolution spatiale et temporelle du phénomène observée et qu'il peut être mis en oeuvre dans presque tous les compartiments de la cellule. Son principal désavantage est lié au fait que l'accepteur, même en l'absence de donneur, est fluorescent [7]. En effet, le couple donneur/accepteur est choisi de sorte que le spectre d'émission du donneur et le spectre d'absorption de l'accepteur se recouvrent ce qui impose généralement que le spectre d'absorption de l'accepteur et du donneur se recouvrent légèrement. Cela pose un problème pour mesurer l'efficacité du FRET qui est généralement obtenue par le rapport de l'intensité de fluorescence de l'accepteur sur celle du donneur. Du fait que ce rapport est non nul même en absence de FRET, sa variation doit être importante pour être significative.

Une alternative pour contourner cette difficulté est de quantifier l'efficacité du FRET en mesurant la variation de la durée de vie du donneur. En effet, du point de vue du donneur, le FRET est un processus de désactivation non radiatif de son état excité et, plus il est efficace, plus la durée de vie de l'état excité du donneur sera courte. Cette approche ne nécessite aucune mesure sur l'accepteur et est donc plus rapide et généralement plus fiable. De plus, la durée de vie est indépendante de la concentration en protéine dans la cellule qui est souvent mal connue et présente de fortes hétérogénéités. Au laboratoire, le groupe de biophysique travaille au développement de ces techniques d'imagerie appelée imagerie de temps de vie de fluorescence (FLIM, *Fluorescence lifetime imaging microscopy*). L'une des principales limites à l'utilisation des techniques de FLIM-FRET est que la mesure du temps de vie peut être complexe car le déclin de la fluorescence peut présenter plusieurs composantes de temps qui dépendent de l'environnement. C'est le cas de la CFP qui est le donneur le plus utilisé et présente un déclin de la fluorescence complexe [35]. L'amélioration de la CFP a conduit à la Cerulean [36,37] qui fait l'objet de l'étude présentée au chapitre II.

Plus récemment, des homologues de la GFP avec de nouvelles propriétés ont été développées [38,

39, 40]. Ces protéines sont dites activables. On distingue deux types : les protéines photoactivables qui, sous l'action d'un laser, passent d'un état excité non fluorescent à un état excité fluorescent ; et les protéines photoconvertibles dont la position du spectre de fluorescence est décalée sous l'action d'un laser. Ces protéines sont de puissants outils pour l'étude de la dynamique des protéines en cellule vivante car elles peuvent être activées à un instant précis.

I.3 Champ de force pour un acide aminé non standard

Au cours de cette thèse, nous avons mis en oeuvre des simulations de dynamique moléculaire de la GFP ainsi que de la Cerulean. Les principes de base de ce type de simulations sont présentés en annexe B. On rappellera simplement ici que les simulations de 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. Le choix du champ de force est donc une des premières étapes avant de faire une simulation de dynamique moléculaire. Dans le programme AMBER utilisé pour les simulations, le champ de force couramment employé est *ff99*. Son expression est rappelée ici :

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

Les protéines de la famille de la GFP possèdent un chromophore qui n'est pas un acide aminé standard. Or, les paramètres des champs de force à notre disposition définissent uniquement les paramètres des acides aminés standard. Il est donc nécessaire de déterminer les paramètres associés au chromophore de ces protéines. Pour cela, certaines contraintes liées à l'utilisation de la suite de programme AMBER doivent être respectées :

- la forme analytique du champ de force utilisé pour le chromophore doit être la même que celle du champ de force *ff99* utilisé pour les acides aminés standard.
- Pour être cohérent avec le champ de force *ff99* utilisé pour les autres acides aminés, les paramètres de l'acide aminé non standard doivent être déterminés d'une manière aussi proche que possible de celle employée pour les autres acides aminés.

Il n'existe pas de champ de force universel. Ainsi, une paramétrisation du champ de force ne peut être valide ni pour l'ensemble de la surface d'énergie potentielle du système, ni pour étudier tout type de propriétés. Les paramètres doivent être calculés pour répondre à un besoin précis. Leurs calculs doivent se faire dans une région de l'espace des configurations qui recouvre au mieux la zone qui sera explorée au cours de la dynamique. Ainsi, plusieurs champs de force ont été mis au point suivant la protéine étudiée et le type d'étude effectuée.

L'étude concernant la Cerulean met en jeu deux conformations du chromophore à l'état électronique fondamental, nommée *cis* et *trans* (figure I.6). Cette nomenclature sera justifiée au chapitre II. Elle

nécessite de décrire correctement la dynamique du chromophore dans les bassins d'équilibre autour de ces deux conformations. Nous avons donc déterminé un champ de force du chromophore à l'état fondamental pour chacune de ces deux conformations. Dans l'étude sur la GFP, nous nous sommes intéressés au comportement du premier état électronique excité du chromophore, particulièrement à la dynamique de la torsion φ (figure I.6 pour la définition de l'angle) sur l'ensemble de son domaine de définition.

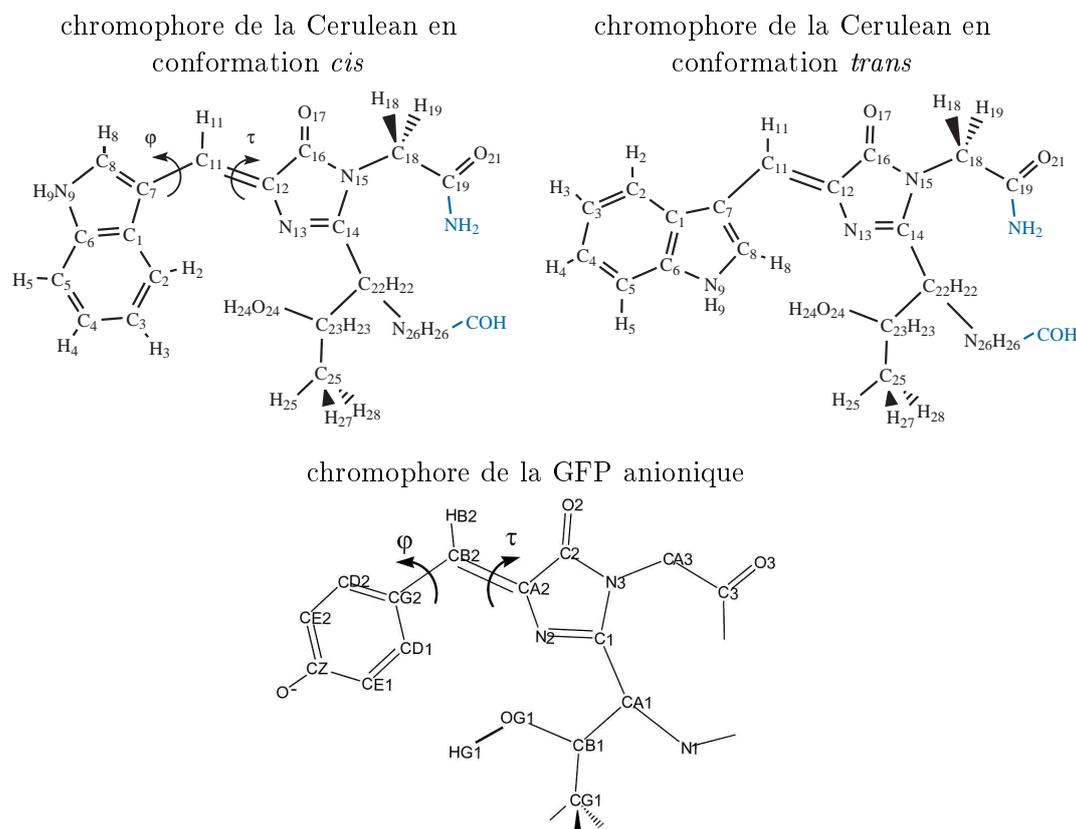


Figure I.6 – Représentation et numérotation des atomes du chromophore de la Cerulean en conformation *cis* et *trans* et du chromophore anionique de la GFP.

Nous allons maintenant détailler la démarche utilisée pour construire ces différents champs de force. Nous commencerons par définir l'entité chromophore pour laquelle les paramètres du champ de force doivent être calculés. Nous présenterons ensuite la paramétrisation des deux champs de force du chromophore de la Cerulean puis du champ de force du chromophore de la GFP.

I.3.1 L'entité chromophore

Le chromophore des GFP résulte de la cyclisation de trois acides aminés (en position 65, 66, 67) [7] et présente les mêmes terminaisons qu'un acide aminé standard. Chacune de ses extrémités est reliée au reste de la protéine par une liaison peptidique. Conventionnellement, la séparation entre deux acides aminés dans l'enchaînement du squelette protéique se fait entre le NH et le CO de chaque liaison peptidique (figure I.7).

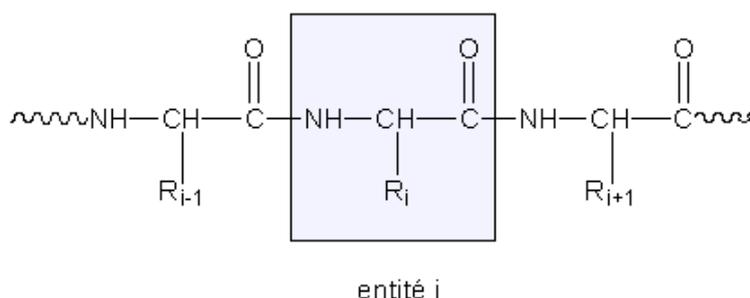


Figure I.7 – Schéma représentant une chaîne d'acides aminés reliés par des liaisons peptidiques. Le cadre délimite l'entité associée à un acide aminé standard dans AMBER.

Dans le champ de force *ff99*, chaque acide aminé, délimité par ses deux extrémités NH et CO (figure I.7), constitue une entité à laquelle est associée une librairie qui contient tous les paramètres du champ de force nécessaire pour l'entité considérée. Dans notre cas, il faut donc construire une nouvelle entité qui correspondra au chromophore.

Certains paramètres des champs de force des chromophores de la GFP et de la Cerulean ont été obtenus à partir de calculs de chimie quantique. Pour faire ces calculs, il est préférable de compléter les liaisons manquantes des atomes terminaux afin qu'ils aient le nombre de liaisons adéquates. De cette manière, l'environnement chimique des atomes terminaux de l'entité chromophore reste proche de celui existant lorsque le chromophore est relié à la protéine. Pour conserver la conjugaison des liaisons peptidiques, un groupement COH est rajouté sur la terminaison NH du chromophore et un groupement NH₂ est rajouté sur sa terminaison CO (en bleu sur la figure I.6). Ce modèle sera utilisé par la suite comme modèle du chromophore pour réaliser les calculs quantiques nécessaires pour déterminer les champs de force du chromophore de la GFP et de la Cerulean. Il sera désigné simplement sous le nom de chromophore.

I.3.2 Charges effectives

Une bonne représentation des interactions électrostatiques est très importantes pour obtenir une bonne description de la structure et de la dynamique des systèmes biologiques. Un moyen simple de décrire ces interactions est d'utiliser des charges effectives représentant la densité électronique de la molécule. Deux approches différentes peuvent être envisagées pour les déterminer. La première consiste à les déterminer de manière empirique de sorte que les simulations, dans lesquelles elles sont utilisées, reproduisent au mieux des données expérimentales. Ce type d'approche est mis à profit dans le cas des charges de type OPLS [41]. L'autre approche consiste à calculer les charges effectives à partir de calcul de mécanique quantique *ab initio*. Les charges effectives sont alors ajustées de façon à reproduire au mieux le potentiel électrostatique calculé au niveau *ab initio* sur une grille de points [42].

Nous avons utilisé cette deuxième approche pour déterminer les charges effectives pour le chromophore de la Cerulean et de la GFP. Après avoir décrit cette approche, les résultats obtenus pour le

chromophore de la Cerulean et celui de la GFP seront présentés.

I.3.2.1 Détermination de charges effectives à partir de calculs *ab initio*

Les charges effectives sont placées sur les atomes. Elles sont ajustées de sorte qu'elles reproduisent le potentiel électrostatique créé par la densité électronique calculée au niveau *ab initio*. On choisit alors une grille de points autour de la molécule et on calcule le potentiel électrostatique, V_k^{ab} , en chaque point k de cette grille par une méthode *ab initio*. On cherche alors les charges qui minimisent l'écart quadratique moyen entre V_k^{ab} et V_k^{charge} , le potentiel créé en chaque point de la grille par la répartition de charges effectives, c'est à dire la fonction χ^2 suivante

$$\chi^2 = \frac{1}{N_{pts}} \sum_k^{N_{pts}} (V_k^{ab} - V_k^{charge})^2 \quad (\text{I.5})$$

où k désigne un point de la grille. Les dérivées premières de χ^2 par rapport aux charges effectives sont des fonctions linéaires des charges effectives qui peuvent donc être obtenues analytiquement, par un calcul matriciel.

La racine carré de la fonction χ^2 , nommé RMS (*Root Mean Square*) donne une indication de la qualité de la répartition de charge obtenue. L'écart quadratique moyen relatif, RRMS, correspond au RMS divisé par la racine de la moyenne des carrés des valeurs du potentiel électrostatique calculées *ab initio* (équation I.6). Il donne une indication relative de la qualité de la répartition de charges obtenues.

$$RRMS = \sqrt{\frac{\sum_k^{N_{pts}} (V_k^{ab} - V_k^{charge})^2}{\sum_k^{N_{pts}} (V_k^{ab})^2}} \quad (\text{I.6})$$

Dans le champ de force *ff99*, le potentiel électrostatique *ab initio* utilisé pour déterminer les charges a été calculé au niveau Hartree-Fock, HF, avec une base 6-31G* [3].

Une des difficultés liées à l'utilisation de cette méthode vient du fait que les charges obtenues dépendent souvent de la grille de points utilisée et de la géométrie de la molécule. L'origine de ces variations est que le nombre de charges effectives utilisées est souvent trop élevé et certaines charges sont alors indéterminées. C'est souvent le cas des atomes "enfouis" comme les atomes de carbone de groupement méthyle. Ceux-ci sont entourés de trois autres charges plus proches des points de la grille. La charge de l'atome de carbone aura donc peu d'effet sur la valeur du potentiel et sera indéterminée. Pour contourner ce problème, différentes approches ont été proposées, la plus répandue étant la méthode RESP [43,44] (*Restraint Electrostatic Potential*). Dans cette méthode, on rajoute dans l'équation I.5 un terme qui permet de contraindre les charges autour d'une valeur

cible. Ces contraintes, sont généralement de forme harmonique ou hyperbolique et permettent de limiter les fluctuations des charges en fonction de la géométrie ou de la grille de points.

B. Lévy et J. Ridard, au laboratoire, ont développé il y a quelques années une autre approche. Elle consiste à ajouter des contraintes linéaires, traitées de manière exacte et permettant de réduire le nombre de paramètres indéterminés [45, 46]. Ces contraintes, permettent de fixer la valeur de certaines charges ou la valeur de la somme d'un groupe de charges. Des contraintes utilisant une force de rappel harmonique peuvent également être employées. Dans ce cas, on rajoute dans la fonction χ^2 que l'on minimise (équation I.5) une expression de la forme :

$$\chi_{\text{harmonique}} = \sum_i k_i (q^i - q_c^i)^2 \quad (\text{I.7})$$

où q_i désigne la charge effective de l'atome i , q_c^i désigne la charge cible vers laquelle doit tendre la charge effective q^i et k_i est un poids qui permet d'ajuster la force avec laquelle s'exerce cette contrainte.

Pour déterminer les charges du chromophore de la GFP et de la Cerulean, nous avons utilisé à la fois des contraintes linéaires et harmoniques. Le calcul de ces charges est présenté ci-dessous.

I.3.2.2 Calcul des charges effectives

La position et le nombre de charges effectives sont indépendants de la topologie de la molécule. Dans certains modèles on utilise plusieurs charges effectives par centre atomique afin de représenter, par exemple, la densité électronique sur un atome par un dipôle. Dans le champ de force *ff99*, une charge effective est positionnée sur chaque centre atomique. Pour que les paramètres calculés pour le chromophore soient compatibles avec le champ de force *ff99*, le potentiel électrostatique du chromophore sera ajusté en plaçant une charge sur chaque centre atomique.

Le chromophore de la Cerulean contient 46 atomes et celui de la GFP en contient 42. La détermination des charges effectives nécessite donc de minimiser une fonction de 45 ou 41 variables si on impose la valeur de la charge totale du chromophore. C'est un problème en partie indéterminé à cause de la présence d'atomes "enfouis" à l'intérieur de la molécule. Nous avons introduit un certain nombre de contraintes dont l'un des effets est de limiter cette indétermination.

a) Choix de la grille de points

La grille de points est déterminée en répartissant des points de manière homogène sur la surface obtenue en fusionnant les surfaces de sphères centrées sur chaque atome du chromophore. Quatre surfaces sont utilisées simultanément. Elles sont construites à partir de sphères de rayons différents centrées sur les atomes. De cette manière, aussi bien des points proches qu'éloignés des centres atomiques peuvent être utilisés. Les charges effectives obtenues représenteront correctement le potentiel électrostatique à différentes distances.

b) Chromophore de la GFP

Comme on souhaite modéliser la surface d'énergie potentielle de l'état électronique excité fluorescent du chromophore de la GFP, les charges effectives ont été ajustées sur le potentiel électrostatique créé par la densité électronique de cet état électronique. La géométrie de l'état excité du chromophore de la GFP a été tout d'abord partiellement optimisée au niveau CIS/6-31G* en fixant l'orientation des acides aminés Thr65 et Gly67 à leur position dans la structure RX. Le potentiel électrostatique a ensuite été calculé au niveau CIS/6-31G*. Les charges effectives de l'état fondamental ont aussi été calculées. Elles ont été ajustées sur le potentiel électrostatique calculé au niveau HF/6-31G* en utilisant la géométrie de l'état fondamental partiellement optimisée au niveau B3LYP/6-31G* en fixant l'orientation des acides aminés Thr65 et Gly67 à leur position dans la structure RX. Les géométries de l'état électronique fondamental et de l'état électronique excité sont peu différentes et le chromophore est quasi plan. Les coordonnées internes les plus modifiées sont les longueurs des liaisons du ponts reliant les deux cycles aromatiques.

Plusieurs commentaires peuvent être faits sur la méthode de calcul utilisée pour l'état excité fluorescent. Tout d'abord, la base employée est petite et ne comporte pas de fonctions diffuses ce qui n'est généralement pas recommandé pour les calculs d'états excités. Cependant, comme le chromophore est confiné à l'intérieur de la protéine, il est probable que l'expansion du nuage électronique lors du passage à l'état électronique excité soit contraint par la présence d'acides aminés proches du chromophore. Limiter la taille de la base a donc ici un sens physique qui est de modéliser le confinement du nuage électronique du chromophore à l'intérieur de la protéine. D'autre part, la méthode CIS n'est pas la méthode la plus précise pour calculer la densité électronique d'un état excité. Les méthodes de type TDDFT ou CASSCF sont généralement plus précises car elles incluent une partie plus importante de la corrélation électronique. Cependant, le calcul de la densité électronique de l'état électronique excité au niveau CIS, pour déterminer les charges effectives du chromophore, est cohérent avec l'utilisation de calcul au niveau HF pour l'état électronique fondamental dans le champ de force *ff99*.

Un ensemble de contraintes linéaires ont été introduites dans le calcul des charges :

1. Le chromophore étant anionique, de charge -1, nous avons imposé que la somme des charges effectives soit égale à -1.
2. La somme des charges effectives des atomes ajoutés pour fermer l'entité chromophore doit être nulle (atomes en bleu sur la figure I.6 page 20).
3. Les trois atomes d'hydrogène du groupement méthyle de la partie correspondant à Thr65 sont chimiquement équivalents. Nous avons donc imposé que les charges des trois hydrogènes soient égales.

Une fois ces contraintes introduites, le RRMS est de 3% pour les charges de l'état fondamental et celles de l'état excité fluorescent. Les résultats numériques sont donnés dans le tableau I.1.

Les charges effectives obtenues pour l'état fondamental sont en bon accord avec celles déterminées par Helms *et al* [24]. Lors du passage de l'état électronique fondamental à l'état électronique excité,

atome	S ₀	S ₁	atome	S ₀	S ₁
<i>Imidazolinone</i>			<i>Partie Thr65</i>		
C1	0.32270	0.29711	CA1	0.01052	-0.06993
N2	-0.52449	-0.66806	H(CA1)	0.11362	0.13086
CA2	-0.07839	0.31875	CB1	0.51998	0.51781
C2	0.41522	0.36771	H(CB1)	-0.05811	-0.04499
O2	-0.61502	-0.59539	CG1	-0.16067	-0.19992
N3	0.04186	0.04186	H(CG1)(×3)	0.03582	0.04634
<i>somme</i>	<i>-0.44</i>	<i>-0.24</i>	OG1	-0.71063	-0.71024
<i>différence S₁ – S₀</i>		<i>+0.20</i>	H(OG1)	0.39301	0.40500
			N1	-0.79897	-0.76867
			H(N1)	0.38194	0.38937
<i>Pont reliant les deux cycles</i>			<i>somme</i>	<i>-0.20</i>	<i>-0.21</i>
CB2	-0.08643	-0.58097	<i>différence S₁ – S₀</i>		<i>-0.01</i>
HB2	0.13608	0.17717			
<i>somme</i>	<i>0.05</i>	<i>-0.40</i>	<i>Partie Gly67</i>		
<i>différence S₁ – S₀</i>		<i>-0.45</i>	CA3	-0.44694	-0.42387
			H(CA3)(×2)	0.20306	0.19896
<i>Phénolate</i>			C3	0.75080	0.75055
CG2	-0.01568	0.43039	O3	-0.64036	-0.64620
CD1	-0.03297	-0.32853	<i>somme</i>	<i>0.07</i>	<i>0.08</i>
HD1	0.18027	0.25163	<i>différence S₁ – S₀</i>		<i>0.01</i>
CE1	-0.59636	-0.46433			
HE1	0.18533	0.17238			
CZ	0.92479	0.81694			
OH	-0.79480	-0.77154			
CE2	-0.53607	-0.37815			
HE2	0.17566	0.16082			
CD2	-0.13400	-0.30281			
HD2	0.16451	0.18365			
<i>somme</i>	<i>-0.48</i>	<i>-0.23</i>			
<i>différence S₁ – S₀</i>		<i>+0.25</i>			

Tableau I.1 – Charges effectives, en unité atomique, obtenues pour le chromophore de la GFP dans son état fondamental, S₀, et dans son premier état excité fluorescent, S₁. La numérotation des atomes est présentée sur la figure I.6 page 20.

il y a un transfert de charges d'environ 0.5 électron des atomes des cycles imidazolinone et phénolate vers l'atome de carbone du pont qui relie ces cycles.

Les différences des charges effectives des paires d'atomes CD1-CD2, CE1-CE2, HD1-HD2 et HE1-HE2 qui sont localement symétrique par rapport à l'axe de symétrie du phénol, n'excèdent pas 0.08 électron. La symétrie locale du phénolate est donc conservée.

Le cycle phénolate seul possède un axe de symétrie (passant par les atomes CG2 et CZ). Cette symétrie pourrait justifier d'imposer que les charges des paires d'atomes CD1-CD2, CE1-CE2, HD1-HD2, HE1-HE2 soient égales. De telles contraintes n'ont pas été introduites car bien que le chromophore présente localement cette symétrie, la présence du cycle imidazolinone et des parties issues de Thr65 et Gly67 brise cette symétrie. Les

Lorsqu'on s'intéresse à des mouvements de grandes amplitudes, comme dans notre cas le long de la coordonnée φ , des transferts de charge importants peuvent avoir lieu. Dans cette étude, nous avons négligé cette possibilité et choisi d'étudier la dynamique de l'état excité du chromophore en utilisant des charges effectives fixes, c'est à dire indépendante de la géométrie. Le développement de champs de force différents permettant d'utiliser des charges variables est en projet.

c) Chromophore de la Cerulean

Les charges effectives ont été calculées pour les conformations *cis* et *trans* du chromophore à l'état électronique fondamental. Nous avons tout d'abord partiellement optimisé la géométrie du chromophore au niveau B3LYP/6-31G*, en fixant l'orientation des acides aminés Thr65 et Gly67 à leur position dans la structure RX. Pour conserver une approche comparable à celle utilisée pour déterminer les charges du champ de force *ff99*, les charges effectives ont été ajustées de sorte qu'elles reproduisent au mieux les valeurs du potentiel électrostatique créées par la densité électronique du chromophore, calculées au niveau HF/6-31G*.

Nous avons imposé un premier groupe de contraintes linéaires :

1. La somme des charges des atomes ajoutés pour fermer l'entité chromophore doit être nulle (atomes en bleu sur la figure I.6 page 20).
2. Dans le champ de force *ff99*, les atomes du groupement carbonyle (C et O) et du groupement NH du squelette protéique ont la même charge pour tous les acides aminés neutres. Pour le chromophore, les charges de ces quatre atomes ont donc été fixées à leurs valeurs dans le champ de force *ff99* (tableau I.2). Cette contrainte n'a pas été utilisée dans le cas de la GFP car le chromophore est à l'état électronique excité.
3. Les charges effectives des trois atomes d'hydrogène du groupement méthyle sont égales.

atome	C	O	N	H
charge	0.5973	-0.5679	-0.4157	0.2719

Tableau I.2 – Charges effectives, en ua (fraction de la charge d'un électron), portées par les atomes de la fonction amide d'un acide aminé neutre dans le champ de force *ff99*.

Ces contraintes dégradent faiblement la valeur du RRMS : il vaut maintenant 8% alors qu'il valait 6% lorsque seule la contrainte de neutralité totale du chromophore était imposée.

Nous avons imposé les contraintes harmoniques suivantes :

1. Des contraintes centrées sur zéro pour les charges des atomes lourds des cycles imidazolinone et indole du chromophore. Les atomes d'hydrogène n'ont pas été contraints.
2. Des contraintes sur les charges effectives des atomes issus de Thr65, avec comme charges cibles, les valeurs qu'elles ont dans le champ de force *ff99*. En effet il semble acceptable de supposer que la densité électronique des atomes de cette partie soit peu influencée, à l'état fondamental, par la cyclisation des acides aminés qui forme le chromophore.

Le choix des valeurs des constantes k_i que nous avons utilisées résulte d'un compromis : on souhaite que le RRMS soit peu dégradé par l'ajout des contraintes et que les charges obtenues soient suffisamment proches des valeurs cibles. Nous avons retenu deux valeurs de k_i : une pour contraindre les charges des atomes de Thr65 ($k_i = 0.05 \text{ B}^{-2}$) et l'autre pour contraindre la valeur absolue des charges des atomes lourds des deux cycles ($k_i = 0.0005 \text{ B}^{-2}$). Une fois ces contraintes introduites, le RRMS est de 9% dans la conformation *trans* et de 8% dans la conformation *cis*, ce qui est acceptable. Les charges ainsi obtenues sont présentées dans le tableau I.3.

On peut noter que le RRMS est significativement plus grand que celui obtenu pour le chromophore de la GFP (de l'ordre de 3%). La raison est que le chromophore de la GFP étant anionique les valeurs absolues du potentiel électrostatique qu'il crée en chaque point de la grille sont plus grandes que celles du potentiel électrostatique créées par le chromophore de la Cerulean qui est neutre. Or, les valeurs du RMS correspondant aux répartitions de charges obtenues pour le chromophore de la Cerulean ou de la GFP sont similaires et de l'ordre de 0.002 ua. Dans l'équation I.6, dans le cas du chromophore de la GFP, le nombre au dénominateur étant plus grand, le RRMS est donc plus petit.

Nous avons vérifié que les charges obtenues conduisent à des valeurs du moment dipolaire des liaisons polaires N9-H9 et O24-H24 similaires à celles du champ de force *ff99*. Ceci est important car ces atomes sont susceptible de former des liaisons hydrogène avec les acides aminés voisins. Une modification trop importante de leur charge pourrait avoir des effets néfastes sur la stabilité des liaisons hydrogène et en conséquence sur la dynamique du chromophore.

Le changement de conformation n'induit pas de transfert de charge important entre les différentes parties du chromophore. La charge totale de la partie issue de Gly67 augmente de 0.14 ua lors du passage de la conformation *trans* à la conformation *cis* alors que la charge totale du cycle imidazolinone, de la partie Thr65 et des atomes C11, H11 diminue de la même quantité. A l'intérieur des cycles, les charges des atomes dont l'environnement est directement affecté par le changement de conformation sont fortement modifiées. Il s'agit des atomes C14, N13 et C12 pour le cycle imidazolinone et des atomes C1, C2, C3, C7 et C8 pour le tryptophane. Lors du passage de la conformation *trans* à *cis*, la charge de l'atome C2 augmente alors que la charge de l'atome C8 diminue. Cette observation est cohérente avec le fait que l'atome C2 est proche du doublet non liant

atome	<i>trans</i>	<i>cis</i>	<i>ff99</i>	atome	<i>trans</i>	<i>cis</i>	<i>ff99</i>
<i>Imidazolinone</i>				<i>Partie Thr65</i>			
N15	0.08538	-0.31807		N26	-0.41570	-0.41570	<i>-0.4157</i>
C16	0.35820	0.58114		H26	0.27190	0.27190	<i>0.2719</i>
O17	-0.51477	-0.55726		C22	-0.03978	-0.03273	<i>-0.0389</i>
C14	0.10406	0.29637		H22	0.12775	0.09196	<i>0.1007</i>
N13	-0.45181	-0.21691		C23	0.39126	0.35661	<i>0.3654</i>
C12	0.11566	-0.10130		H23	-0.00726	-0.02332	<i>0.0043</i>
<i>somme</i>	<i>-0.30328</i>	<i>-0.31603</i>		C25	-0.25322	-0.24863	<i>-0.2438</i>
<i>Pont reliant les deux cycles</i>				H25	0.06909	0.06948	<i>0.0642</i>
C11	-0.10649	-0.13672		H27	0.06909	0.06948	<i>0.0642</i>
H11	0.15347	0.14281		H27	0.06909	0.06948	<i>0.0642</i>
<i>somme</i>	<i>0.04698</i>	<i>0.00609</i>		O24	-0.57699	-0.64716	<i>-0.6761</i>
<i>Indole</i>				H24	0.36742	0.41580	<i>0.4102</i>
				<i>somme</i>	<i>0.07265</i>	<i>-0.02283</i>	
				<i>Tryptophane</i>			
C7	-0.15501	0.03980	<i>-0.1415</i>	<i>Partie Gly67</i>			
C8	0.04407	-0.14575	<i>-0.1638</i>	C18	-0.29518	0.00444	<i>-0.0252</i>
H8	0.12827	0.21142	<i>0.2062</i>	H18	0.16620	0.10397	<i>0.0698</i>
N9	-0.40727	-0.36993	<i>-0.3418</i>	H19	0.16620	0.07353	<i>0.0698</i>
H9	0.37511	0.36616	<i>0.3412</i>	C19	0.59730	0.59730	<i>0.5973</i>
C6	0.12979	0.14055	<i>0.1380</i>	O21	-0.56790	-0.56790	<i>-0.5679</i>
C5	-0.26222	-0.25499	<i>-0.2601</i>	<i>somme</i>	<i>0.06662</i>	<i>0.21134</i>	
H5	0.17076	0.16315	<i>0.1572</i>	<i>dipôle en Debye</i>			
C4	-0.12565	-0.13566	<i>-0.1134</i>	<i>charges</i>	9.914	9.319	
H4	0.14248	0.14880	<i>0.1417</i>	RHF	9.964	9.298	
C3	-0.16315	-0.23503	<i>-0.1972</i>	<i>RRMS</i>			
H3	0.14727	0.15561	<i>0.1447</i>		0.09	0.08	
C2	-0.26883	-0.06659	<i>-0.2387</i>				
H2	0.18665	0.05953	<i>0.1700</i>				
C1	0.17474	0.04436	<i>0.1243</i>				
<i>somme</i>	<i>0.11701</i>	<i>0.12143</i>					

Tableau I.3 – Charges effectives, en unité atomique, obtenues pour le chromophore de la Cerulean en conformation *trans* et en conformation *cis*. La valeur du dipôle de la ligne *charges* correspond au dipôle créé par les charges effectives. La valeur de la ligne RHF correspond au dipôle calculé au niveau RHF/6-31G*. Colonnes *ff99* : valeurs des charges correspondantes dans le champ de force *ff99*.

de l'azote N13 dans la conformation *cis* alors que c'est le carbone C8 qui prend sa place dans la conformation *trans*.

Les charges des atomes N15 et C18 présentent également des variations importantes, d'amplitudes semblables et de signes opposés. Ces variations ne semblent pas avoir de signification physique, et sont vraisemblablement dues à une instabilité de l'algorithme de minimisation car C18 est un atome "enfoui".

I.3.3 Champs de force pour le chromophore de la Cerulean

I.3.3.1 Interactions entre atomes non liés

Les interactions entre atomes non liés sont de deux types : les interactions de van der Waals, et les interactions électrostatiques. Les interactions électrostatiques sont modélisées au travers des charges effectives dont le calcul a été présenté au paragraphe I.3.2.2. Nous avons pris les valeurs des paramètres des interactions de van der Waals dans le champ de force *ff99*. Elles découlent du "type d'atome" déclaré pour chaque atome. C'est du type d'atome que dépendent également les paramètres des interactions entre atomes liés. Ces types sont décrits dans ce qui suit.

I.3.3.2 Interactions entre atomes liés

Pour construire le champ de force du chromophore, il faut donc définir les types des atomes qu'il contient et, si nécessaire, définir de nouveaux types et tous les paramètres qui leurs sont associés.

a) Type d'atome dans le champ de force *ff99*

Le type d'un atome donné est relié à ses propriétés chimiques. Il est attribué en fonction du nombre et du type de liaisons qu'il forme. La théorie de l'hybridation classant les atomes de type sp^3 (l'atome forme 4 liaisons simples et adopte une géométrie tétraédrique), sp^2 (l'atome forme 2 liaisons simples et une liaison double toutes dans un même plan) ou sp^1 (l'atome forme une liaison simple et une liaison triple colinéaire), est généralement utilisée. Par exemple, le type d'un carbone d'une chaîne aliphatique, sp^3 , sera différent de celui d'un carbone d'un système π , sp^2 . Ensuite, pour une hybridation donnée, le type d'atome dépend de la fonction chimique à laquelle il appartient et donc des atomes auxquels il est lié. Le type d'un carbone d'un groupement carbonyle est différent du type d'un atome de carbone d'un cycle indole, bien qu'ils appartiennent tous les deux à un système π .

Le chromophore étant constitué de trois acides aminés, un type d'atome peut être attribué pour la plupart des atomes du chromophore par analogie avec ceux du champ de force *ff99*. Ainsi, pour les atomes issus de Thr65 et Gly67, nous avons attribué le même type d'atome que dans les acides aminés thréonine et glycine. De la même manière, pour les atomes du cycle indole, nous avons attribué les types des atomes du cycle indole du tryptophane. Le type O, de l'oxygène du carbonyle du squelette protéique a été attribué à l'oxygène O17 de l'imidazolinone. Le type HA qui correspond

à un hydrogène de groupements aromatiques sans groupements électro-attracteurs a été attribué à l'hydrogène H11.

Par contre, le type de l'atome de carbone C11 reliant les deux cycles et des atomes C12, N13, C14, N15 et C16 du cycle imidazolinone, ne peuvent pas être choisis parmi ceux disponibles dans le champ de force *ff99* car aucun acide aminé ne présente ce type de topologie. De nouveaux types doivent donc être définis.

b) Définitions de nouveaux types d'atome

Pour définir un nouveau type d'atome, il faut préciser de quel élément chimique il s'agit et donner tous les paramètres du champ de force (à l'exception des charges atomiques, qui sont définies atome par atome) dans lesquels il intervient : les paramètres des interactions de van der Waals, les constantes de force et les valeurs d'équilibre pour toutes les élongations, les pliages et les torsions.

Les paramètres de van der Waals ont été pris égaux à ceux du même élément chimique avec le même nombre de liaisons dans le champ de force *ff99*. Les valeurs d'équilibres des liaisons, des angles de valence et des angles dièdres ont été déduites directement de l'optimisation de géométrie *ab initio* du chromophore. Pour les atomes du cycle imidazolinone, C12, N13, C14, N15 et C16, toutes les constantes de forces ont été choisies dans le champ de force *ff99* par analogie avec d'autres types d'atome existants. Le cycle imidazolinone étant similaire au cycle d'un acide aminé histidine, les paramètres ont été pris parmi ceux des types d'atome d'une histidine. De la même manière, les paramètres faisant intervenir l'oxygène du groupement carbonyle de l'imidazolinone ont été pris parmi ceux faisant intervenir celui du squelette protéique.

c) Cas particulier des atomes reliant les deux cycles

Une attention particulière a été donnée aux paramètres des atomes C7, C11 et C12 qui relient les deux cycles aromatiques car ils sont particulièrement importants pour décrire la dynamique du chromophore.

A l'état fondamental, la liaison C11-C7 est une liaison simple tandis que la liaison C11-C12 est une liaison double. Les constantes d'élongation ont été choisies de sorte qu'elles reproduisent ce caractère de simple ou double liaison. Ces trois atomes étant impliqués dans le système π reliant les deux cycles, la constante de force d'élongation correspondant à une liaison simple entre un atome sp^2 relié à un atome de carbone d'un cycle aromatique a été attribuée à la liaison C11-C7. Cette constante de force est donnée dans le champ de force *gaff* [47] (*General Amber force field*) qui est un champ de force similaire à *ff99* mais qui contient un plus large choix de paramètres, adaptés aux molécules organiques en général. La constante de force d'une liaison double entre deux atomes de carbone sp^2 d'un cycle aromatique a été attribuée à la liaison C11-C12 et a été prise dans le champ de force *ff99*.

Dans le champ de force *ff99*, la constante de force liée à un pliage entre trois atomes sp^2 comme C12, C11 et C7 est la même quelque soit le type d'atome et a été attribuée au pliage entre ces trois atomes.

Dans le champ de force *ff99* l'expression analytique de l'énergie potentielle de torsion pour un angle de torsion ϕ quelconque est la suivante :

$$E_{diedre}(\phi) = \sum_n \frac{V_n}{k} [1 + \cos(n\phi - \phi_0)] \quad (\text{I.8})$$

La valeur de k est le nombre d'angles dièdres associés à la torsion et vaut 4 dans le cas d'atomes impliqués dans un système π . Plusieurs paramètres doivent être déterminés pour chaque torsion : Les constantes de force, V_n qui définissent la raideur de la torsion, n la périodicité et ϕ_0 la valeur d'équilibre. Pour cette étude, comme nous souhaitons simplement décrire la surface d'énergie potentielle du chromophore au voisinage de ses deux positions d'équilibre, l'utilisation de plusieurs termes de périodicités différentes est inutile. Nous avons donc pris simplement le terme $n=1$.

Nous avons déterminé ces paramètres pour les deux torsions, φ et τ , autour des deux liaisons C7-C11 et C11-C12, qui relient les deux cycles aromatiques (figure I.6 page 20). La torsion φ sera caractérisé par l'angle dièdre C12-C11-C7-C8 et l'angle τ sera caractérisé par l'angle dièdre N13-C12-C11-C7. Dans le cas présent les valeurs à l'équilibre, φ_0 et τ_0 , des deux angles de torsion φ et τ valent 0° ou 180° car le chromophore est plan. Il reste donc à déterminer les constantes de force de torsion V_φ et V_τ associées aux angles de torsion φ et τ dans chacune des deux conformations *cis* ($\varphi=0^\circ$ et $\tau=0^\circ$) et *trans* ($\varphi=180^\circ$ et $\tau=0^\circ$).

Ces constantes ont été déterminées en ajustant un ensemble d'énergies du chromophore calculées avec le champ de force *ff99* sur un ensemble d'énergies calculées par une méthode *ab initio*. Par exemple, pour la détermination de V_φ en conformation *cis*, nous avons tout d'abord fait des calculs *ab initio* du chromophore en conformation *cis* correspondant à différentes valeurs de φ , de -30° à $+30^\circ$ par pas de 10° . Pour chacun de ces calculs, nous avons imposé la valeur de φ , et optimisé la géométrie (au niveau B3LYP/6-31G*) en imposant que τ conserve sa valeur à l'équilibre et que les angles dièdres des parties Thr65 et Gly67 conservent leurs valeurs dans la structure RX. Nous obtenons ainsi une courbe $E_{ab\ initio}(\varphi)$ (figure I.8).

Ensuite, pour le même ensemble de valeurs de φ , nous avons calculé, pour plusieurs valeurs de la constante de force V_φ , l'énergie totale du chromophore avec l'expression analytique du champ de force *ff99* :

$$E_{ff99} = E_{liaison} + E_{angle} + E_{diedre} + E_{vdW} + E_{elec} \quad (\text{I.9})$$

Nous avons ensuite cherché la valeur de cette constante pour laquelle les variations de la courbe E_{ff99} reproduisent au mieux celles de la courbe $E_{ab\ initio}$.

Les géométries du chromophore utilisées dans les calculs de E_{ff99} ne sont pas identiques à celles des calculs *ab initio*. Ces dernières sont partiellement optimisées pour chaque valeur de l'angle φ : les longueurs de liaisons et les angles varient en fonction de φ . Par contre, les énergies E_{ff99} sont calculées pour des géométries du chromophore où seul φ varie, les longueurs de liaison et les angles de liaison restant fixés à leurs valeurs optimisées *ab initio*. Cette façon de faire découle du fait que dans la dynamique les coordonnées interne ne sont pas couplées. Donc les deux premiers termes de

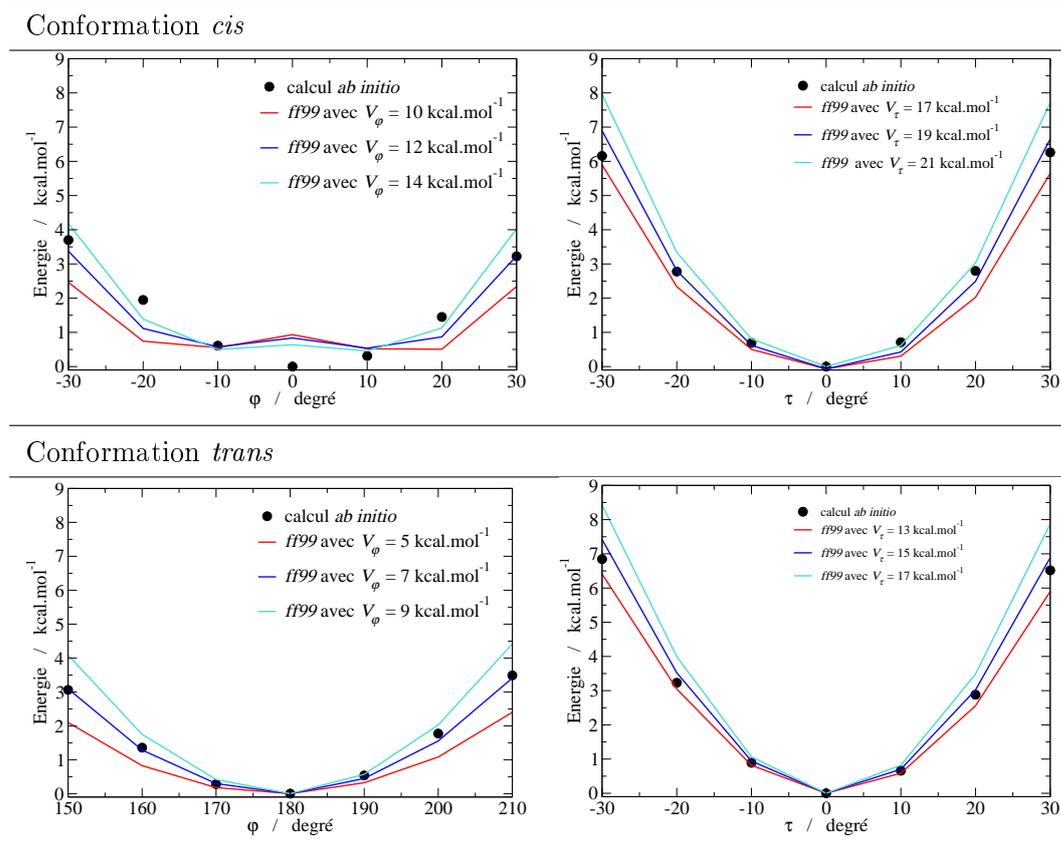


Figure I.8 – Courbes d'énergie potentielle des torsions autour des liaisons C12-C11 (angle dièdre τ) et C11-C7 (angle dièdre φ) pour les deux conformations *cis* et *trans* du chromophore. Les calculs *ab initio* ont été fait au niveau B3LYP/6-31G* sur des géométries partiellement optimisées. *ff99* : valeurs calculées avec le champ de force *ff99*.

E_{ff99} (énergie d'élongation et de pliage) ne dépendent pas de l'angle φ , mais le terme de van der Waals et le terme électrostatique en dépendent.

Nous avons procédé de cette façon pour les constantes de force V_φ et V_τ pour les conformations *cis* et *trans*. Les résultats sont présentés dans le tableau I.4.

La constante de force de la torsion φ est systématiquement plus faible que celle de la torsion τ , ce qui traduit le fait que l'indice de liaison de C11-C7 est plus faible que celui de C11-C12 (tableau I.4). Ce résultat est cohérent avec les caractères de liaison simple et de liaison double de ces liaisons dans la structure de Lewis du chromophore à l'état électronique fondamental (figure I.6 page 20). On remarque sur la figure I.8 que dans la conformation *cis*, pour l'angle φ , la courbe E_{ff99} présente un faible maximum local en 0° . Il est dû à la partie répulsive des forces de van der Waals entre l'atome H2 du cycle indole et l'atome N13 du cycle imidazolinone. Il n'est pas présent au niveau *ab initio* car, dans ce dernier, la géométrie a été optimisée et cette interaction a été relaxée. Ce maximum local de E_{ff99} n'est pas observé dans la conformation *cis* pour l'angle de torsion τ car la constante de force de cette torsion est plus forte et la variation d'énergie liée à la torsion autour de cette liaison est plus importante que le gain d'énergie dû à la relaxation de l'interaction entre H2 et N13. Pour la conformation *trans*, il n'y a pas de maximum en 0° ni pour l'angle φ ni pour l'angle

Conformation	V_φ (kcal · mol ⁻¹)	V_τ (kcal · mol ⁻¹)	φ_0	τ_0
<i>trans</i>	7.	15.	0°	0°
<i>cis</i>	12.	19	180°	0°

Tableau I.4 – Paramètres de l'énergie potentielle de torsion autour des liaisons C12-C11 (angle dièdre φ) et C11-C7 (angle dièdre τ) pour les conformations *cis* et *trans* du chromophore. L'angle φ est défini comme l'angle dièdre entre les atomes C12-C11-C7-C8 et l'angle τ est défini comme l'angle dièdre entre les atomes N13-C12-C11-C7.

τ car il n'y a pas de terme de van der Waals fortement répulsif dans la géométrie plane.

Le champ de force du chromophore de la Cerulean est maintenant complet pour les deux conformations *cis* et *trans* et peut être utilisé dans une simulation de dynamique moléculaire.

I.3.4 Champs de force pour le chromophore de la GFP

Nous avons déterminé un champ de force pour le chromophore anionique de la GFP dans son état excité fluorescent en utilisant une approche similaire à celle employée pour la Cerulean. Mais, contrairement à la Cerulean, dans le cas de la GFP, nous ne nous sommes pas limités à l'étude de la dynamique du chromophore de la GFP autour de sa position d'équilibre. Nous avons étudié la dynamique de l'état excité du chromophore le long de la coordonnée φ . Ceci a nécessité de connaître de manière plus complète la courbe d'énergie potentielle dans l'état excité fluorescent du chromophore anionique en fonction de cette coordonnée.

I.3.4.1 Interactions entre atomes non liés

Les charges effectives utilisées pour décrire les interactions électrostatiques ont été déterminées séparément et sont présentées au paragraphe I.3.2.2. Les valeurs des paramètres des interactions de van der Waals ont été prises dans le champ de force *ff99*. Nous avons fait le choix de ne pas modifier les valeurs des rayons de van der Waals pour plusieurs raisons. Tout d'abord, nous avons fait ce choix pour conserver une homogénéité dans les interactions de van der Waals entre le chromophore et les acides aminés de la protéine et celles entre les acides aminés standard. D'autre part, dans notre cas, il est acceptable de supposer que le passage à l'état excité fluorescent n'induit pas une variation importante du rayon de van der Waals. En effet, le chromophore étant confiné à l'intérieur de la protéine, l'extension du nuage électronique lors du passage de l'état fondamental à l'état excité est limitée. Cette hypothèse est également renforcée par le fait que l'état excité fluorescent du chromophore de la GFP est de type $\pi\pi^*$ et consiste principalement à peupler l'orbitale LUMO qui a une expansion peu différente de la HOMO (figure I.4 page 16).

I.3.4.2 Interactions entre atomes liés

Les observations détaillées dans le cas du chromophore de la Cerulean sont directement transposables dans le cas du chromophore de la GFP. De nouveaux types d'atome ont été définis pour les atomes du cycle imidazolinone C1, C2, CA2, N2, N3 et pour le carbone CB2 du pont reliant les deux cycles aromatiques. Les positions d'équilibre des coordonnées internes impliquant ces atomes ont été extraites de la géométrie de l'état excité partiellement optimisée au niveau CIS. Concernant les constantes de force des coordonnées internes faisant intervenir ces atomes, nous avons utilisé les mêmes que dans le champ de force du chromophore de la Cerulean (paragraphe I.3.3.2). Elles sont déduites par analogie avec les types d'atome du champ de force *ff99*, exceptées les constantes de force des torsions φ et τ et des pyramidalisations des atomes de carbone CA2 et CG2 notée pyr_{CA2} et pyr_{CG2} ^a.

Dans l'étude de la GFP, nous nous sommes intéressés à la rotation du phénolate autour de son axe CG2-CZ. Ce mouvement peut être décrit comme une rotation d'angle φ , les atomes de carbone CA2 et CG2 restant plans. Il est donc nécessaire de choisir de façon appropriée les constantes de forces associées à ces angles. Nous avons ajouté des contraintes harmoniques sur les angles dièdres caractérisant la torsion τ et les angles dièdres impropres caractérisant les pyramidalisations pyr_{CA2} et pyr_{CG2} de la forme $k(\phi - \phi_0)^2$ où ϕ représente un angle dièdre quelconque. Une constante de force de $k=40 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ a été utilisée pour chaque coordonnée. Le tableau I.5 rassemble les paramètres des contraintes harmoniques utilisées ainsi que les angles dièdres impliqués.

Coordonnée	τ		PYR_{CA2}	PYR_{CG2}
Angle dièdre	N2-CA2-CB2-CG2	C2-CA2-CB2-CG2	N2-C2-CA2-CB2	CD1-CD2-CG2-CB2
k kcal · mol ⁻¹ · rad ⁻²	20	20	40	40
ϕ_0	0°	180°	0°	0°

Tableau I.5 – Valeurs de la constante de force k et de la position du minimum ϕ_0 de la contrainte harmonique utilisée, de la forme $k(\phi - \phi_0)^2$, pour les coordonnées τ , pyr_{CA2} et pyr_{CG2} . L'angle de torsion τ est caractérisé par les angles dièdres N2-CA2-CB2-CG2 et C2-CA2-CB2-CG2. Les pyramidalisations pyr_{CA2} et pyr_{CG2} sont données par les angles dièdres impropres N2-C2-CA2-CB2 et CD1-CD2-CG2-CB2.

Les paramètres de l'énergie de torsion en fonction de φ ont été ajustés de sorte que l'énergie de type *ff99* reproduise au mieux l'énergie du premier état électronique excité, calculée au niveau CASPT2 par Toniolo *et al* [48]. Ces calculs ont constitué une référence pour ce travail.

Les charges effectives, les paramètres van der Waals, les valeurs d'équilibre et les paramètres des interactions entre atomes liés ayant été choisis, il reste à déterminer la forme et les paramètres du potentiel de torsion φ . Nous les avons déterminé de façon à ce que l'énergie totale du chromophore calculée avec l'expression analytique du champ de force *ff99* reproduise les variations d'énergie

^aPour simplifier la lecture des résultats on prendra comme convention qu'un angle de pyramidalisation de 0° correspond à une géométrie plane.

obtenues au niveau CASPT2 par Toniolo *et al.* L'expression suivante, avec deux fonctions cosinus, de périodicité π et $\pi/2$ (qui correspondent aux termes $n=2$ et $n=4$ de l'équation I.8) est bien adaptée.

$$E_{dièdre}(\varphi) = \frac{V_2}{k}(1 + \cos 2\varphi) + \frac{V_4}{k}(1 + \cos 4\varphi)$$

Ici la constante k vaut 4 comme dans le cas de la Cerulean. Nous avons obtenu les valeurs $V_2 = 2.3 \text{ kcal} \cdot \text{mol}^{-1}$ et $V_4 = -2.5 \text{ kcal} \cdot \text{mol}^{-1}$. La figure I.9 présente la comparaison entre la courbe d'énergie potentielle de l'état électronique excité en fonction de φ obtenue par Toniolo *et al* et celle obtenue avec les paramètres que nous avons déterminés. L'écart quadratique moyen entre ces deux courbes est de $1.2 \text{ kcal} \cdot \text{mol}^{-1}$ et conduit à un RRMS de 6%. Ceci est de l'ordre de grandeur de la précision des calculs d'énergie *ab initio* d'états électroniques excités [49].

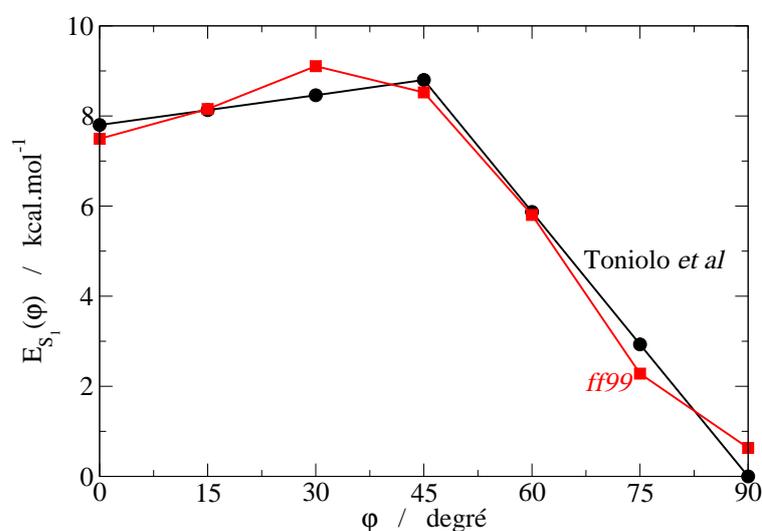


Figure I.9 – Comparaison entre la courbe d'énergie potentielle de l'état excité du chromophore calculée par Toniolo *et al* et celle calculée avec le champ de force *ff99* après ajustement des paramètres associés à la torsion φ .

Ce champ de force de l'état électronique excité fluorescent du chromophore anionique de la GFP sera utilisé au chapitre III, lors de l'étude de l'effet de la dynamique de la torsion φ sur le déclin de la fluorescence de la GFP. L'utilisation de ce champ de force, pour décrire l'état électronique excité du chromophore de la GFP, implique de négliger plusieurs effets. Entre autre, l'effet de l'environnement sur la surface d'énergie potentielle, le transfert de charges dû à la rotation selon la coordonnée φ et la présence d'autres états électroniques excités ne sont pas pris en compte. Bien qu'il correspondent au niveau de description de l'état excité le plus simple, nous verrons au chapitre III qu'il permet d'obtenir des résultats en bon accord avec les observations expérimentales.

I.4 Calcul d'énergie libre – *Umbrella Sampling*

La plupart des propriétés physico-chimiques ou biochimiques d'un système peuvent être interprétées directement ou indirectement à partir des variations de l'énergie libre ou de l'enthalpie libre

du système qui, de ce fait, sont des quantités très importantes. Par exemple, les préférences conformationnelles d'une molécule ou d'une protéine, la solvatation, les constantes d'association ou de dissociation de complexes sont directement reliées à la différence d'énergie libre entre deux états.

Après avoir défini plus précisément l'énergie libre d'un système, la technique du *umbrella sampling* [50, 51], permettant de calculer un profil d'énergie libre en fonction d'une coordonnée de réaction, sera présentée.

I.4.1 Considérations générales

L'énergie libre est une fonction d'état, qui dépend du nombre de particules, N , de la température, T et du volume, V , du système. Lorsqu'on s'intéresse à des systèmes en phase condensée qui sont peu compressibles, comme une protéine en phase aqueuse, l'énergie libre et l'enthalpie libre peuvent être confondues. Par la suite, nous utiliserons uniquement l'énergie libre qui sera représentée par la lettre \mathcal{F} que l'on se trouve dans l'ensemble canonique ou dans l'ensemble (N, P, T) . Dans l'ensemble canonique, l'énergie libre est définie à partir de la fonction de partition \mathcal{Q} du système par :

$$\mathcal{F} = -k_B T \ln \mathcal{Q} \tag{I.10}$$

où T est la température du système et k_B la constante de Boltzmann.

L'énergie libre est une propriété statistique. Elle est directement reliée à la probabilité de trouver le système dans un état donné et au volume de l'espace des phases accessible au système. Pour obtenir une bonne estimation de l'énergie libre absolue, il serait donc nécessaire d'échantillonner la totalité de l'espace des phases et de calculer la fonction de partition du système. Ceci est impossible car le nombre d'états d'un système est infini alors que toute simulation ou expérience a une durée finie et conduit donc à un échantillonnage limité.

La différence d'énergie libre entre deux états du système est plus facile à calculer. Elle correspond au rapport des probabilités de trouver le système dans deux états que l'on considère. De nombreuses techniques de simulations permettant de calculer des différences d'énergie libre existent : l'intégration thermodynamique, la perturbation thermodynamique ou encore le *umbrella sampling* que nous avons utilisé pour réaliser ce travail et que nous allons détailler.

Lorsqu'on s'intéresse à une réaction ou un processus physico-chimique associé à une coordonnée de réaction, q , il est utile de s'intéresser à la probabilité d'obtenir des configurations dans lesquelles la coordonnée q prend des valeurs proches d'une valeur cible q_c . La coordonnée q peut être, par exemple, une distance entre deux molécules, un angle de valence, un angle dièdre d'une molécule ou encore la longueur d'un polymère ou d'une structure en hélice. La densité de probabilité d'une

coordonnée q est donnée par la relation :

$$\rho(q) = \frac{\int \delta(q - q(r^N)) \exp\left(-\frac{\mathcal{U}(r^N)}{k_B T}\right) dr^N}{\int \exp\left(-\frac{\mathcal{U}(r^N)}{k_B T}\right) dr^N} \quad (\text{I.11})$$

$$\rho(q) = \frac{\mathcal{Q}^p(q)}{\mathcal{Q}}$$

où le dénominateur, \mathcal{Q} , est la fonction de partition du système et $\mathcal{U}(r^N)$ est l'énergie potentielle du système pour une configuration définie par la donnée des N coordonnées du système notées r^N . Le terme $\mathcal{Q}^p(q)$ est, en fait, la fonction de partition qu'aurait le système si on supprimait le degré de liberté correspondant à la coordonnée q en la fixant à une valeur donnée pour toutes les configurations.

La densité de probabilité peut alors s'écrire :

$$\rho(q) = \exp\left(-\frac{\mathcal{F}(q) - \mathcal{F}}{k_B T}\right) \quad (\text{I.12})$$

où \mathcal{F} est l'énergie libre du système et $\mathcal{F}(q)$ est l'énergie libre partielle en fonction de la coordonnée q , appelée aussi énergie libre de Landau. Elle est définie par :

$$\mathcal{F}(q) = -k_B T \ln \mathcal{Q}_p(q) \quad (\text{I.13})$$

L'énergie libre partielle est donc une fonction de la coordonnée de réaction qui est directement reliée à la probabilité d'échantillonner une configuration dans laquelle la coordonnée q prend une valeur donnée. En pratique, l'équation I.12 est utilisée pour calculer l'énergie libre. On la calcule à une constante près à partir de la densité de probabilité $\rho(q)$ en utilisant la relation :

$$\mathcal{F}(q) = -k_B T \ln \rho(q) + C^{ste} \quad (\text{I.14})$$

où la constante, égale à $-\ln \mathcal{Q}$, est indépendante de q . La densité de probabilité $\rho(q)$ peut alors être obtenue par simulation en calculant l'histogramme de cette coordonnée au cours d'une simulation.

I.4.2 Principe du *umbrella sampling*

Il est fréquent qu'une simulation moléculaire ne permette pas d'échantillonner l'ensemble du domaine de variation de la coordonnée de réaction. Le potentiel intra-moléculaire ou l'environnement du système peuvent induire des contraintes telles que seulement une petite partie du domaine de variation de la coordonnée de réaction est échantillonnée. De plus, l'échantillonnage est limité compte tenu du fait que la durée des simulations n'est pas infinie. L'utilisation directe de l'équation I.14 ne permet donc pas d'obtenir, dans tous les cas, la variation d'énergie libre entre la configuration d'équilibre et une configuration d'intérêt.

La technique du *umbrella sampling*, consiste à ajouter à l'énergie potentielle du système $\mathcal{U}(r^N)$ un terme supplémentaire appelé potentiel de biais, V^b , qui dépend de la coordonnée de réaction q que l'on considère. La forme généralement employée est une fonction quadratique :

$$V^b(q) = k(q - q_0)^2$$

où k est une constante de force et q_0 caractérise la position du minimum du potentiel de biais. Les simulations faites en rajoutant le potentiel de biais sont appelées simulations biaisées. Elles permettent, en choisissant les valeurs de k et q_0 , d'échantillonner la région de l'espace des configurations voulue. Une simulation correspondant à une valeur de q_0 donnée est appelée une "fenêtre".

Pour obtenir la densité de probabilité du système, il faut alors retirer l'effet du potentiel de biais sur l'échantillon de configurations extrait des simulations biaisées. Introduisons un potentiel de biais $V^b(r^N)$ dans l'équation I.11

$$\rho(q) = \frac{\int \delta(q - q(r^N)) \exp\left(-\frac{\mathcal{U}(r^N) + V^b(r^N) - V^b(r^N)}{k_B T}\right) dr^N}{\int \exp\left(-\frac{\mathcal{U}(r^N)}{k_B T}\right) dr^N} \quad (\text{I.15})$$

Comme le potentiel de biais ne dépend que de la coordonnée q dont on cherche la densité de probabilité, le terme dépendant du potentiel de biais peut être factorisé de la façon suivante :

$$\rho(q) = \frac{\exp\left(\frac{V^b(q)}{k_B T}\right) \int \delta(q - q(r^N)) \exp\left(-\frac{\mathcal{U}(r^N) + V^b(r^N)}{k_B T}\right) dr^N}{\int \exp\left(-\frac{\mathcal{U}(r^N)}{k_B T}\right) dr^N} \quad (\text{I.16})$$

Soit \mathcal{Q}^b la fonction de partition du système soumis au potentiel $\mathcal{U}(r^N) + V^b(r^N)$.

$$\mathcal{Q}^b = \int \exp\left(-\frac{\mathcal{U}(r^N) + V^b(r^N)}{k_B T}\right) dr^N \quad (\text{I.17})$$

En multipliant par \mathcal{Q}^b le numérateur et le dénominateur de l'équation I.16, on fait apparaître au numérateur la densité de probabilité biaisé de la coordonnée q , notée $\rho^b(q)$.

$$\rho(q) = \rho^b(q) \exp\left(\frac{V^b(q)}{k_B T}\right) \frac{\mathcal{Q}^b}{\mathcal{Q}} \quad (\text{I.18})$$

Cette équation permet donc de calculer la distribution $\rho(q)$ à partir de la distribution biaisée $\rho^b(q)$ obtenu au cours de la simulation biaisée, à un facteur multiplicatif près. Il est important de remarquer que le rapport $\mathcal{Q}^b/\mathcal{Q}$ est une constante indépendante de la coordonnée q . Elle peut être considérée comme un coefficient de normalisation et, dans un premier temps, il est inutile de la calculer. On peut ensuite déduire de l'équation I.18 l'expression de l'énergie libre partielle $\mathcal{F}_1(q)$ en utilisant

l'équation I.14.

$$\begin{aligned}\mathcal{F}_1(q) &= -k_B T \ln \rho^b(q) - V^b(q) + C^{ste} \\ \mathcal{F}_1(q) &= \mathcal{F}^b(q) - V^b(q) + C^{ste}\end{aligned}\tag{I.19}$$

L'indice 1 se rapporte à la simulation biaisée et l'indice 0 se rapportera à la simulation non biaisée. Afin de pouvoir calculer la variation d'énergie libre entre la géométrie d'équilibre et les géométries d'intérêt échantillonnées à l'aide du potentiel de biais, il faut placer la courbe $\mathcal{F}_1(q)$ obtenue à partir de l'équation I.19 par rapport à la courbe $\mathcal{F}_0(q)$ obtenue à partir de la simulation non biaisée.

I.4.3 Construction du profil d'énergie libre

Pour positionner les courbes d'énergie libre partielle $\mathcal{F}_0(q)$ et $\mathcal{F}_1(q)$ l'une par rapport à l'autre, on utilise le fait que l'énergie libre est une fonction d'état et que les densités de probabilité obtenues à partir des simulations moléculaires (simulation de dynamique moléculaire ou simulation Monte Carlo) correspondent à l'état d'équilibre du système. Il en découle qu'une seule valeur d'énergie libre existe pour un état donné du système. Considérons, par exemple, que le système est dans un état défini par le fait que la coordonnée q prend la valeur q_c . Les valeurs $\mathcal{F}_0(q_c)$ et $\mathcal{F}_1(q_c)$ doivent donc être égales. Cela permet de placer les deux courbes $\mathcal{F}_0(q)$ et $\mathcal{F}_1(q)$ l'une par rapport à l'autre et de calculer les différences d'énergie libre voulues.

Cependant, pour pouvoir imposer que les valeurs $\mathcal{F}_0(q_c)$ et $\mathcal{F}_1(q_c)$ soient égales il faut remplir certaines conditions. Tout d'abord, les densités de probabilité $\rho_0(q)$ et $\rho_1^b(q)$ obtenues par la simulation non biaisée et par la simulation biaisée doivent bien entendu se recouvrir afin qu'il existe un intervalle de valeurs de q échantillonné par les deux simulations. En pratique, cette condition est remplie en faisant une succession de simulations biaisées, d'indice i , en faisant varier progressivement la position du minimum du potentiel de biais. Cependant, le fait que deux densités de probabilité biaisées $\rho_i^b(q)$ et $\rho_{i+1}^b(q)$ se recouvrent ne garantit pas que l'état du système caractérisé par une valeur q_c de la coordonnée q est le même dans les deux simulations. En effet, dans les systèmes complexes la surface d'énergie potentielle peut présenter de nombreux minimum locaux. Deux configurations issues des fenêtres i et $i + 1$ peuvent donc correspondre à la même valeur q_c mais à une situation différente pour le système en raison d'une modification de l'environnement. Dans ce cas, il est impossible de positionner les courbes $\mathcal{F}_i(q)$ et $\mathcal{F}_{i+1}(q)$ l'une par rapport à l'autre car quelque soit la valeur de q , elles ne correspondent pas à un même état du système et les deux valeurs de l'énergie libre n'ont aucune raison d'être égales. Ce type de problème peut avoir comme origine, un échantillonnage insuffisant dû au fait que les configurations échantillonnées dans la fenêtre $i + 1$ sont très peu probables dans la fenêtre i . Une solution pour remédier à ce problème peut être d'intercaler une fenêtre supplémentaire entre les fenêtre i et $i + 1$. Il est donc nécessaire de veiller à ce que, d'une fenêtre à l'autre, les zones de l'espace des configurations échantillonnées aient une partie commune. Ceci assure que l'état caractérisé par une valeur q_c de la coordonnée de réaction est bien le même dans les deux fenêtres. Si cette condition est remplie, les courbes $\mathcal{F}_i(q)$ étant définies à une constante

près indépendante de q , elles doivent être parallèles deux à deux dans leur intervalle commun de valeur de q . Un exemple de courbes $\mathcal{F}_i(q)$ est présenté sur la figure I.10 a.

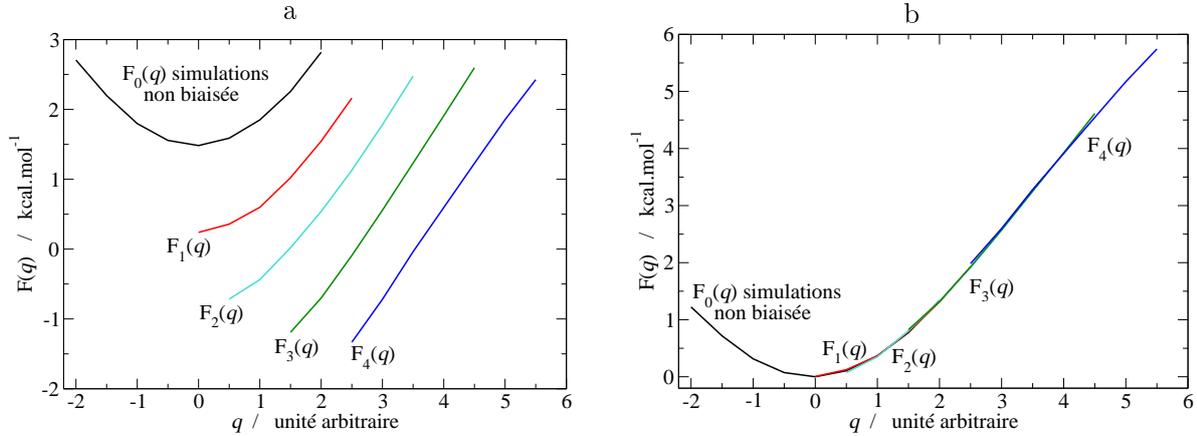


Figure I.10 – Courbes $\mathcal{F}_i(q)$ obtenues à partir d'une simulation sans contraintes et de simulations biaisées.

Pour construire le profil d'énergie libre complet, on doit placer toutes les courbes $\mathcal{F}_i(q)$ les unes par rapport aux autres. Cela revient à les raccorder en calculant les constantes, C_i^{ste} , de sorte que les parties communes des courbes se superposent (figure I.10 b). Par défaut, on impose la valeur de la constante C_0^{ste} telle que $\min[\mathcal{F}_0(q)] = 0$ et on calcule les autres constantes C_{i+1}^{ste} en minimisant la fonction suivante :

$$\chi_{i+1}^2 = \sum_k w_k^i w_k^{i+1} (\mathcal{F}_i(q_k) - \mathcal{F}_{i+1}(q_k) - C_{i+1}^{ste})^2 \quad (\text{I.20})$$

où la somme porte sur les valeurs q_k de la coordonnée de réaction q échantillonnées à la fois dans la fenêtre i et dans la fenêtre $i + 1$ et w_k est une fonction de poids qui sera discutée plus tard.

Le minimum de la fonction χ^2 possède une expression analytique qui est

$$C_{i+1}^{ste} = \frac{\sum_k w_k^i w_k^{i+1} (\mathcal{F}_i(q_k) - \mathcal{F}_{i+1}(q_k))}{\sum_k w_k^i w_k^{i+1}} \quad (\text{I.21})$$

Il s'agit en fait de la moyenne des écarts entre les courbes $\mathcal{F}_i(q)$ et $\mathcal{F}_{i+1}(q)$ pondérées par les poids w_k .

L'introduction de poids dans le calcul des constantes C_i^{ste} permet de sélectionner l'intervalle des valeurs de q à considérer pour chaque simulation biaisée et de pondérer leur importance. Les courbes $\mathcal{F}_i(q)$ sont calculées à partir des densités de probabilité $\rho_i^b(q)$, assimilées à l'histogramme des valeurs de q dans la fenêtre i . L'échantillonnage des configurations est insuffisant aux extrémités des densités de probabilité. Dans ces zones, la densité de probabilité obtenue est donc très incertaine et ne devrait pas être utilisée pour construire le profil d'énergie libre. De plus, les courbes $\mathcal{F}_i(q)$ sont obtenues à partir du logarithme des densités de probabilité. Or, aux extrémités, ces dernières oscillent près

de zéro et la fonction logarithme diverge, Ceci provoque de fortes oscillations des courbes $\mathcal{F}_i(q)$ et rend difficile le calcul des constantes C_i^{ste} . Ce phénomène est illustré sur la figure I.11.

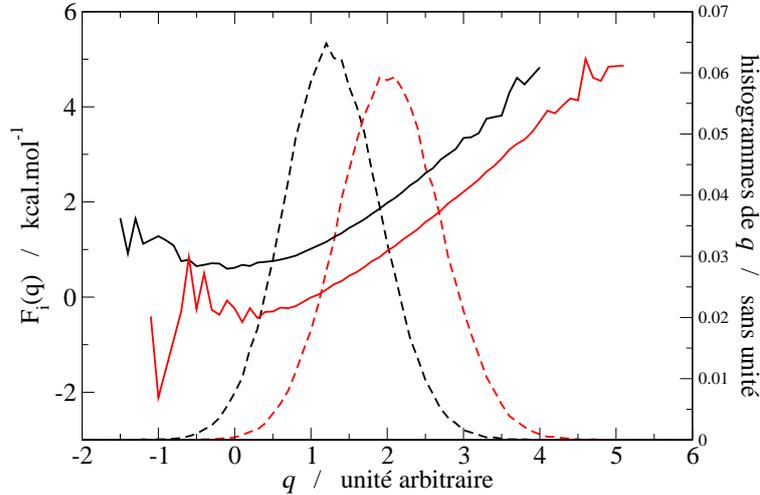


Figure I.11 – Deux courbes $\mathcal{F}_i(q)$ issues de deux fenêtres consécutives et les deux histogrammes biaisés des valeurs de q correspondants.

Nous avons utilisé des poids définis de deux façons différentes. La façon la plus simple est de tronquer l'histogramme à partir du moment où sa hauteur pour une valeur de q est inférieure à un certain seuil. De cette manière, on supprime les parties insuffisamment échantillonnées. Par exemple, on peut n'utiliser que la partie de l'histogramme où, pour chaque valeur de q , il prend des valeurs supérieures à 5% de son maximum. De cette manière, on définit un intervalle de valeurs de q pour chaque fenêtre i et les poids w_k^i valent 1 à l'intérieur de cet intervalle et sont nuls à l'extérieur. Une autre façon de faire, décrite dans la référence [52], est d'utiliser comme fonction de poids l'histogramme biaisé lui-même. De cette manière, les valeurs de q les plus échantillonnées sont largement favorisées par rapport aux autres et les termes qui ont le plus de poids dans le calcul des constantes C_i^{ste} sont ceux pour lesquels le recouvrement entre les histogrammes biaisés est maximal. L'équation I.21 s'écrit alors de la façon suivante

$$C_{i+1}^{ste} = \frac{\sum_k \rho_i^b(q_k) \rho_{i+1}^b(q_k) (\mathcal{F}_i(q_k) - \mathcal{F}_{i+1}(q_k))}{\sum_k \rho_i^b(q_k) \rho_{i+1}^b(q_k)} \quad (\text{I.22})$$

D'autres méthodes existent pour construire le profil d'énergie libre partielle, la plus utilisée étant la méthode WHAM (*Weighted Histogram Analysis Method* [53,54]). C'est une méthode itérative basée sur la minimisation de l'erreur statistique due à l'échantillonnage limité dans chaque fenêtre. On trouve également une méthode de *Umbrella integration* [55] qui utilise la dérivée des courbes $\mathcal{F}_i(q)$ par rapport à q ce qui évite le calcul des constantes C_i^{ste} .

I.5 Distribution de protons dans une protéine

Parmi les 20 acides aminés qui constituent les protéines, 7 d'entre eux présentent une fonction acido-basique et peuvent être présent dans la structure de la protéine sous une forme acide (protoné) ou basique (non protoné) (tableau A.1 page 159 en annexe). La forme sous laquelle se trouve l'acide aminé est l'état de protonation individuel. Le micro-état de protonation d'une protéine résulte de l'ensemble des états de protonation individuels des acides aminés de cette protéine. Chaque micro-état de protonation est caractérisé par un vecteur $X = (x_1, x_2, \dots, x_n)$ où les x_i représentent l'état de protonation individuel de chaque acide aminé. Il vaut 1 si l'acide aminé est protoné et 0 s'il est non protoné. Suivant l'état de protonation individuel de chacun de ces acides aminés, le potentiel électrostatique qu'ils créent est différent ce qui peut avoir des conséquences importantes sur la structure, les propriétés ou l'activité des protéines [56]. Connaître l'état de protonation individuel de chaque acide aminé acido-basique est donc crucial pour comprendre le fonctionnement de certaines protéines. Or, les structures obtenues par diffraction des rayons X n'ont pas une résolution suffisante pour définir la position des protons. D'autres techniques doivent donc être utilisées pour calculer l'état de protonation d'un acide aminé dans la protéine.

Après avoir considéré le calcul de la probabilité d'un état de protonation d'une molécule qui présente une fonction acido-basique en solution aqueuse, nous présenterons le calcul de cette probabilité dans une protéine par résolution de l'équation de Poisson-Boltzmann.

I.5.1 Etat de protonation en phase aqueuse

Dans ce paragraphe, nous rappellerons la relation entre la probabilité de l'état de protonation d'une molécule en solution aqueuse et le pH de la solution. Comme on l'a vu au paragraphe I.4, le rapport des probabilités entre deux états est directement liée à leur différence d'énergie libre $\Delta\mathcal{F}$. La relation entre le $\Delta\mathcal{F}$ associé au changement d'état de protonation d'un acide AH et le pH de la solution, se déduit de la constante d'équilibre K_a .



$$K_a = \frac{[A^-][H^+]}{[AH]} \quad (\text{I.23})$$

En utilisant la relation $pX = -\log(X)$ ^a, on obtient à partir de l'équation I.24

$$pH - pK_a = \log\left(\frac{[A^-]}{[AH]}\right) \quad (\text{I.24})$$

Le pK_a correspond à la valeur du pH pour laquelle la quantité de molécules A sous forme basique et acide est la même. La probabilité pour que la molécule A soit sous forme acide ou basique est

^aLa fonction \log représente le logarithme en base 10. Le logarithme en base e sera noté \ln

donc de 1/2. Le rapport $[A^-]/[AH]$ est le rapport de la probabilité pour que la molécule A soit sous sa forme basique et de la probabilité pour que la molécule A soit sous sa forme acide. On a donc

$$\frac{[A^-]}{[AH]} = \exp\left(-\frac{\Delta\mathcal{F}}{k_B T}\right) \quad (\text{I.25})$$

où $\Delta\mathcal{F}$ est l'énergie libre de changement d'état de protonation. En combinant les équations I.25 et I.24 on obtient la relation cherchée entre le pH et la différence d'énergie libre $\Delta\mathcal{F}$ entre les deux états de protonation

$$\Delta\mathcal{F} = -k_B T \ln(10) (pK_a - pH) \quad (\text{I.26})$$

Cette relation en phase aqueuse est utile car d'une part elle permet de relier le $\Delta\mathcal{F}$ et le pH de façon simple et d'autre part, en phase aqueuse les valeurs de pK_a pour l'ensemble des acides aminés sont connues (tableau A.1 page 159 en annexe). On pourra donc calculer, de façon approchée, la différence entre les énergies libres de protonation dans la protéine et dans l'eau au lieu d'avoir à évaluer directement l'énergie libre de protonation dans la protéine.

I.5.2 Etat de protonation dans la protéine

Une protéine est un polyacide. L'état de protonation individuel de chaque acide aminé dépend alors de la position et de l'état de protonation individuel des autres acides aminés de la protéine. A chaque micro-état de protonation est associé une énergie libre $\Delta\mathcal{F}$ calculée par rapport à un micro-état de référence dans lequel tous les acides aminés acido-basiques sont neutres (les acides carboxyliques, les alcools et les thiols sont protonés et les amines sont non protonées). L'énergie libre d'un micro-état par rapport à l'état de référence, $\Delta\mathcal{F}(X)$, est donnée par la formule de Bashford et Karplus [57].

$$\Delta\mathcal{F}(X) = -k_B T \sum_i x_i \ln 10 (pK_{intr,i} - pH) + \frac{1}{2} \sum_{i,j} (x_i + q_i^0) (x_j + q_j^0) W_{i,j} \quad (\text{I.27})$$

où les indices i et j représentent les acides aminés acido-basiques. Le pK_{intr} est le pK_a "intrinsèque" d'un acide aminé, c'est à dire le pK_a d'un acide aminé quand tous les autres sont dans leur état de référence. Le pK_{intr} dépend du pK_a de l'acide aminé i en phase aqueuse, de l'état de protonation individuel des acides aminés dans le micro-état défini par le vecteur X et de l'interaction entre le potentiel électrostatique créé par chaque acide aminé i , d'une part, et leur propre champ de réaction, les charges des autres acides aminés de la protéine dans leur état de référence (neutre) et le potentiel électrostatique créé par le solvant, d'autre part. Le second terme est un terme de paire qui dépend de l'état de protonation individuel des acides aminés i et j dans le micro-état de protonation défini par le vecteur X .

L'état de protonation individuel moyen d'un acide aminé est donné par

$$\langle x_i \rangle = \frac{\sum_X x_i \exp\left(-\frac{\Delta\mathcal{F}(X)}{k_B T}\right)}{\sum_X \exp\left(-\frac{\Delta\mathcal{F}(X)}{k_B T}\right)} \quad (\text{I.28})$$

La courbe des valeurs de $\langle x_i \rangle$ en fonction du pH est appelée courbe de titration. Elle donne la probabilité qu'un acide aminé soit sous sa forme acide ou basique en fonction du pH quelque soit l'état de protonation individuel des autres acides aminés. Le $\text{pK}_{\frac{1}{2}}$ d'un acide aminé est défini comme la valeur du pH pour laquelle $\langle x_i \rangle$ vaut 1/2.

Pour calculer les valeurs de $\Delta\mathcal{F}(X)$, connaissant le vecteur X et les charges portées par chaque atome, il faut calculer les valeurs de $\text{pK}_{intr,i}$ et W_{ij} . Ce calcul est fait en résolvant l'équation de Poisson-Boltzmann. Elle permet d'évaluer le potentiel électrostatique créé par la densité de charges de la protéine, $\rho(r)$ et par le solvant, traité comme un milieu diélectrique continu :

$$\nabla[\epsilon(r)\nabla V(r)] = -\rho(r) - \sum_i q_i c_i \exp\left(-\frac{q_i V(r)}{k_B T}\right) \quad (\text{I.29})$$

$\epsilon(r)$ représente la constante diélectrique, $V(r)$ est le potentiel électrostatique que l'on cherche et c_i est le nombre de charges q_i de type i présentes dans le solvant par unité de volume. Le dernier terme de l'équation de Poisson-Boltzmann représente des charges du solvant en équilibre thermique avec les charges de la protéine. Le cas le plus simple et le plus couramment utilisé consiste à considérer des ions monovalents (seuls deux types de charges q_i sont alors à considérer $q_i = \pm e$). De plus, on linéarise l'équation de Poisson-Boltzmann en considérant que l'énergie potentielle créée par les charges q_i , est faible devant l'énergie thermique.

$$\nabla[\epsilon(r)\nabla V(r)] = -\rho(r) + \frac{2e^2 \mathcal{N}_A I}{1000 k_B T} V(r) \quad (\text{I.30})$$

\mathcal{N}_A est le nombre d'Avogadro, e est la charge électrique élémentaire et I est la force ionique^a qui définit la concentration des charges q_i dans le solvant. La fonction $\epsilon(r)$ est approchée par une fonction marche : elle vaut $\epsilon_{solvant}$ pour les valeurs de r qui correspondent à l'extérieur de la protéine et une $\epsilon_{proteine}$ pour les valeurs de r qui correspondent à l'intérieur de la protéine. Le potentiel $V(r)$ est alors calculé par différences finies sur une grille de points.

I.5.3 Le serveur de calcul H++

Pour déterminer les micro-états de protonation les plus probables à un pH donné et les courbes de titration, nous avons utilisé le serveur de calcul H++ [58, 59]. Ce serveur résout l'équation de Poisson-Boltzmann linéarisée et calcule, pour l'ensemble des acides aminés acido-basiques, les courbes de titration (équation I.28) et les valeurs de $\text{pK}_{\frac{1}{2}}$. Trois paramètres doivent être précisés :

^aDans le cas d'ions monovalent la force ionique est en fait égale à la concentration c_i des ions en solution. Dans l'équation I.30, elle est convertie en $\text{mol} \cdot \text{L}^{-1}$ ce qui fait apparaître le facteur 1000 et le nombre d'Avogadro

la force ionique et la valeur de la constante diélectrique à l'intérieur de la protéine et sa valeur dans le solvant. La constante diélectrique du solvant est généralement celle de l'eau ($\epsilon_{\text{solvant}} = 80$). La force ionique et la constante diélectrique à l'intérieur de la protéine sont plus difficiles à déterminer. Les valeurs habituellement prises dans la littérature et recommandée par le serveur H++ sont $\epsilon_{\text{protéine}} = 4$ pour la constante diélectrique à l'intérieur de la protéine et $1.5 \text{ mol} \cdot \text{L}^{-1}$ pour la force ionique.

H++ donne également la liste des micro-états de protonation de la protéine de plus basse énergie et leur énergie libre associée à condition que le nombre d'acide aminés acido-basiques soit inférieur à 25. Ces calculs sont faits sur une géométrie donnée de la protéine, en précisant les charges effectives et les rayons atomiques portés par chaque atome.

D'autres serveurs permettant de calculer les micro-états de protonation existent. On peut notamment citer le serveur PROPKA qui se base sur une méthode empirique [60,61].

De plus, pour un pH donné plusieurs micro-états de protonation peuvent avoir des valeurs de $\Delta\mathcal{F}(X)$ 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 accessible.

I.6 Calcul d'une incertitude par la méthode du *bootstrap*

La méthode du *bootstrap* [62,63] vise à évaluer l'incertitude statistique liée à l'utilisation d'un échantillon de taille limité, sans rajouter d'information supplémentaire à celles contenues dans l'échantillon dont on dispose. Cette méthode permet donc de calculer les incertitudes statistiques sur la valeur moyenne, la valeur médiane ou encore les intervalles de confiances d'une grandeur (c'est à dire l'un des estimateurs statistiques de la densité de probabilité). Dans les études réalisées par simulation moléculaire ou par chimie quantique, la taille des échantillons est limitée, notamment par le temps de calcul. La densité de probabilité est donc obtenue de manière très approchée. Le *bootstrap* est alors bien adapté à l'évaluation des incertitudes statistiques sur les estimateurs statistiques.

Le *bootstrap* sera utilisé dans le chapitre II pour évaluer l'incertitude sur les spectres d'absorption de la Cerulean, à partir des calculs de transition électronique d'un échantillon de configurations extraites d'une dynamique moléculaire classique.

I.6.1 La méthode

Le *bootstrap* est synonyme de ré-échantillonnage. On note x la variable échantillonnée. Soit θ la vraie valeur, en général inconnue, de l'estimateur statistique (par exemple la valeur moyenne) que l'on cherche à déterminer. On note E_i un échantillon initial contenant N éléments, de valeur x_k^i , obtenu par une technique d'échantillonnage quelconque à partir de l'espace complet des éléments. On note θ_i la valeur de l'estimateur statistique θ , obtenue à partir de l'échantillon initial. La méthode du *bootstrap* consiste à construire N_b échantillons de *bootstrap*, notés E_b , de la même taille que

l'échantillon initial, en utilisant les éléments de l'échantillon initial. Les N éléments de l'échantillon E_b , de valeur x_k^b , sont choisis aléatoirement parmi les N éléments de l'échantillon E_i avec remise, c'est à dire qu'un même élément peut être choisi plusieurs fois. A chaque tirage, chaque élément de l'échantillon initial a la même probabilité $1/N$ d'être choisi. De cette manière la densité de probabilité construite à partir des valeurs x_k^b des éléments de E_b conserve approximativement la même forme que celle construite à partir des valeurs x_k^i des éléments de E_i . En effet, plus la probabilité d'obtenir une valeur donnée x_0 est grande, plus le nombre d'éléments de l'échantillon E_i qui ont une valeur x_k^i proche de x_0 sera grand et donc, lors de la construction des échantillons E_b , la probabilité de choisir un élément de E_i dont la valeur x_k^i est proche de x_0 sera grande. La figure I.12 résume le fonctionnement de la méthode du *bootstrap*.

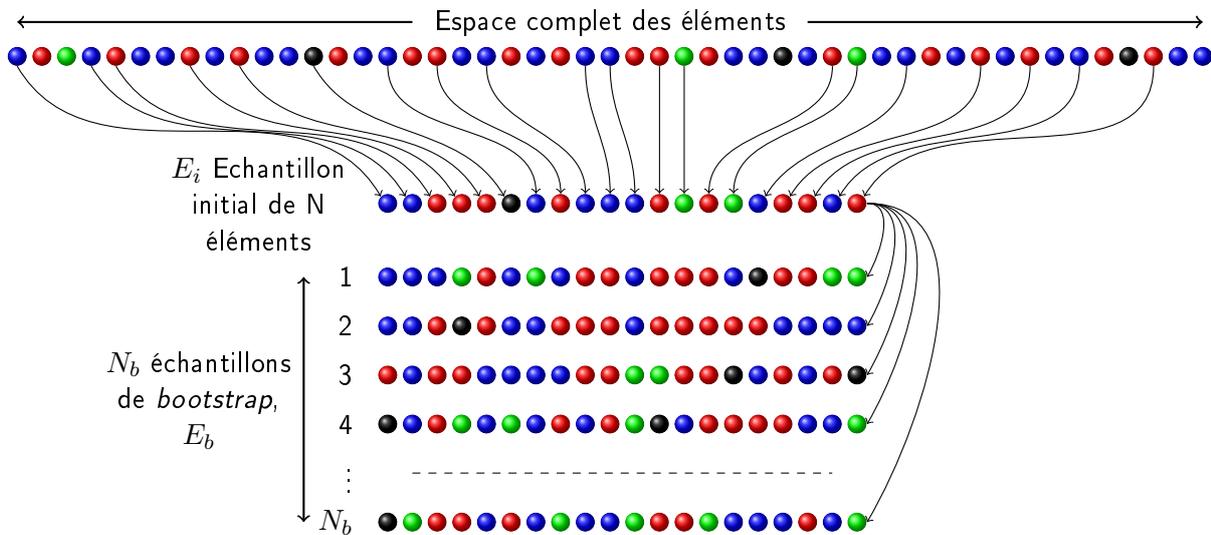


Figure I.12 – Schéma permettant de visualiser la méthode du *bootstrap*. L'échantillon initial E_i est obtenu à partir de l'espace complet des éléments par une méthode d'échantillonnage (dynamique moléculaire, tirage aléatoire ...). Les échantillons de *bootstrap* sont obtenus en tirant aléatoirement N éléments parmi les N éléments de l'échantillon initial avec remise : un élément peut donc être tiré plusieurs fois.

Pour chacun des N_b échantillons E_b , la valeur θ_b de l'estimateur θ peut être calculée. On obtient ainsi la distribution $\mathcal{D}^b(\theta)$ des valeurs de l'estimateur θ extraites des N_b échantillons E_b . D'autre part, notons $\mathcal{D}^{all}(\theta)$ la distribution des valeurs de l'estimateur θ extraites de tous les échantillons contenant N éléments qui peuvent être construit à partir de l'espace complet des éléments. Il a été démontré par B. Efron qu'à la limite où le nombre N d'éléments de l'échantillon initial tend vers l'infini, les distributions $\mathcal{D}^{all}(\theta)$ et $\mathcal{D}^b(\theta)$ sont égales [62].

La largeur de la distribution $\mathcal{D}^{all}(\theta)$ donne une estimation des fluctuations de l'estimateur θ auxquelles on peut s'attendre compte tenu du fait qu'on utilise un échantillon ne contenant que N éléments. Ces fluctuations sont alors interprétées comme l'incertitude statistique sur l'estimateur θ . Le *bootstrap* permet de construire la distribution $\mathcal{D}^b(\theta)$ et d'obtenir l'incertitude sur l'estimateur θ en supposant qu'elle est similaire à la distribution $\mathcal{D}^{all}(\theta)$, qui n'est pas calculable.

L'incertitude $u(\theta)$ sur l'estimateur θ est obtenue directement à partir de l'écart entre θ_i et la moyenne

de θ_b ainsi que des de l'écart type de θ_b . Les échantillons E_b de *bootstrap* étant obtenus directement à partir de l'échantillon initial, l'incertitude sur l'estimateur θ est obtenue sans ajouter d'information supplémentaire à l'échantillon initial. Elle est donnée par :

$$u(\theta) = \sqrt{(\theta_s - \langle \theta_b \rangle_{N_b})^2 + \sigma_{N_b}^2} \quad (\text{I.31})$$

La valeur de l'écart entre θ_i et la moyenne de θ_b est généralement très faible et l'incertitude sur θ est alors directement reliée à l'écart type σ_{N_b} qui quantifie les fluctuations de θ_b autour de θ_i .

I.6.2 Un exemple d'application de la méthode du *bootstrap* sur une distribution modèle

Considérons une variable x dont la densité de probabilité $\rho(x)$ est une distribution gaussienne de largeur σ_x et de moyenne $\langle x \rangle$.

$$\rho(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(\frac{-(x - \langle x \rangle)^2}{2\sigma_x^2}\right)$$

Le but de cet exemple est de montrer comment la méthode du *bootstrap* permet de calculer l'incertitude due à la taille de l'échantillon utilisé, sur la valeur moyenne ou l'écart type.

Les N valeurs de x de l'échantillon initial ont été obtenues par la méthode du rejet [64, p 281] à partir de la distribution $\rho(x)$. L'avantage d'utiliser une densité de probabilité connue est que l'on pourra comparer les vraies valeurs de la moyenne et de l'écart type, notés respectivement $\langle x \rangle$ et σ_x , aux valeurs de la moyenne et de l'écart type obtenues à partir de l'échantillon en tenant compte des incertitudes obtenues par le *bootstrap*. Dans cet exemple les valeurs $\langle x \rangle = 0$ et $\sigma_x = 1$ ont été utilisées. Un des paramètres importants lors de la mise en oeuvre de la méthode du *bootstrap* est le nombre N_b d'échantillons de *bootstrap* générés. Ce nombre doit être choisi de sorte que la moyenne sur les échantillons de *bootstrap* et l'écart type aient convergé, c'est-à-dire soient devenues indépendants de N_b . Le nombre à utiliser dépend de la taille et du type d'échantillon que l'on considère. Dans cet exemple, on a utilisé 5000 échantillons de *bootstrap*.

L'incertitude sur la valeur moyenne de x , $u(\langle x \rangle)$, et sur l'écart type, $u(\sigma_x)$, ont été calculées en fonction du nombre N de valeurs de x dans l'échantillon initial en utilisant l'équation I.31. Les variations de ces incertitudes en fonction de N sont présentées figure I.13 a. Cette figure montre, comme attendu, que plus le nombre N de valeurs de x contenues dans l'échantillon initial augmente, plus l'incertitude sur la valeur moyenne et l'écart type diminuent. On a obtenu les valeurs suivantes pour la moyenne de x et l'écart type de x ainsi que pour leurs incertitudes associées, obtenues dans deux cas particuliers, $N=50$ et $N=1000$:

vraies valeurs	$N= 50$	$N= 1000$
$\langle x \rangle = 0$	$\langle x \rangle = -0.14 \pm 0.14$	$\langle x \rangle = 0.00 \pm 0.03$
$\sigma_x = 1$	$\sigma_x = 0.97 \pm 0.13$	$\sigma_x = 0.98 \pm 0.02$

Les valeurs et les incertitudes obtenues sont en bon accord avec la vraie valeur de $\langle x \rangle$ et σ_x .

La figure I.13 b illustre les résultats obtenus. Elle présente, dans le cas où $N=50$ et $N=1000$, les distributions $\mathcal{D}^b(\langle x \rangle)$ obtenues à partir des échantillons de *bootstrap*. On remarque que la distribution obtenue pour $N=50$ est plus large que celle obtenue pour $N=1000$. Cette largeur traduit le fait que dans le cas où $N=50$, l'échantillon initial est de faible qualité et que l'incertitude statistique sur la valeur moyenne est importante.

La méthode du *bootstrap* permet donc de calculer une incertitude statistique sur une grandeur donnée et de quantifier la qualité d'un échantillon.

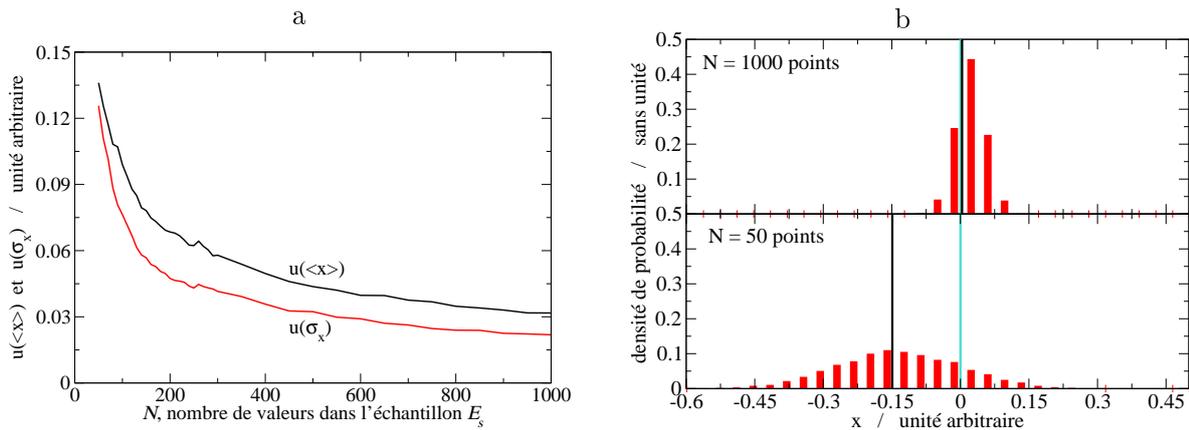


Figure I.13 – a) Incertitude $u(\langle x \rangle)$ sur la valeur moyenne de x (en noir) et $u(\sigma_x)$ sur l'écart type de x (en rouge), en fonction du nombre N de valeurs de x dans l'échantillon initial. b) En rouge, distribution $\mathcal{D}^b(\langle x \rangle)$ des valeurs moyennes calculées sur les N_b échantillons de *bootstrap* dans le cas où l'échantillon initial contient $N=1000$ valeurs de x (en haut) et $N=50$ valeurs de x (en bas). Le trait vert représente la vraie valeur moyenne de la distribution utilisée et le trait noir représente la valeur moyenne de l'échantillon initial.

Chapitre II

Relation entre pH, structure et spectre d'absorption de la Cerulean

Sommaire

II.1 Introduction	50
II.1.1 La <i>Cyan Fluorescent Protein</i>	50
II.1.2 La mutation H148D - de la ECFP à la Cerulean	51
II.1.3 Origines possibles de la dépendance en pH du spectre d'absorption de la Cerulean	53
II.1.4 Objectifs et présentation de l'étude	55
II.2 Dynamique moléculaire de la Cerulean	55
II.2.1 Distribution des protons à pH 5 et pH 8	56
II.2.2 Coordonnées initiales	59
II.2.3 Structure et dynamique de la Cerulean à pH 5	61
II.2.4 Dynamique globale de la Cerulean - mouvement des brins 6 et 7	69
II.2.5 Structure et dynamique à pH 8	79
II.3 Spectre d'absorption en fonction du pH	84
II.3.1 Calcul d'un spectre d'absorption par simulation moléculaire	85
II.3.2 Spectres d'absorption de la Cerulean	87
II.3.3 Facteurs influençant la forme et la position du spectre d'absorption théorique	88
II.3.4 Transitions électroniques impliquées dans les spectres d'absorption	95
II.4 Vers une description plus complète de l'effet du pH	99
II.4.1 Simulation Monte Carlo dans l'espace des micro-états de protonation	100
II.4.2 Densité de probabilité des micro-états de protonation	100
II.4.3 Effet des micro-états de protonation sur une observable	102
II.5 Conclusion du chapitre	103

Ce chapitre présente une étude de la Cerulean, un mutant de la GFP, qui est un bon candidat pour remplacer la CFP dans les expériences d'imagerie cellulaire utilisant le FRET. L'objectif de ce travail était de comprendre la relation entre la structure de la Cerulean et son spectre d'absorption afin de pouvoir argumenter les hypothèses faites par Malo *et al* [37] pour expliquer les résultats expérimentaux concernant leur dépendance au pH.

II.1 Introduction

La Cerulean et la 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 [27].

II.1.1 La *Cyan Fluorescent Protein*

La CFP couramment utilisée est en fait la *Enhanced Cyan Fluorescent Protein*, ECFP, qui est obtenue à partir de la GFP sauvage en faisant la mutation Y66W et cinq mutations supplémentaires (S65T, F64L, N146I, M153T, V163A) permettant entre autre d'augmenter le rendement quantique de fluorescence [65].

Une étude par RMN du ^{19}F de la ECFP a mis en évidence l'existence de deux conformations qui s'échangent à une échelle de temps de l'ordre de la milliseconde [66]. La résolution de la structure de la ECFP par diffraction des rayons X [67] a confirmé l'existence d'une conformation A majoritaire (60%) et d'une conformation B (40%) minoritaire. Ces structures diffèrent par l'orientation des chaînes latérales des acides aminés His148 et Tyr145 qui sont proches du chromophore (figure II.1 a, les distances entre le carbone α des acides aminés His148 et Tyr145 et l'atome pontant entre les deux cycles du chromophore sont respectivement de 8.8 et 11.7 Å). Dans la conformation A, la chaîne latérale de Tyr145 est nettement à l'intérieur de la protéine alors que celle de His148 est nettement à l'extérieur. Par contre dans la conformation B la chaîne latérale de Tyr145 est dans le solvant, avec une orientation approximativement parallèle à l'axe du tonneau. Celle de His148 est elle aussi à peu près parallèle à l'axe du tonneau et dans la zone frontière entre la protéine et le solvant.

Le spectre d'absorption de la ECFP est caractérisé par deux maxima à 430 nm et à 450 nm [67] et est peu dépendant du pH [68]. Bae *et al* [67] avaient émis l'hypothèse que ces deux maxima correspondent aux deux conformations A et B. Cependant, lors d'une étude antérieure par dynamique moléculaire effectuée au laboratoire [69], nous avons montré que les spectres d'absorption associés à chaque conformation sont presque identiques, ce qui permet d'écarter le rôle des deux conformations dans la double bosse du spectre expérimental. Cette double bosse est vraisemblablement due à une structure vibrationnelle, mais cette hypothèse n'a jusqu'à présent jamais été vérifiée.

Les propriétés de fluorescence de la ECFP ne sont pas optimales. Son rendement quantique est assez faible, proche de 0.4 [25,28] et son déclin de fluorescence est très complexe. Des études expérimentales

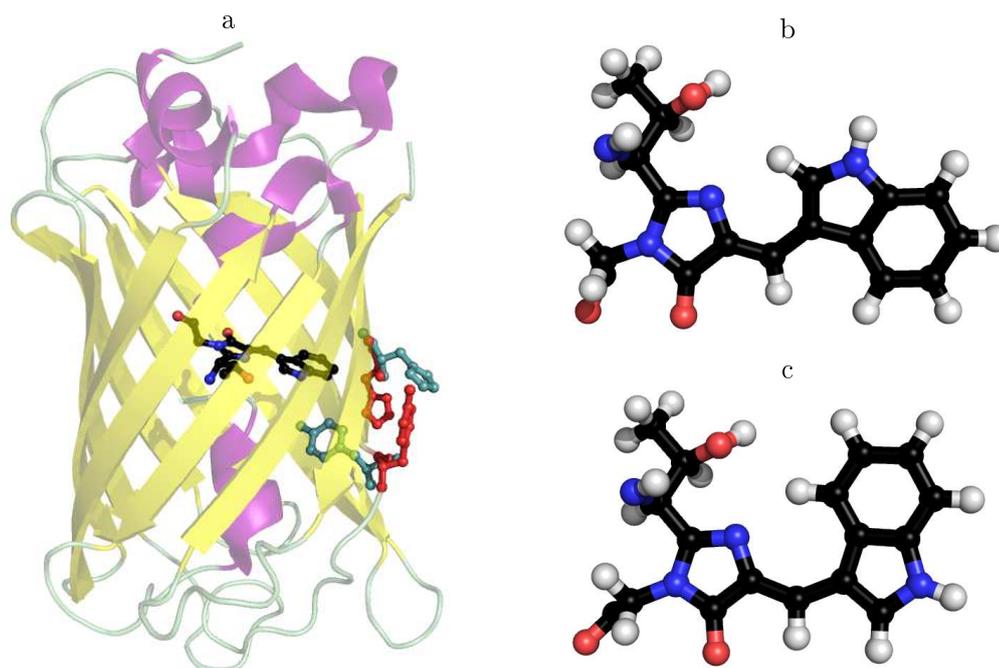


Figure II.1 – a : Structure de la ECFP. Tyr145 et His148 sont représentés en bleu pour la conformation A (code PDB 1OXD) et en rouge pour la conformation B (code PDB 10XE). b : Chromophore dans la conformation *trans*. c : Chromophore dans la conformation *cis*.

[70,35] ont mis en évidence deux types d'hétérogénéité. D'une part, il est fortement multiexponentiel, plusieurs composantes de temps entre 0.1 et 10 ns sont nécessaires pour représenter le déclin de fluorescence. D'autre part ces composantes de temps et la durée de vie moyenne de l'état excité de la ECFP sont dépendantes de la longueur d'onde d'émission. De plus ces propriétés de la fluorescence de la ECFP présentent également une forte dépendance au pH et à la température. La durée de vie moyenne de la ECFP est constante entre pH 11 et pH 7 et diminue de 2.5 ns à 1.5 ns quand le pH varie de pH 7 à pH 5.5. Quatre composantes de temps sont nécessaires pour représenter le déclin de la fluorescence à pH 5.5 alors qu'une seule suffit à pH 11. L'effet de la température sur la durée de vie moyenne est plus régulier : elle diminue linéairement de 3.5 ns à 1.5 ns quand la température augmente de 275 K à 315 K. Le nombre de composantes dans le déclin de fluorescence est de quatre à 273 K alors qu'une composante domine à 310 K.

II.1.2 La 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. Rizzo *et al* [36] en se basant sur l'hypothèse que la complexité du déclin de fluorescence était due à la présence des deux conformations A et B, ont proposé de faire une mutation impliquant His148 et/ou Tyr145 dans le but de supprimer l'une de ces deux conformations. En effet, la substitution de l'un de ces acides aminés par un acide aminé hydrophile, devrait stabiliser fortement la conformation

dans laquelle cet acide aminé est orienté vers le solvant. Dans cette optique ils ont remplacé l'histidine 148 par un acide aspartique (mutation H148D). Cet acide aminé a un pKa de 3.8 et aux pH physiologiques il est presque toujours sous sa forme anionique et donc très hydrophile. Cette mutation produit effectivement le résultat recherché : le déclin de fluorescence est plus proche d'un déclin exponentiel que celui de la ECFP. Rizzo *et al* ont introduit deux mutations supplémentaires visant à accroître les capacités de maturation de la ECFP/H148D qui ont conduit à la Cerulean (ECFP/S72A/Y145A/H148D) [36]. La mutation Y145A substitue un acide aminé tyrosine, hydrophile, par un acide aminé alanine, hydrophobe. La combinaison des mutations Y145A/H148D devrait donc fortement favoriser la conformation A, dans laquelle la chaîne latérale de l'acide aminé 145 est orientée vers l'intérieur de la protéine et la chaîne latérale de l'acide aminé en position 148 orientée vers l'extérieur de la protéine (figure II.1 a).

En 2007 la structure RX de la Cerulean a été déterminée à pH 5 par Malo *et al* [37]. Ces auteurs n'ont pas pu obtenir de cristaux à pH plus élevé et, à notre connaissance, aucune structure de la Cerulean à pH neutre ou basique n'a été publiée à ce jour. Cette structure RX à pH 5 présente plusieurs singularités. Le chromophore de la Cerulean est en conformation *cis* alors qu'il adopte une conformation *trans*^a dans la structure RX de la ECFP à pH 8 [67] (figure II.1 b et c). La chaîne latérale de Asp148 supposée être exposée au solvant [36] est en fait à l'intérieur de la protéine, en liaison hydrogène avec l'indole du chromophore de la Cerulean (figure II.2 a). 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* 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.

Malo *et al* ont déterminé le spectre d'absorption de la Cerulean pour des valeurs du pH allant de 5 à 10. A tous pH le spectre présente deux maxima. A pH 8 il est quasi identique à celui de la ECFP [36]. Il est pratiquement inchangé entre pH 8 et 10, mais, lorsque le pH diminue de 8 à 5 les deux maxima se déplacent vers le bleu (vers les courtes longueurs d'ondes) de respectivement 11 et 15 nm (figure II.2 b). Ce décalage est très lent et se produit en environ 1 h. Une telle dépendance du spectre au pH n'a pas été observé pour la ECFP.

Les caractéristiques de la fluorescence de la Cerulean sont plus avantageuses pour l'imagerie cellulaire que celles de la ECFP [36,37]. Son rendement quantique est 1.7 fois plus grand et son coefficient d'absorption 1.5 fois plus grand, ce qui fait qu'elle est 2.5 fois plus brillante^b que la ECFP. Son déclin de fluorescence se rapproche plus d'un simple déclin mono-exponentiel. Cette propriété est clairement due à la mutation H148D. En effet, le mutant ECFP/H148D, qui présente les mêmes caractéristiques photophysiques que la Cerulean, a été étudié dans le groupe de Biophysique du laboratoire [35] et les résultats montrent qu'une composante de temps domine largement les autres

^aLa nomenclature des conformères obtenue par une rotation de l'angle φ est ambiguë car l'angle φ correspond à la torsion autour d'une liaison simple dans l'état fondamental. La nomenclature *cis/trans* adoptée ici est celle utilisée par Malo *et al*.

^bPar convention la brillance est définie comme le produit du rendement quantique de fluorescence par le coefficient d'extinction molaire.

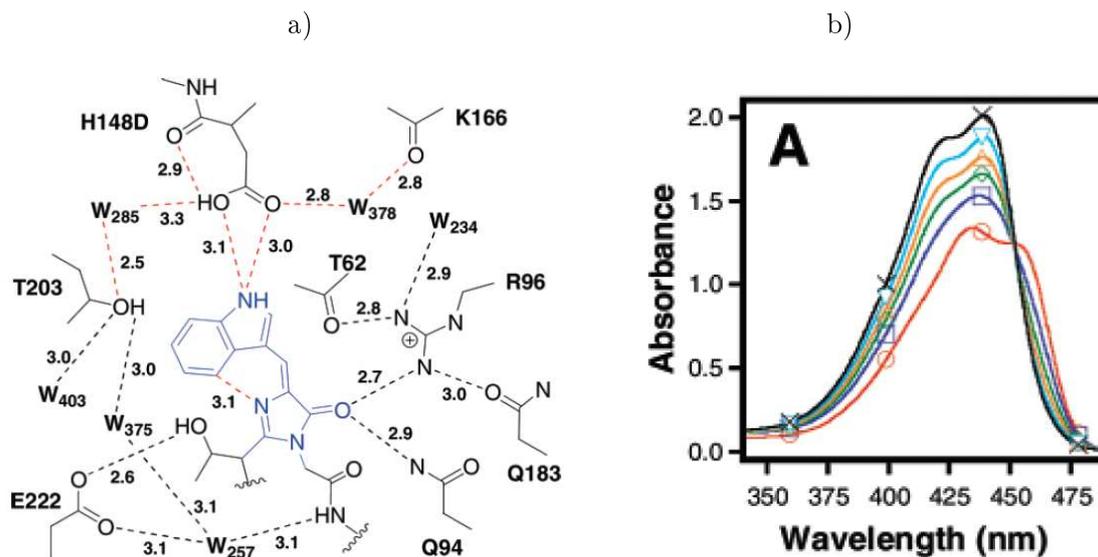


Figure et résultats reproduits à partir de l'article de Malo *et al* [37]

Figure II.2 – a) 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*. b) Spectre d'absorption de la Cerulean en solution en fonction du temps après un saut de pH de pH 8 à pH 5. A $t = 0$ (rouge ○) pH 8, 20 min (violet □), 40 min (vert ◇), 60 min (orange △), 90 min (bleu ▽), 150 min (noir ×) pH 5.

dans le déclin de la fluorescence. La durée de vie moyenne de l'état excité de la Cerulean ne dépend pas significativement de la longueur d'onde d'émission. Cependant les caractéristiques de sa fluorescence sont encore complexes. L'intensité de fluorescence diminue d'un facteur 4 lorsque le pH est abaissé de pH 8 à pH 5. La durée de vie moyenne présente des dépendances au pH et à la température similaire à celles de la ECFP [35].

II.1.3 Origines possibles de la dépendance en pH du spectre d'absorption de la Cerulean

Malo *et al* proposent deux hypothèses non exclusives l'une de l'autre pour expliquer la dépendance en pH du spectre d'absorption de la Cerulean : une réorganisation de la distribution des protons des acides aminés voisins du chromophore (effet électrostatique) et/ou une isomérisation du chromophore.

Il est très probable que Asp148 soit sous sa forme basique à pH 8. S'il est sous sa forme acide à pH 5, comme le suppose Malo *et al*, il y a effectivement un changement important dans l'environnement électrostatique du chromophore qui pourrait à lui seul expliquer un décalage d'une dizaine de nm. Il y a également des arguments en faveur d'une possible isomérisation du chromophore. L'un de ces arguments vient du fait que le décalage du spectre est très lent, de l'ordre de 1 heure. Ceci indique que lors du décalage du spectre, le système doit franchir une barrière élevée. Or l'isomérisation du

chromophore à l'intérieur de la protéine se heurte effectivement à une barrière importante. D'une part, la rotation de l'indole d'une conformation plane (conformation *cis*) à l'autre (conformation *trans*) implique que le système franchisse une barrière intramoléculaire élevée. De plus, cette rotation est gênée par les acides aminés voisins. Dans le cas des protéines photoconvertibles, l'isomérisation spontanée à l'état fondamental se fait en des temps allant de quelques minutes à quelques heures [71]. Le temps caractéristique observé ici est bien de cet ordre de grandeur.

Une autre possibilité pour expliquer le décalage du spectre d'absorption pourrait être un changement d'état de protonation du chromophore lui-même. Dans la gamme de pH étudiée, la perte du proton du groupe NH de l'indole n'est pas envisageable. En effet, l'indole formerait une amidure qui est une espèce très réactive (le pKa d'un indole est de l'ordre de 17 en solution aqueuse). Par contre l'ajout d'un proton sur l'azote N13 du cycle imidazolinone à pH acide est une hypothèse plus réaliste. Pour la tester, nous avons fait des calculs quantiques de l'énergie de transition du chromophore neutre en conformation *cis* ou *trans* et du chromophore cationique en conformation *cis*. Dans les trois cas la géométrie du chromophore a été optimisée au niveau B3LYP/6-31G* et l'énergie de transition a été calculée au niveau TDDFT. Les résultats sont rassemblés sur la figure II.3.

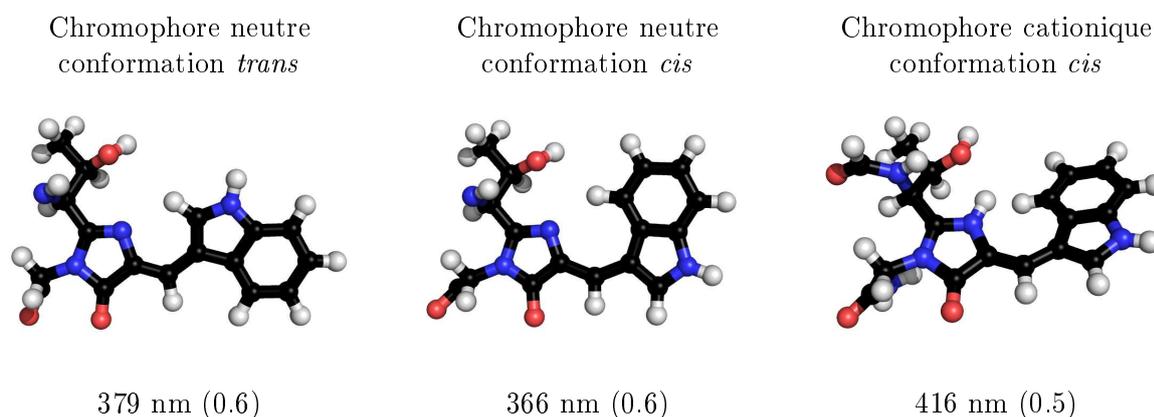


Figure II.3 – Longueur d'onde de la première transition électronique et force d'oscillateur associée (entre parenthèses) pour la conformation *cis* du chromophore neutre et du chromophore cationique et la conformation *trans* du chromophore neutre.

On voit qu'en supposant que le chromophore est toujours neutre, la longueur d'onde diminue de 13 nm quand on passe de la conformation *trans* à la conformation *cis*, en accord avec les calculs de la littérature [72, 73]. Par contre, si on suppose que le chromophore est cationique à pH 5, le décalage de la longueur d'onde quand on passe de pH 8 à pH 5 serait un décalage vers le rouge de l'ordre 40 ou 50 nm suivant la conformation *trans* ou *cis* du chromophore à pH 8. Ce décalage est en sens opposé à celui du décalage expérimental et d'amplitude nettement plus grande. Bien que ces résultats ne tiennent pas compte de l'environnement protéique, ils nous permettent d'écartier l'hypothèse d'un chromophore cationique à pH 5.

Il est donc raisonnable de s'en tenir aux hypothèses faites par Malo *et al* sur l'origine du décalage du spectre d'absorption. Ces auteurs ont proposé une description qualitative de ce qui pourrait se passer quand le pH augmente de 5 à 8, qui est la suivante. Asp148 pourrait perdre son proton (le pKa d'un acide aspartique est de 3.9 en solution) et sa chaîne latérale pourrait se tourner vers l'extérieur

de la protéine pour stabiliser sa charge négative (c'est cette hypothèse qui avait conduit Rizzo *et al* à proposer la mutation H148D afin de stabiliser une des deux conformations de la ECFP [36]). L'indole du chromophore, en conformation *cis*, serait alors dépourvu de liaison hydrogène et dans une cavité plus large du fait de l'absence de Asp148. Ces deux effets donneraient au chromophore une liberté de mouvement plus importante pouvant conduire à son isomérisation en conformation *trans*. Cette conformation serait alors stabilisée car l'indole pourrait former de nouvelles liaisons hydrogène et le chromophore pourrait se trouver dans une situation similaire à celle qu'il adopte dans la ECFP.

II.1.4 Objectifs et présentation de l'étude

L'étude que nous avons réalisée sur la Cerulean a pour objectif d'apporter des éléments qui permettent de confirmer ou infirmer les hypothèses faites par Malo *et al*. Tout d'abord, nous avons étudié l'état de protonation et les conformations préférentielles de Asp148 suivant la forme acido-basique qu'il adopte. Ensuite, nous avons étudié l'effet sur le spectre d'absorption de la Cerulean du changement de conformation du chromophore et de l'effet électrostatique lié au changement de pH et donc de la distribution des protons dans la protéine.

Dans cette optique, nous avons réalisé trois simulations de dynamique moléculaire de la Cerulean. Ces trois cas seront nommés pH5-*cis*, pH8-*cis* et pH8-*trans* en fonction du pH et de la conformation du chromophore considérée.

pH5-*cis* : Structure et dynamique de la Cerulean à pH 5 avec le chromophore en conformation *cis*.

pH8-*cis* : Structure et dynamique de la Cerulean à pH 8 avec le chromophore en conformation *cis*.

pH8-*trans* : Structure et dynamique de la Cerulean à pH 8 avec le chromophore en conformation *trans*.

Dans ce qui suit, sont présentés les résultats de cette étude. Dans un premier temps, le choix que nous avons fait pour modéliser le pH dans nos simulations sera discuté. Ensuite les simulations moléculaires réalisées dans les trois cas présentés ci-dessus seront décrites et les résultats sur la structure et la dynamique de la protéine, du chromophore et de Asp148 seront discutés. Dans une dernière partie nous discuterons de la façon d'utiliser ces simulations pour calculer le spectre d'absorption de la Cerulean à différents pH.

II.2 Dynamique moléculaire de la Cerulean

Dans cette partie, les trois simulations de dynamique moléculaire seront présentées. Le choix des coordonnées de départ et la détermination de l'état de protonation le plus probable des acides aminés acido-basiques, selon le pH, sont décrits. L'analyse de la structure de la Cerulean en fonction du pH et la dynamique de la chaîne latérale de Asp148 sont discutées.

II.2.1 Distribution des protons à pH 5 et pH 8

Les acides aminés dont la chaîne latérale porte une fonction acido-basique peuvent être sous leur forme acide (protoné) ou sous leur forme basique (non protoné) 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. La probabilité de chacune de ces distributions dépend du pH. L'état de protonation de la protéine à un pH donné correspond alors à un ensemble de distributions de protons.

Dans les simulations classiques standard, les liaisons chimiques sont définies *a priori* et ne sont pas modifiables au cours de la simulation, il est donc indispensable de choisir une distribution de protons. En général on choisit la plus probable. Une alternative est de réaliser des simulations à pH constant dans lesquelles la distribution des protons dans la protéine est fréquemment modifiée [74]. Cette approche n'a pas été utilisée dans ce travail.

Dans cette partie nous avons choisi de modéliser le rôle du pH en utilisant le micro-état de protonation le plus probable à pH 5 ou à pH 8. Dans la Cerulean, 77 acides aminés présentent une fonction acido-basique : il s'agit des différents acides aspartiques, acides glutamiques, histidines, lysines, arginines, cystéines et tyrosines contenus dans cette protéine. Les distributions de protons et leur probabilité ont été obtenues par résolution de l'équation de Poisson-Boltzmann à l'aide du serveur de calcul H++ [58, 59] (paragraphe I.5 page 42) et font l'objet des paragraphes suivants. Dans ces calculs, nous avons employé les paramètres les plus courants : la constante diélectrique à l'intérieur de la protéine est de 4, celle dans le solvant de 80, et la force ionique du milieu est fixée à 0.15 mol/L.

II.2.1.1 Courbe de titration de la chaîne latérale de Asp148

La courbe de titration de Asp148 calculée par H++ dans la géométrie de la structure RX à pH 5 est présentée figure II.4. Elle indique que le $pK_{\frac{1}{2}}$ de Asp148 vaut 6.4 et que, à pH 5 la probabilité qu'il soit protoné est très proche de 1 (0.9). Ce résultat conforte l'hypothèse faite par Malo *et al.* En solution aqueuse le pKa de l'acide aspartique vaut 3.4. Asp148 est donc un acide plus faible dans la protéine qu'en solution et a tendance à perdre moins facilement son proton.

II.2.1.2 Distributions de protons à pH 5

Nous avons cherché à déterminer les micro-états de protonation de plus basse énergie libre. Cependant, H++ détermine les micro-états de protonation seulement si le nombre d'acides aminés acido-basiques est inférieur à 25. Pour réduire le nombre d'acides aminés acido-basiques inclus dans le calcul par H++ des micro-états de protonation, nous avons assigné l'état de protonation des acides aminés dont le $pK_{\frac{1}{2}}$ ou la courbe de titration permettait de le définir sans ambiguïté. C'est le cas, pour l'ensemble des acides aminés cystéine, tyrosine, lysine et arginine, soit 36 acides aminés

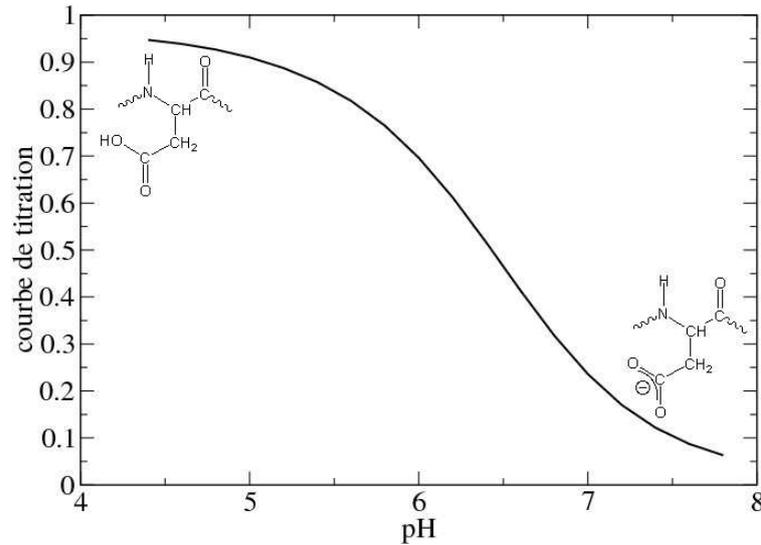


Figure II.4 – Courbe de titration de Asp148 dans la Cerulean. L’ordonnée représente la probabilité que Asp148 soit dans sa forme acide (protoné et neutre). Les deux molécules en haut à gauche et en bas à droite représentent respectivement la forme acide et la forme basique d’un acide aspartique.

car leur $pK_{\frac{1}{2}}$ est supérieur à 8. Ces acides aminés sont donc pris dans leur forme acide à pH 5. De la même manière les acides aminés : Glu5, Asp21, His25, Asp36, His81, Asp82, Asp103, Glu111, Asp155, His169, His181, His199, Asp210, Glu213, Asp216 et Glu222 ont pour $pK_{\frac{1}{2}}$ une valeur inférieure à 3 ou supérieure à 8. La forme basique a été assignée aux acides aminés dont le $pK_{\frac{1}{2}}$ est inférieur à 3 et la forme acide a été assignée à ceux dont le $pK_{\frac{1}{2}}$ est supérieur à 8. De cette manière l’état de protonation reste indéterminé pour seulement 24 acides aminés et H^{++} fournit alors la liste des 100 micro-états de protonation de plus basse énergie. Nous avons pu remarquer que les courbes de titration de ces 24 acides aminés sont peu modifiées par rapport au calcul H^{++} initial prenant en compte l’ensemble des acides aminés acido-basiques. Les valeurs de $pK_{\frac{1}{2}}$ varient de moins de 0.3 unité. Les 20 micro-états de plus basse énergie sont donnés dans le tableau II.1.

L’intervalle des valeurs des énergies libres associées aux vingt micro-états du tableau II.1 est de $0.29 \text{ kcal} \cdot \text{mol}^{-1}$ et il est seulement de $0.6 \text{ kcal} \cdot \text{mol}^{-1}$ pour les 100 micro-états de plus basse énergie. Le rapport des probabilités de Boltzmann à température ambiante (pour une température de 300 K, $k_B T = 0.59 \text{ kcal} \cdot \text{mol}^{-1}$) entre le micro-état le plus bas et le 20^{ème} micro-état est de l’ordre de 0.4. Ce résultat souligne le fait qu’un unique micro-état de protonation n’est pas représentatif des distributions de protons à un pH donné dans une protéine. Cependant, dans le cas présent, parmi les 100 micro-états de plus basse énergie, seuls 11 acides aminés changent d’état de protonation : Glu6, Glu17, Asp19, Glu32, Glu34, His77, Glu90, Glu124, His139, Glu172 et Asp180. Leurs chaînes latérales sont orientées dans le solvant et aucun d’entre eux n’est proche du chromophore ou de Asp148. On peut donc supposer que leur état de protonation aura peu d’influence sur la dynamique du chromophore et le spectre d’absorption. Le micro-état de plus basse énergie a donc été retenu pour décrire la distribution des protons dans la protéine à pH 5.

Dans ce micro-état de protonation, tous les acides aminés Cys, Tyr, Lys et Arg sont dans leur forme

Liste des 20 micro-états de protonation de plus basse énergie à pH 5

0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0	Energie :-6.237765
0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0	Energie :-6.241885
0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0	Energie :-6.250822
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.253264
0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.287508
0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.287881
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.303818
0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.305521
0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.311610
0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.315763
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0	Energie :-6.329690
0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.347528
0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.393938
0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.397347
0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.403605
0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.433411
0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.437868
0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.445058
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.473899
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0	Energie :-6.523533

Tableau II.1 – Liste des 20 micro-états de protonation de plus basse énergie à pH 5. Chaque ligne représente un micro-état dans lequel 0 signifie que l'acide aminé est dans sa forme basique (non protoné) et 1 signifie qu'il est dans sa forme acide (protonée). Chaque colonne correspond à un acide aminé. Ils sont donnés par numéro croissant : Glu6, Glu17, Asp19, Glu32, Glu34, His77, Glu90, Glu95, Asp102, Glu115, Asp117, Glu124, Asp129, Glu132, Asp133, His139, Glu142, Asp148, Glu172, Asp173, Asp180, Asp190, Asp197, His217. La dernière colonne est l'énergie libre associée au micro-état en kcal·mol⁻¹.

acide, les acides aminés Asp et Glu sont dans leur forme basique, sauf Asp19, Asp148 et Glu142 qui sont dans leur forme acide (neutre, protoné). Les histidines en position 25, 77, 139 et 199 sont sous leur forme acide (cationique, protonée), celles en position 81, 169 et 181 sont sous leur forme basique neutre, leur unique proton étant en position ϵ et l'histidine en position 217 est neutre, son unique proton étant en position δ . Le choix de la position du proton acide des acides aminés histidines neutres est fait de manière à favoriser la formation de liaisons hydrogène avec des acides aminés proches. L'orientation des chaînes latérales permet généralement de supposer la présence ou non d'un proton. Les sites de protonation des histidines neutres précédemment définis sont identiques à ceux trouvés, selon le même raisonnement par Helms *et al* [24] dans la GFP.

II.2.1.3 Distribution de protons à pH 8

La géométrie de la Cerulean à pH 8 n'est pas connue. Nous ne disposons donc que des courbes de titration précédentes, obtenues dans la géométrie à pH 5. Or ces courbes dépendent significativement de la géométrie. Le point le plus important pour notre étude est l'état de protonation de Asp148. A pH 8, sa chaîne latérale peut être à l'intérieur de la protéine, comme à pH 5, ou (plus probablement) dans le solvant. Si elle est à l'intérieur de la protéine on peut supposer que la courbe de titration précédente est valable et on en déduit alors que Asp148 est non protoné à pH 8. Si la chaîne latérale est dans le solvant, alors le $pK_{\frac{1}{2}}$ de Asp148 doit probablement être voisin du pKa standard en phase aqueuse (3.8) et donc Asp148 est également non protoné. On peut donc affirmer que Asp148 est très probablement non protoné à pH 8. Les états de protonation individuels des autres acides aminés sont moins cruciaux et nous les avons déterminés à partir des courbes de titration obtenues précédemment.

L'analyse des courbes de titration montre que peu d'acides aminés acido-basiques ont un $pK_{\frac{1}{2}}$ proche de 8. Ceci est probablement lié au fait que c'est également le cas en solution aqueuse, hormis pour les histidines. Seuls 7 acides aminés ont un $pK_{\frac{1}{2}}$ supérieur à 5.5 ou inférieur à 10.5 : Asp19, His25, Tyr39, Lys101, Glu142, Lys158 et His199. De la même manière que ce qui a été fait à pH 5, les acides aminés dont le $pK_{\frac{1}{2}}$ est inférieur à 5.5 ont été supposés non protonés alors que ceux dont le $pK_{\frac{1}{2}}$ est supérieur à 10.5 ont été définis sous leur forme protoné. De cette manière l'état de protonation de seulement 7 acides aminés reste indéterminé et le serveur H++ fournit les micro-états de plus basse énergie. Les 20 micro-états de plus basse énergie sont présentés dans le tableau II.2.

Contrairement à la situation observée à pH 5, en raison du nombre limité d'acides aminés inclus dans le calcul des micro-états de protonation, l'amplitude de variation d'énergie libre est importante. La gamme d'énergie libre a une amplitude de $1.6 \text{ kcal} \cdot \text{mol}^{-1}$, si on considère les 20 micro-états de plus basse énergie et de $3.6 \text{ kcal} \cdot \text{mol}^{-1}$ si on considère les 100 micro-états de plus basse énergie. D'autre part, seuls His25, Lys158 et His199, changent d'état de protonation dans les 20 premiers micro-états. Ces acides aminés étant éloignés du chromophore, il est raisonnable de penser que leur état de protonation influe peu sur le chromophore.

En résumé, nous avons utilisé le micro-état de protonation de plus basse énergie pour modéliser la distribution de proton à pH 8. Tous les acides aminés acides sont non protonés, tous les acides aminés basiques sont protonés. Les histidines His25 et His199 sont sous leur forme acide (cationique, doublement protonée). Les histidines en position 77, 81, 139, 169 et 181 sont neutres et leur proton acide est en position ϵ , His217 est neutre et son proton acide est en position δ .

II.2.2 Coordonnées initiales

La configuration du système choisie pour initier le calcul de la dynamique moléculaire doit représenter une configuration probable du système dans les conditions de température, de pression et de pH choisies pour la simulation.

Liste des 20 micro-états de protonation de plus basse énergie à pH 8

0 0 1 1 1 0 0 1	Energie	:-4.665819
0 1 1 1 1 0 0 1	Energie	:-4.782417
0 0 1 1 1 0 1 1	Energie	:-4.844590
0 1 1 1 1 0 1 1	Energie	:-4.946409
0 0 1 1 0 0 0 0	Energie	:-5.282663
0 0 1 1 0 0 0 0	Energie	:-5.357990
0 1 1 1 0 0 0 0	Energie	:-5.427706
0 1 1 1 0 0 0 0	Energie	:-5.501660
0 0 1 1 0 0 0 1	Energie	:-5.608031
0 1 1 1 0 0 0 1	Energie	:-5.743129
0 0 1 1 0 0 1 1	Energie	:-5.745722
0 0 1 1 0 0 1 0	Energie	:-5.761141
0 0 1 1 0 0 1 0	Energie	:-5.791964
0 1 1 1 0 0 1 1	Energie	:-5.866040
0 1 1 1 0 0 1 0	Energie	:-5.891404
0 0 1 1 0 0 0 1	Energie	:-5.920547
0 1 1 1 0 0 1 0	Energie	:-5.920855
0 1 1 1 0 0 0 1	Energie	:-6.057017
0 0 1 1 0 0 1 1	Energie	:-6.102741
0 1 1 1 0 0 1 1	Energie	:-6.224432

Tableau II.2 – Liste des 20 micro-états de protonation de plus basse énergie à pH 8. Chaque ligne représente un micro-état dans lequel 0 signifie que l'acide aminé est non protoné et 1 signifie qu'il est protoné. Chaque colonne correspond à un acide aminé. Ils sont donnés par numéro croissant : Asp19, His25, Tyr39, Lys101, Glu142, Asp148, Lys158, His199. La dernière colonne est l'énergie libre associée au micro-état en kcal·mol⁻¹.

II.2.2.1 Simulations à pH 5 et à pH 8 en conformation *cis*

La Cerulean a été cristallisée à pH 5 (code 1Q57 de la *Protein Data Bank* [37]) et sa structure a été déterminée avec une résolution de 2 Å. Bien que faible, cette résolution est suffisante pour pouvoir utiliser la structure pour faire une simulation, car la position des atomes lourds (C, N, O et S) est bien identifiée. Le chromophore est caractérisé dans sa conformation *cis*. Cette structure a été utilisée comme point de départ des simulations dans les cas pH5-*cis* et pH8-*cis*. L'obtention d'un point de départ pour la simulation pH8-*trans* nécessite une étude spécifique.

Afin de compléter la structure initiale on ajoute les 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 *Protein Data Bank* aux structures et géométries contenues dans les bibliothèques de chaque acide aminé. L'état de protonation le plus probable des acides aminés acido-basiques défini précédemment est utilisé pour chaque pH.

II.2.2.2 Simulation à pH 8 en conformation *trans*

Aucune structure RX de la Cerulean avec le chromophore en conformation *trans* n'est encore actuellement disponible. Nous avons utilisé les résultats de la simulation réalisée dans le cas pH8-*cis* pour obtenir une configuration de départ utilisable dans le cas pH8-*trans*. Dans un premier temps nous avons extrait un ensemble de configurations (1150 configurations) de la simulation réalisée dans le cas pH8-*cis*. Pour chacune de ces configurations, nous avons modifié les coordonnées du chromophore pour qu'il adopte une conformation *trans* (par une rotation de l'angle de torsion φ de 180°) et nous avons identifié les atomes de la protéine situés à moins de 2 Å d'un atome du cycle indole. Une distance inférieure à 2 Å indique généralement une forte interaction répulsive.

Parmi tous les atomes du cycle indole, seulement quelques uns ont des distances inférieures à 2 Å avec des atomes d'acides aminés voisins du chromophore. Ils correspondent à la partie la plus externe du cycle indole (plus précisément du cycle benzène) qui est celle qui se déplace le plus lors d'une isomérisation de la conformation *cis* à la conformation *trans*. Ainsi les atomes C3, H3, C4 et H4 (figure I.6 page 20 pour le nom des atomes) se placent, dans 20 à 50% des configurations, à des distances inférieures à 2 Å de Val150, Phe165 et Asp148, essentiellement des atomes de groupements méthyle ou de chaînes aliphatiques. D'autres atomes, C2, H2, H5 et H8 se trouvent également à des distances inférieures à 2 Å d'atomes d'acides aminés proches mais seulement dans 10% des configurations au plus.

La gêne stérique engendrée par l'isomérisation du chromophore est donc relativement limitée. De plus, les répulsions entre le chromophore et les atomes des acides aminés voisins sont facilement supprimées par des mouvements de torsion de petite amplitude des chaînes aliphatiques, ce qui ne nécessite pas une déformation globale de la protéine. La cavité de la protéine qui contient le chromophore en conformation *cis* semble donc facilement s'adapter au chromophore en conformation *trans*. Ainsi, quelque soit le choix de la configuration de la simulation à pH 8 avec le chromophore en conformation *cis* utilisée pour créer le point de départ de la simulation pH8-*trans*, les résultats de la trajectoire à l'équilibre devraient être identiques. En conséquence, les coordonnées de départ pour la simulation pH8-*trans* ont été obtenues en changeant la conformation du chromophore dans une configuration extraite, au hasard, de la simulation à pH 8 avec le chromophore en conformation *cis*. L'énergie du système a ensuite été minimisée puis, en suivant le protocole indiqué au paragraphe B.2.6 page 173, le système a été porté à température ambiante et à pression atmosphérique avant de faire une simulation de 3 ns dans l'ensemble (N,P,T).

II.2.3 Structure et dynamique de la Cerulean à pH 5

A partir de la structure cristalline et du micro-état de protonation choisi pour représenter l'état de protonation de la protéine à pH 5, des simulations de dynamique moléculaire ont été réalisées à pH 5 en conformation *cis*. L'analyse de la structure RX conduit Malo *et al* [37] à proposer que Asp148 est sous sa forme acide et, en raison de l'existence possible d'une liaison hydrogène, il conclut sur la position préférentielle du proton acide sur un des atomes d'oxygène de la fonction acide carboxylique de Asp148. Dans nos simulations, nous avons choisi d'étudier systématiquement les

différentes situations de départ selon la position du proton acide de Asp148 à partir de la structure RX.

Après avoir présenté les 4 situations initiales possibles et vérifié la stabilité de la structure au cours des simulations, la structure d'équilibre sera analysée et comparée aux données expérimentales. Nous décrirons en particulier l'orientation de la chaîne latérale de l'acide aminé Asp148, le réseau de liaisons hydrogène impliquant le chromophore et Asp148 et la dynamique des brins du cylindre de feuillets β formant le *cleft*.

II.2.3.1 Position initiale du proton acide sur Asp148

Sur une fonction acide carboxylique le proton acide peut être placé de plusieurs manières : sur l'un ou l'autre des atomes d'oxygène de la fonction acide carboxylique et, sur chaque atome d'oxygène, il existe deux minima correspondant à l'orientation du proton, en direction de l'autre atome d'oxygène du groupe COOH ou orienté dans la direction opposée. Cela conduit aux quatre possibilités qui sont présentées sur la figure II.5.

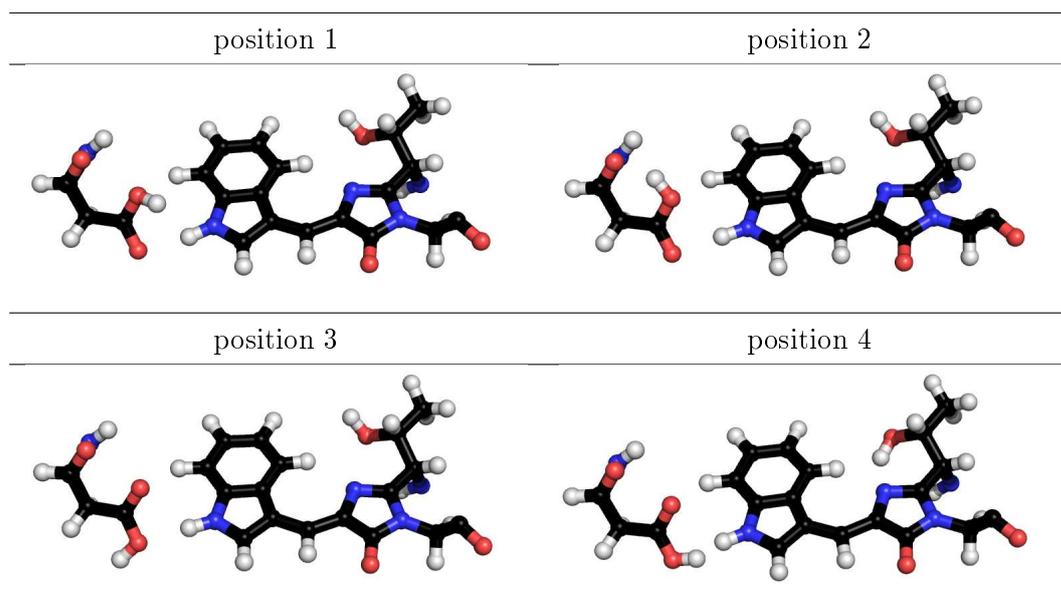


Figure II.5 – Les quatre positions initiales envisagées pour le proton acide de Asp148.

La position 1 est obtenue à partir du modèle utilisé par AMBER par défaut lorsqu'il ajoute un hydrogène sur le groupement carboxylique d'un acide aspartique. La position 2 est en accord avec celle proposée par Malo *et al* [37] dans laquelle une liaison hydrogène intra-acide aminé existe entre l'oxygène du squelette protéique de Asp148 et l'oxygène de la fonction acide carboxylique qui en est le plus proche (figure II.2 page 53). Les positions 3 et 4 correspondent à la protonation de l'autre atome d'oxygène de la fonction acide carboxylique. Conventionnellement les atomes d'oxygène de la fonction acide carboxylique d'un acide aspartique s'appellent OD1 et OD2. Cette notation n'est pas adaptée à notre étude car il est préférable dans notre cas de pouvoir suivre la dynamique des atomes d'oxygène en sachant lequel porte le proton acide indépendamment de leur position dans la

structure RX de départ. Pour clarifier et simplifier la discussion, nous appellerons par la suite ODH l'atome d'oxygène de la fonction acide carboxylique qui porte le proton acide, OD l'autre atome d'oxygène de cette fonction et HD le proton acide. Les noms des atomes de Asp148 sont résumés sur la figure II.6.

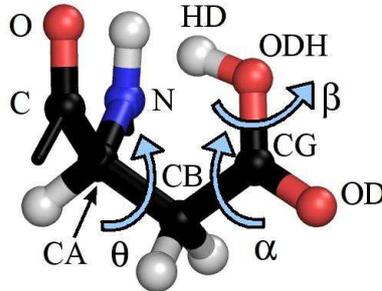


Figure II.6 – Noms des atomes employés pour Asp148. Les angles θ , α et β correspondent respectivement aux torsions autour des liaisons CA-CB, CB-CG et CG-ODH. La torsion θ est mesurée par l'angle dièdre C-CA-CB-CG, la torsion α est mesurée par l'angle dièdre CA-CG-CB-ODH et la torsion β est mesurée par l'angle dièdre CB-CG-ODH-HD.

Pour les autres acides aminés acide aspartique ou acide glutamique protonés, le modèle contenu dans les bibliothèques de AMBER a été utilisé. Une simulation a été réalisée pour chacune des quatre positions initiales du proton acide de Asp148 présentées sur la figure II.5. Par la suite la simulation ayant comme point de départ la position i sera nommée simulation i . Pour chaque cas, nous avons minimisé l'énergie, puis le système a été progressivement porté à température ambiante et à pression atmosphérique en suivant les étapes décrites dans le paragraphe B.2.6 page 173. Une simulation de 4 ns dans l'ensemble (N,P,T) à 300 K et sous une pression de 1 bar a ensuite été réalisée. Les résultats présentés par la suite ont été extraits des 3.5 dernières nanosecondes des simulations. Les 500 premières picosecondes ont pour intérêt de laisser le système relaxer vers sa situation d'équilibre et ne sont pas prises en compte.

II.2.3.2 Stabilité de la structure au cours des simulations

L'amplitude des déformations de la protéine dans son ensemble au cours d'une simulation de dynamique moléculaire peut être mesurée au travers de l'évolution en fonction du temps de la moyenne du déplacement carré moyen des atomes de la protéine, RMSd, équation II.1 :

$$\text{RMSd} = \sqrt{\frac{1}{M_{\text{totale}}} \sum_{iat}^{N_{\text{atome}}} m_{iat} (\vec{r}_{iat}^c - \vec{r}_{iat}^{\text{ref}})^2} \quad (\text{II.1})$$

où M_{totale} et N_{atome} sont la masse totale et le nombre des atomes inclus dans le calcul du RMSd, iat représente l'indice d'un atome, m_{iat} est la masse de l'atome iat , c représente l'indice d'une configuration extraite de la simulation, ref représente l'indice de la configuration de référence par rapport à laquelle on mesure le déplacement et \vec{r}_{iat} est le vecteur position de l'atome iat dans une configuration donnée. Généralement la structure de référence est la structure cristalline. Le calcul

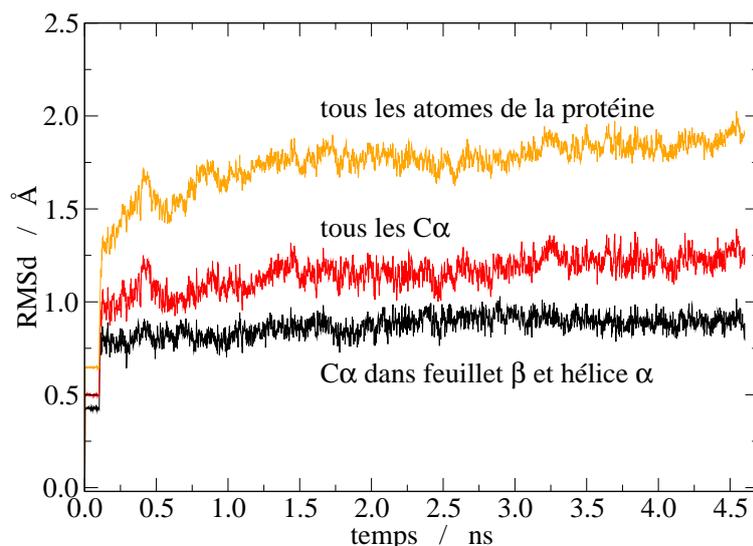


Figure II.7 – Déplacement carré moyen en fonction du temps des C_{α} des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β en noir, de l'ensemble des C_{α} en rouge et de l'ensemble des atomes de la protéine en orange dans la simulation 1.

du RMSd indique alors si au cours de la simulation la structure de la protéine dévie fortement de la structure cristalline. Une déviation trop importante est généralement le signe d'une trajectoire non physique comme par exemple un dépliement partiel de la protéine. Dans le cas où des atomes de nature différente sont inclus dans le calcul du RMSd, celui-ci est généralement pondéré par la masse des atomes comme indiqué dans l'équation II.1. La figure II.7 présente l'évolution du RMSd en fonction du temps pour trois groupes d'atomes dans le cas de la simulation 1.

Le calcul du RMSd présenté sur la figure II.7 a été calculé sur l'ensemble de la simulation 1 en incluant les parties de mise à l'équilibre de la température et de la pression. Les trois courbes obtenues (i) pour l'ensemble des C_{α} constituant les feuillets β et l'hélice α (ii) pour l'ensemble des C_{α} et (iii) pour l'ensemble des atomes de la protéine ont une allure similaire. Elles présentent un premier palier entre 0 et 0.2 ns qui correspond à la mise à l'équilibre de la température pendant laquelle des contraintes sont imposées sur les atomes de la protéine. Ces contraintes ont clairement pour effet de limiter le déplacement des atomes pendant cette étape. A partir du moment où les contraintes sont relâchées le RMSd augmente puis reste à peu près constant pour les trois groupes d'atomes ce qui indique que la structure est stable tout au long de la simulation (la dérive est de l'ordre de 0.3 Å pour l'ensemble des C_{α}).

Le RMSd correspondant aux C_{α} inclus dans une structure secondaire (courbe noire) est le plus faible. Ces atomes correspondent à la partie rigide de la protéine, ils sont contraints par les liaisons covalentes et les liaisons hydrogène qui maintiennent les structures secondaires. Le RMSd de l'ensemble des C_{α} , courbe rouge, est légèrement plus élevé, dû à la présence de boucles souples dans la structure. Le RMSd des C_{α} oscille entre 0.7 Å et 1.3 Å sur l'ensemble de la simulation. Ces valeurs sont en bon accord avec celles obtenues pour d'autres simulations de la GFP ou de mutants [69, 24, 75] et indiquent que l'ensemble de la structure est conservée tout au long des simu-

lations. Les valeurs de RMSd pour les mêmes groupes d'atomes et pour les trois autres simulations sont similaires et ne sont pas présentés ici.

II.2.3.3 Conformations de la chaîne latérale de Asp148

Les conformations de la chaîne latérale de Asp148 peuvent être caractérisées en considérant les valeurs moyennes et les fluctuations de plusieurs torsions autour des liaisons de la chaîne latérale ou de distances qui sont données dans le tableau II.3.

	β	α	θ	O-ODH	O-OD
référence [37]	/	166	-53	2.9	4.5
simulation 1	0 (11)	3 (22)	-54 (11)	5.0 (0.2)	3.3 (0.2)
simulation 2	-7 (11)	3 (18)	-55 (9)	4.9 (0.2)	3.2 (0.2)
simulation 3	-1 (10)	9 (18)	-56 (9)	5.1 (0.2)	3.3 (0.2)
simulation 4	177 (12)	200 (84)	66 (11)	4.6 (0.4)	4.4 (0.4)

Tableau II.3 – Valeur moyenne et déviation standard (entre parenthèses) des torsions (en degrés) et des distances (en Å) qui permettent de caractériser les conformations de la chaîne latérale de Asp148 obtenues le long des quatre simulations effectuées à pH 5. Les noms des atomes et les torsions sont définis sur la figure II.6. Les données de la ligne *référence* sont extraites de la structure RX de la Cerulean. Sur cette ligne l'atome d'oxygène ODH est défini comme étant celui qui est supposé porter un proton dans la structure RX.

Il apparaît dans ce tableau que les valeurs moyennes et les fluctuations des différentes coordonnées internes caractérisant la dynamique et la position moyenne de la chaîne latérale de Asp148 sont très similaires dans les simulations 1,2 et 3, et très différentes dans la simulation 4. Les résultats de cette dernière simulation seront discutés séparément.

a) Simulation 1 à 3

Concernant les simulations 1 à 3, les valeurs moyennes de l'angle θ sont similaires dans les trois simulations et proches de la valeur mesurée dans la structure RX. Les valeurs de θ caractérisent le déplacement du groupe COOH de Asp148. Des valeurs de l'ordre de -55° correspondent à une position de la fonction acide carboxylique à l'intérieur de la protéine et dirigée vers l'atome N9 de l'indole du chromophore.

Les valeurs moyennes des distances O-ODH et O-OD ainsi que celle de l'angle de torsion α sont très proches les unes des autres et significativement différentes des valeurs mesurées dans la structure cristalline. L'angle de torsion α décrit la rotation du groupement COOH autour de la liaison $H_2C-COOH$. La géométrie d'équilibre dans les simulations correspond à une situation dans laquelle les atomes d'oxygènes OD et ODH ont échangé leur position par rapport à la structure RX. Pour les simulations 1 et 2 (positions de départ 1 et 2, figure II.5) la position de départ de l'atome ODH est proche de l'oxygène du squelette protéique, la déformation conduisant à la situation d'équilibre

observée dans les simulations est principalement une variation de 180° de l'angle α (figure II.8). Dans la simulation 3, la position initiale de la fonction acide carboxylique est proche de la situation d'équilibre obtenue dans les simulations. De ce fait α conserve la même valeur.

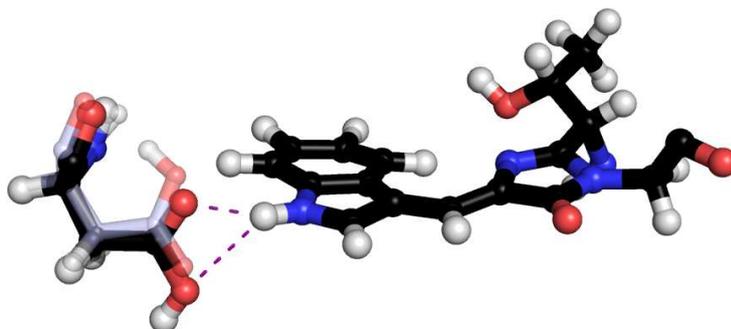


Figure II.8 – En noir, orientation de la chaîne latérale de Asp148 dans les simulations 1, 2 et 3. Il n'y a pas de liaison intra-acide aminé entre l'acide carboxylique et le squelette protéique de Asp148, l'hydrogène HD est orienté à l'extérieur du groupe carboxylique et les deux atomes d'oxygène du groupe carboxylique sont proches du groupe NH de l'indole du chromophore. La géométrie bleu transparente est issue de la structure cristalline, le proton acide étant positionné selon l'hypothèse d'une liaison hydrogène intra-acide aminé.

Un dernier point concerne l'angle de torsion β . Cet angle caractérise la position du proton HD par rapport à l'atome OD : soit la liaison ODH-HD est orientée à l'intérieur du groupe COO, en direction de OD ($\beta = 180^\circ$), soit dans la direction opposée ($\beta = 0^\circ$). Dans les simulations 1 à 3, la valeur de β est proche de zéro. Le proton HD est donc dans la direction opposé à OD, c'est à dire à l'extérieur du groupement acide carboxylique. Cette position de l'atome HD est stabilisée par la formation d'une liaison hydrogène avec un acide aminé proche, la lysine 166 (figure II.9).

En résumé, pour les simulations 1 à 3, quelque soit la position du proton dans la géométrie de départ, la conformation d'équilibre de la chaîne latérale de Asp148 est celle présentée sur la figure II.8 : la chaîne latérale de Asp148 est à l'intérieur de la protéine, l'atome ODH est l'atome d'oxygène de la fonction acide carboxylique qui est le plus éloigné de l'atome d'oxygène du squelette protéique, la liaison hydrogène intra-acide aminé n'est pas observée et l'hydrogène HD est orienté à l'extérieur de la fonction acide carboxylique. La différence entre la conformation observée expérimentalement et celle issue de la simulation sera analysée à l'aide du réseau de liaison hydrogène et de la structure du *cleft* (ouverture entre deux brins du tonneau de feuillets β de la protéine).

b) Simulation 4

Les valeurs moyennes des coordonnées internes de Asp148, présentées dans le tableau II.3, sont différentes pour cette simulation de celles obtenues pour les autres simulations et la structure RX. La situation moyenne obtenue dans la simulation 4 se caractérise par une valeur moyenne de l'angle de torsion θ de 66° . La chaîne latérale de Asp148 s'oriente vers l'extérieur de la protéine par une rotation autour de la liaison CA-CB (angle de torsion θ), les atomes C, CA, N restant fixes du fait des contraintes dues aux liaisons hydrogène impliquant le squelette protéique de Asp148 dans un feuillet β .

Il résulte de ce changement d'environnement que la dynamique de la chaîne latérale de Asp148 est plus libre, ce qui se traduit par des fluctuations au moins deux fois plus importantes dans cette simulation que dans les trois autres. Les fluctuations de l'angle de torsion α sont très importantes car au cours de la simulation le plan du groupe acide carboxylique tourne plusieurs fois de 180° . Enfin, contrairement aux autres simulations, l'atome d'hydrogène HD est à l'intérieur du groupe acide carboxylique. Dans cette simulation, le groupe COOH étant dans le solvant il n'y a plus de raisons structurales, comme la formation d'une liaison hydrogène qui nécessitent que HD soit à l'extérieur de la fonction acide carboxylique.

Les résultats obtenus à partir de cette dernière simulation sont à considérer avec réserve. Il faut remarquer que le point de départ de la simulation 4 (figure II.5) est énergétiquement défavorable à cause de la courte distance entre le proton HD de Asp148 et le proton H9 du chromophore (1.22 \AA). Malgré les précautions prises lors de la mise à l'équilibre du système, cette courte distance induit probablement de fortes répulsions et donc une vitesse initiale élevée pouvant conduire le système dans des régions de l'espace des configuration en réalité inaccessibles. Cette simulation montre cependant qu'à pH 5 il existe un bassin d'énergie libre dans lequel la chaîne latérale de Asp148 se place à l'extérieur de la protéine. On peut alors se poser les deux questions suivantes : quel est le poids des configurations explorées par la simulation 4 par rapport aux poids des configurations explorées dans les simulations 1, 2 ou 3 ? et sont-elles thermiquement accessibles ? Pour répondre à ces questions il faudrait évaluer la différence d'énergie libre entre la région de l'espace des configurations explorée dans la simulation 4 et celle explorée dans les simulations 1, 2 et 3, ainsi que la barrière qui sépare ces deux régions. Le calcul de la différence d'énergie libre est en cours. Dans ce qui suit, compte tenu de la position de Asp148 clairement identifiée à l'intérieur du tonneau dans la structure RX à pH 5, la simulation 4 ne sera pas prise en compte.

II.2.3.4 Réseau de liaisons hydrogène au voisinage du chromophore

Le réseau de liaisons hydrogène au voisinage du chromophore décrit la façon dont celui-ci est maintenu au sein de la protéine. Le réseau de liaisons hydrogène 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 d'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 que ce sont les chemins les plus probables des transferts de protons ou d'électrons.

Le réseau de liaisons hydrogène a été calculé sur les 3.5 dernières nanosecondes de chacune des simulations 1 à 3 en utilisant les critères décrits au paragraphe B.2.5 page 173. La figure II.9 présente le réseau de liaisons hydrogène obtenu à partir de la trajectoire de la simulation 1. Les résultats obtenus pour les simulations 2 et 3 sont très similaires et ne sont pas présentés.

Le réseau de liaisons hydrogène impliquant le cycle imidazolinone, la fonction alcool de la partie Thr65 du chromophore et les acides aminés Glu222, Ser205 et Thr203 est très similaire à celui décrit dans la structure RX mais aussi dans les protéines de la famille de la GFP en général [69, 24].

Les atomes d'oxygène de la fonction acide carboxylique de Asp148 sont en interaction avec la

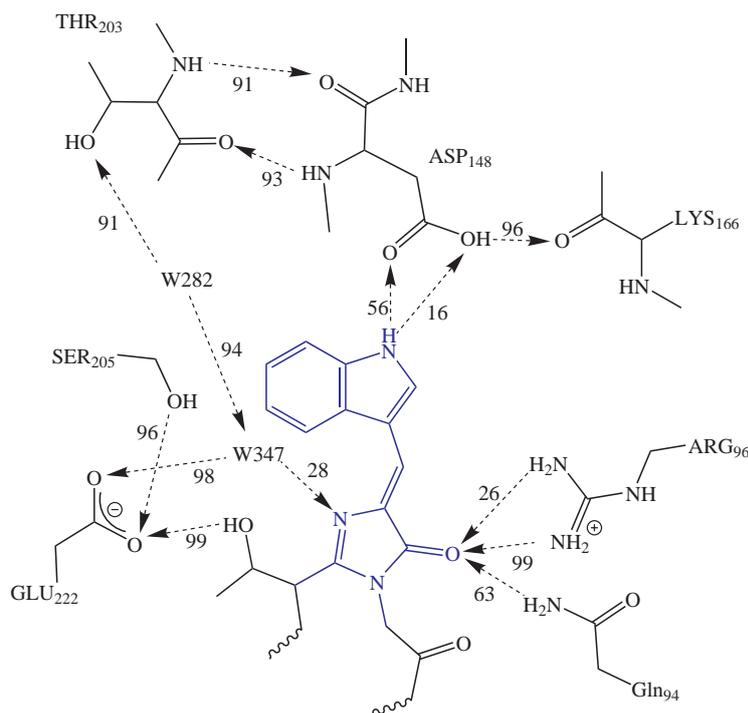


Figure II.9 – Réseau de liaisons hydrogène dans l’environnement du chromophore obtenu à partir de la simulation 1 à pH 5. Les flèches indiquent la présence d’une liaison hydrogène et sont orientées du donneur vers l’accepteur d’hydrogène. Les nombres associés à chaque liaison hydrogène indiquent en pourcentage le temps pendant lequel la liaison hydrogène existe. Les critères utilisés pour déterminer l’existence d’une liaison hydrogène sont de 3.2 Å pour la distance et 40° pour l’angle.

fonction indole du chromophore, comme cela a été proposé par Malo *et al* [37]. Le point le plus remarquable concerne le groupement OH acide de Asp148. Sa position est différente de celle de la structure RX et le proton HD s’oriente à l’extérieur du groupement COOH. Cette conformation est stabilisée par une liaison hydrogène très stable avec le groupement carbonyle du squelette protéique de Lys166. Le réseau de liaisons hydrogène de la fonction acide carboxylique, très stable au cours de la simulation, justifie la convergence vers la même structure d’équilibre des 3 simulations 1, 2 et 3. Ces résultats soulèvent deux questions, l’une concernant l’état de protonation de Asp148, l’autre concernant la dynamique du brin portant l’acide aminé Lys166 (paragraphe II.2.4).

II.2.3.5 Etat de protonation de Asp148

Bien que les acides aspartiques soient en général sous leur forme basique à pH 5 [76], deux raisons nous ont amenés à considérer la forme acide, protonée, de Asp148. Ces raisons sont d’une part l’hypothèse d’une liaison hydrogène intra-acide aminé $O_{148}-ODH$ faite par Malo *et al*, et d’autre part, les résultats des calculs H++. Les calculs effectués en utilisant H++ prédisent que Asp148 est sous sa forme acide dans la structure RX à pH 5. Ce résultat est probablement dû au fait que, dans la géométrie de la structure RX, la distance $O_{148}-ODH$ est courte, ce qui stabilise la présence d’un proton. Or cette liaison hydrogène intra-acide aminé n’est pas observée dans nos simulations.

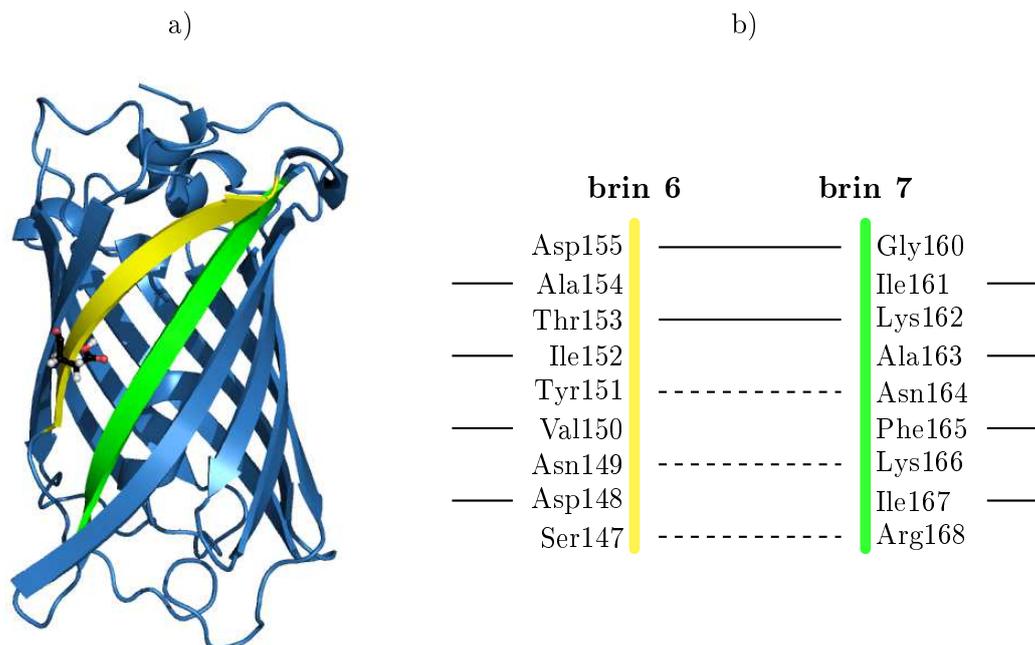


Figure II.10 – a) Représentation du *cleft* entre les brins 6 (en jaune) et 7 (en vert) dans la Cerulean. La chaîne latérale de Asp148 (brin 6) est également représentée. b) Schéma représentant les acides aminés constituant les brins 6 et 7 et les liaisons entre les fonctions amides de chaque paire d'acides aminés. Les traits pleins représentent les liaisons hydrogène présentes dans la structure RX. Les tirets représentent les liaisons hydrogène manquantes.

Faut-il pour autant considérer Asp148 sous sa forme basique ? Nos résultats suggèrent que non car la liaison hydrogène entre l'atome ODH de Asp148 et le carbonyle de Lys166 est observée 96% du temps de la simulation (97% et 98% dans les simulations 2 et 3). La présence d'un proton sur Asp148 est donc justifiée au même titre que précédemment. Asp148 est donc certainement sous sa forme acide à pH 5 bien que cela ne soit pas la conséquence de la liaison hydrogène intra-acide aminé.

Les acides aminés Asp148 et Lys166 sont séparés par une molécule d'eau dans la structure RX. La formation de la liaison hydrogène entre Asp148 et Lys166 implique un déplacement de Lys166 vers Asp148, et de la molécule d'eau. Nous avons étudié plus précisément les déformations du brin 6 qui porte Asp148 et du brin 7 qui porte Lys166, ces deux brins constituant un des 11 feuillets β de la Cerulean.

II.2.4 Dynamique globale de la Cerulean - mouvement des brins 6 et 7

Dans la Cerulean, le cylindre de feuillets β formant le tonneau de la protéine présente une ouverture entre les brins 6 et 7, que nous appelons *cleft*. Entre ces brins, plusieurs paires d'acides aminés ne forment pas les liaisons hydrogène habituelles entre deux brins formant un feuillet β (voir la structure d'un feuillet β , figure A.4 page 161), la distance entre les brins est trop importante ce qui crée une ouverture dans le cylindre (figure II.10).

Cette faiblesse de la structure est également présente dans les structures RX de la GFP [24, 18] et celles d'une grande partie de ses mutants, tels que la EYFP [29] et la ECFP [67]. Cependant

l'ouverture est significativement plus importante dans la structure RX de la Cerulean que dans celles de la GFP, de la EYFP ou de la ECFP.

II.2.4.1 Description de l'ouverture du cylindre de feuillet β entre les brins 6 et 7

Le tableau II.4 donne les distances entre les atomes d'azote du groupe NH et l'oxygène du carbonyle du squelette protéique des paires d'acides aminés des brins 6 et 7 (figure A.4 page 161).

		Thr153 ^a -Lys162		Tyr151-Asn164		Asn149-Lys166		Ser147-Arg168	
1	GFP [18]	3.0	2.7	3.1	3.1	6.0	5.1	6.7	6.7
2	GFP/H148D [77]	3.0	2.9	5.0	4.5	6.8	6.4	7.3	7.3
3	EYFP [29]	3.0	2.7	3.2	3.2	6.1	5.2	6.6	6.8
4	EYFP/H148G [29]	3.2	3.2	4.6	3.2	7.3	5.3	7.1	6.2
5	ECFP(A') [67]	2.9	2.9	3.9	3.4	5.9	5.1	4.5	7.0
6	ECFP(B') [67]	2.9	2.8	3.4	3.9	5.6	5.1	6.5	6.8
7	Cerulean [37]	3.0	2.8	5.0	4.3	6.8	7.0	7.5	6.9
8	Simulation 1	3.0	2.9	3.4	3.2	5.6	5.4	5.7	5.4

^aMet153 dans la GFP, la GFP/H148D, la EYFP et la EYFP/H148G

Tableau II.4 – Distances, en Å, entre l'atome d'oxygène du carbonyle du squelette protéique et l'atome d'azote de la fonction amide des paires d'acides aminés constituant les brins 6 et 7, (figure II.10 b page 69). Chaque paire d'acides aminés peut former deux liaisons hydrogène (figure A.4 page 161). La première valeur est la distance entre l'atome d'azote du premier acide aminé et l'atome d'oxygène du second acide aminé, la seconde valeur est la distance entre l'atome d'oxygène du premier acide aminé et l'azote du second acide aminé. Les lignes 1 à 6 concernent les distances mesurées dans la structure RX de la référence indiquée. Les valeurs de la ligne 7 sont les moyennes obtenues pour chaque distance dans la simulation 1.

En se basant sur le critère selon lequel une liaison hydrogène existe si la distance entre l'atome donneur d'hydrogène et l'atome accepteur d'hydrogène est inférieure à 3.2 Å, quelque soit la protéine, les deux liaisons hydrogène entre les paires d'acides aminés Asp155 - Gly160 et Thr153 - Lys162 sont conservées et les deux liaisons hydrogène entre les paires d'acides aminés Ser147 - Arg168 et Asn149 - Lys166 sont rompues. La présence des deux liaisons hydrogène entre Tyr151 et Asn164 dépend de la protéine considérée. Dans la structure RX de la Cerulean et de la GFP/H148D (mutation de His148 en Asp148), elles sont rompues alors qu'elles sont présentes dans la EYFP et la GFP. Les structures RX de la ECFP et la EYFP/H148G (mutation de His148 en Gly148) présentent une situation intermédiaire dans laquelle une des deux liaisons hydrogène existe et l'autre est rompue.

L'origine du *cleft* dans la structure RX de la Cerulean pourrait être associée aux mutations du chromophore et de l'acide aminé en position 148. Le chromophore des GFPs est fixé sur l'hélice α au centre de la protéine et s'étend vers la surface de la protéine, son extrémité étant située en face du feuillet formé par les brins 6 et 7. Dans le cas de la ECFP et de la Cerulean, le cycle phénol est substitué par un cycle indole plus étendu. Les effets stériques induits par le cycle indole

pourrait avoir comme effet de déstabiliser le réseau de liaisons hydrogène entre les brins 6 et 7. Cela peut expliquer que dans la Cerulean et les deux conformations de la ECFP, le *cleft* soit plus large que dans la EYFP et la GFP. On remarque que la mutation H148G dans la EYFP déstabilise également le réseau de liaisons hydrogène entre les brins 6 et 7. Or la Cerulean est un variant de la ECFP obtenu notamment par une mutation en position 148. La Cerulean cumule donc deux caractéristiques susceptibles de déstabiliser le réseau de liaisons hydrogène entre les brins 6 et 7 : un chromophore plus grand et une mutation en position 148.

II.2.4.2 Dynamique des brins 6 et 7 dans les simulations

La dernière ligne du tableau II.4 donne les valeurs moyennes, obtenues dans la simulation 1, des distances entre les atomes d'azote et d'oxygène des acides aminés portés par les brins 6 et 7. Ces valeurs doivent être comparées à celle de la ligne au dessus qui sont celles mesurées sur la structure RX et correspondent donc aux valeurs initiales de ces distances. Les valeurs moyennes obtenues dans la simulation sont toutes inférieures, de 1.2 à 1.8 Å, aux valeurs mesurées sur la structure RX. De plus les déviations standard de ces distances dans la simulation 1 sont faibles, elles sont comprises entre 0.1 et 0.6 Å. Le rapprochement des brins 6 et 7 est donc conservé tout le long de la simulation. Cette situation est observée quelque soit la simulation (simulation 1 à 3).

Ce rapprochement des brins 6 et 7 rend effective les liaisons hydrogène entre les atomes du squelette protéique des acides aminés Tyr151 et Asn164. La liaison hydrogène entre l'atome d'azote de Tyr151 et l'atome d'oxygène de Asn164 est présente 39% du temps et celle entre l'atome d'oxygène de Tyr151 et l'atome d'azote de Asn164 est présente 63% du temps dans la simulation 1. La situation moyenne obtenue dans les simulations est en fait celle observée dans les structure RX des autres GFP (tableau II.4). De plus, le rapprochement des brins 6 et 7 positionne le groupement carbonyle de Lys166 à une distance favorable pour former une liaison hydrogène avec l'atome ODH de Asp148 (figure II.9). Ces deux liaisons hydrogène stabilisent la région entre les brins 6 et 7. On peut donc conclure que la présence d'un *cleft* plus large dans la Cerulean que dans les autres GFP, n'est pas la conséquence d'effets stériques dus à la taille du chromophore et à la mutation H148D. Si c'était le cas, le rapprochement des brins ne serait pas observé dans la dynamique.

Dans l'ensemble de nos simulations, le rapprochement des brins 6 et 7 se produit dans les 100 premières picosecondes qui suivent le relâchement des contraintes (lors de la mise à l'équilibre de la protéine paragraphe B.2.6 page 173). Nous avons donc analysé plus en détail les trajectoires de nos simulations pendant l'augmentation de la pression et de la température et nous avons testé trois protocoles de mise à l'équilibre afin de voir si le rapprochement des brins 6 et 7 se manifeste indépendamment du protocole utilisé. Les trois protocoles débutent tous par une augmentation de la température du système au cours d'une simulation dans l'ensemble (N,V,T) avec des contraintes harmoniques sur les atomes de la protéine. Cette simulation se déroule à basse pression (la densité vaut environ 0.8). Les protocoles diffèrent par l'instant où les contraintes sont retirées. Cet instant sera défini par temps $t=0$:

Protocole 1 A l'issue de la simulation (N,V,T) avec contraintes, à basse pression, les contraintes

sont supprimées et la pression est augmentée jusqu'à 1 atm lors d'une simulation dans l'ensemble (N,P,T) sans contraintes. Ce protocole est le protocole utilisé pour l'ensemble de nos simulations et décrit au paragraphe B.2.6 page 173.

Protocole 2 A l'issue de la simulation (N,V,T) avec contraintes, à basse pression, les contraintes sont supprimées et on poursuit la simulation dans l'ensemble (N,V,T), à basse pression, pendant 200 ps. Ensuite, on augmente la pression jusqu'à 1 atm en faisant une simulation dans l'ensemble (N,P,T) sans contraintes.

Protocole 3 A l'issue de la simulation (N,V,T) avec contraintes, à basse pression, on fait une simulation dans l'ensemble (N,P,T) avec les mêmes contraintes, au cours de laquelle on augmente la pression jusqu'à 1 atm. Les contraintes sont ensuite relaxées au cours d'une simulation dans l'ensemble (N,P,T) sans contraintes.

Nous avons utilisé ces trois protocoles dans le cas de la simulation 1. La figure II.11 présente les trajectoires des distances entre l'atome d'oxygène ODH de Asp148 et l'atome d'oxygène du carbonyle de Lys166 d'une part, et la moyenne des trajectoires des deux longueurs des deux liaisons hydrogène entre Tyr151 et Asn164 dans la simulation 1 d'autre part, pour chacun des trois protocoles. La variation de la densité du système est également représentée. Nous avons obtenu des courbes similaires pour les autres simulations.

Le rapprochement des brins 6 et 7 et l'augmentation de la densité liée à la mise à l'équilibre de la pression ont lieu simultanément lorsqu'on utilise le *protocole 1*. L'augmentation de la densité a lieu après le rapprochement des brins 6 et 7 lorsqu'on utilise le *protocole 2* et avant le rapprochement des brins 6 et 7 lorsqu'on utilise le *protocole 3* (figure II.11). Ces résultats montrent que le rapprochement des brins 6 et 7 est indépendant du protocole utilisé et de l'augmentation de la densité dans la boîte de simulation. Il a lieu dès que l'on relâche les contraintes.

Ces observations sur la dynamique globale des feuillettes de la Cerulean montrent que les différences observées sur le réseau de liaisons hydrogène dans le voisinage du chromophore, entre la structure RX et nos simulations, sont la conséquence d'une déformation de la structure de la protéine et pas uniquement du déplacement de la chaîne latérale de Asp148. Dans la structure RX, la distance entre ODH de Asp148 et l'atome d'oxygène du carbonyle de Lys166 est trop importante pour qu'une liaison hydrogène puisse se former. Le rapprochement des brins 6 et 7 réduit cette distance et favorise la formation de cette liaison au détriment de la liaison hydrogène intra-acide aminé entre ODH et l'atome d'oxygène du carbonyle de Asp148.

II.2.4.3 Origine de la différence de positions des brins 6 et 7 entre les structures expérimentale et théorique.

Les différences observées entre la structure RX et la structure moyenne obtenue à partir de nos simulations peuvent avoir principalement trois origines : un défaut du champ de force utilisé, de la dynamique moléculaire ou un effet d'empilement cristallin (la structure RX est obtenue à partir d'une phase solide à 100 K, alors que les simulations de la protéine ont été faites en phase aqueuse à 300 K) . Pour tester les deux premières hypothèses, nous avons analysé les dynamiques du *cleft* de

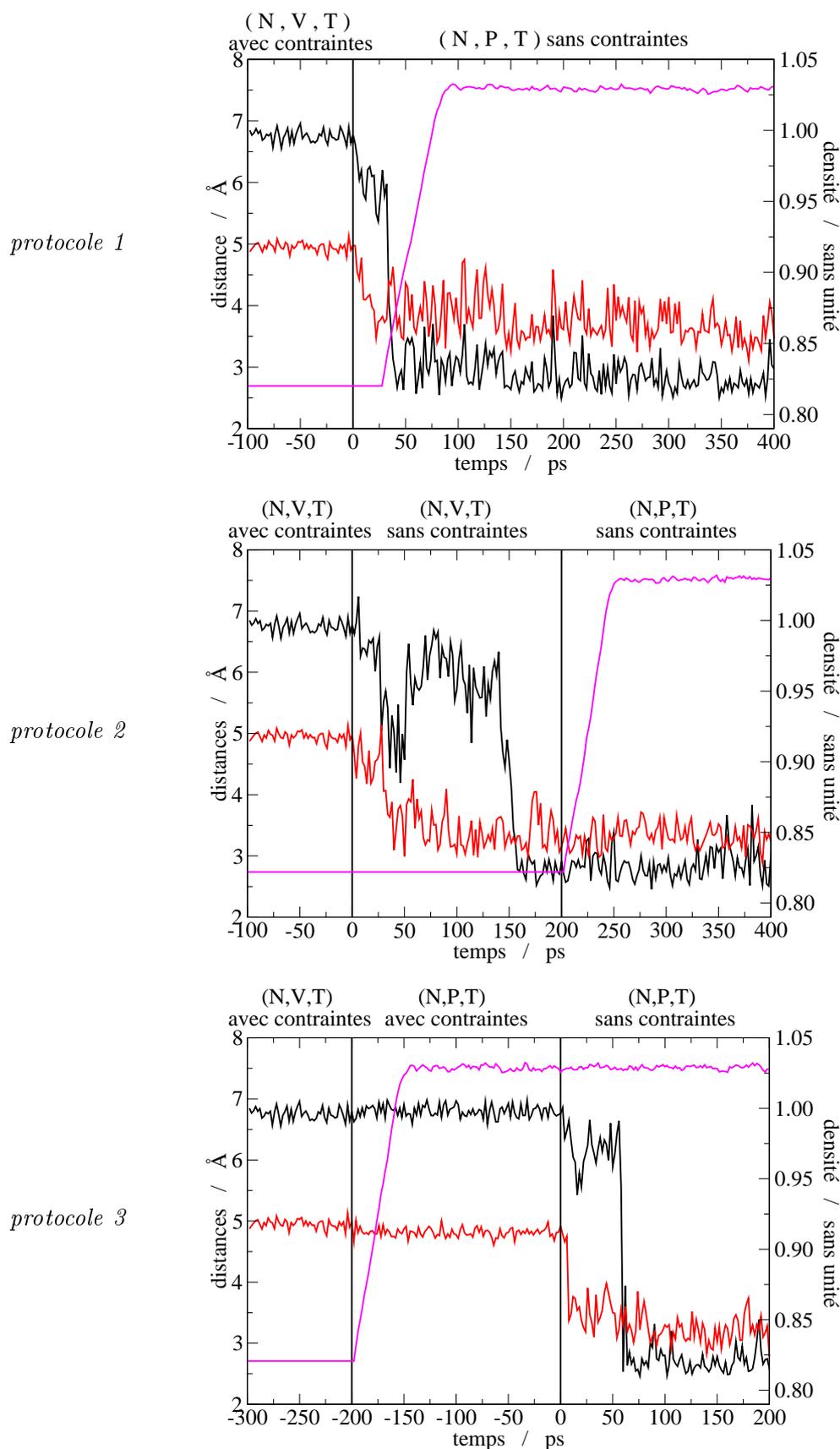


Figure II.11 – Trajectoires de la distance entre l'atome ODH de Asp148 et l'atome d'oxygène du carbonyle de Lys166 (en noir), de la moyenne des deux longueurs des liaisons hydrogènes entre Tyr151 et Asn164 (en rouge) et de la densité de la boîte de simulation (en violet). Le temps $t=0$ correspond au relâchement des contraintes et les barres verticales séparent les différents type de simulation.

la GFP et de la ECFP (conformation A et B) au cours de simulations réalisées précédemment au laboratoire avec le même champ de force [69, 78]. Les distances moyennes entre les paires d'acides aminés listés dans le tableau II.4 obtenues à partir des simulations de ces protéines sont proches des distances mesurées dans la structure RX, la plus grande variation est de 0.7 Å. Ces simulations n'indiquent donc aucune déformation d'une ampleur de celle observée dans la Cerulean. Le champ de force utilisé dans ces simulations reproduit donc bien l'équilibre entre les interactions attractives et répulsives présentes entre les deux brins qui forment le *cleft* dans les structures RX de la GFP et de la ECFP. Il semble peu vraisemblable que le champ de force soit mis en défaut spécifiquement dans le cas de la Cerulean alors qu'elle présente une structure tridimensionnelle similaire à celles de la GFP et de la ECFP. Il semble donc peu probable que la différence entre la structure moyenne de la Cerulean obtenue à partir des simulations et la structure RX provienne d'un défaut du champ de force ou de la dynamique moléculaire.

En ce qui concerne l'hypothèse de l'effet d'empilement cristallin, nous avons analysé les interactions entre une protéine donnée du cristal (nommée par la suite protéine centrale) et ses plus proches voisins dans la phase cristalline de la Cerulean, puis, pour comparaison, dans celle de la GFP. Dans le cas de la GFP, la structure RX correspondant à l'entrée 1EMA [18] de la *Protein Data Bank* a été utilisée. Ce travail a nécessité de construire les répliques de la protéine centrale à partir de la symétrie cristalline.

Les caractéristiques du cristal à partir desquelles la structure RX a été obtenue sont indiquées dans le fichier de la *Protein Data Bank*. La plupart des GFPs cristallisent dans un réseau orthorhombique caractérisé par une maille parallélépipédique (les trois angles α , β et γ du cristal sont droits et les trois arêtes ont des longueurs différentes). Chaque arête de la maille conventionnelle est colinéaire à un des axes du cristal représentés par les vecteurs \vec{a} , \vec{b} et \vec{c} de normes a , b et c (ici $a \neq b \neq c$). Le groupe de symétrie du cristal est $P2_12_12_1$ ^a. Il est assez complexe de construire une maille et un certain nombre de ses répliques à partir du groupe de symétrie auquel elle appartient. Il est plus commode d'utiliser les positions de Wyckoff qui sont données pour chaque groupe de symétrie dans les tables internationales et sont également précisées dans le fichier contenant la structure RX. A chaque position de Wyckoff correspond une opération de symétrie (une rotation plus une translation) permettant de construire une entité de la maille à partir des coordonnées d'une entité de base. Pour le groupe de symétrie $P2_12_12_1$, il y a 4 positions de Wyckoff. Les cristaux des GFP présentent donc 4 entités par maille (ici une entité est une protéine). Voici les positions de Wyckoff

^aLe nom d'un groupe de symétrie est constitué d'une lettre définissant le type de maille du cristal suivie des opérations de symétrie du cristal. La lettre P signifie primitif. Le symbole 2_1 représente une opération de symétrie. Il s'agit d'un axe de symétrie hélicoïdal correspondant à une rotation d'un angle π suivie d'une translation d'une unité selon un des vecteurs du réseau qui porte l'axe de symétrie. Dans le cas d'un groupe de symétrie d'un réseau orthorhombique, la première opération de symétrie est associée aux axes \vec{a} , \vec{b} et \vec{c} du réseau, la seconde est associée à l'axe défini par le vecteur $\vec{a} + \vec{b} + \vec{c}$ et la troisième opération de symétrie est associée aux axes $\vec{a} + \vec{b}$, $\vec{a} + \vec{c}$ et $\vec{b} + \vec{c}$.

du groupe $P2_12_12_1$:

$$\begin{aligned} \mathcal{O}_1 : & \quad x \quad , \quad y \quad , \quad z \\ \mathcal{O}_2 : & \quad 1/2 - x \quad , \quad -y \quad , \quad 1/2 + z \\ \mathcal{O}_3 : & \quad -x \quad , \quad 1/2 + y \quad , \quad 1/2 - z \\ \mathcal{O}_4 : & \quad 1/2 + x \quad , \quad 1/2 - y \quad , \quad -z \end{aligned}$$

L'opération 1 est l'identité et correspond en fait à la structure donnée dans le fichier de la *Protein Data Bank*. L'opération 2 est une rotation suivie d'une translation qui peuvent être représentées par des opérateurs matriciels de la façon suivante :

$$\mathcal{O}_2 \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} a/2 \\ 0 \\ c/2 \end{pmatrix}$$

Il s'agit en fait d'une rotation d'un angle π autour de l'axe \vec{c} suivie d'une translation suivant l'axe $\vec{a} + \vec{c}$. Les répliques de la protéine centrale sont obtenues en appliquant aux coordonnées de la structure RX les quatre opérations de symétries données par les positions de Wyckoff. On obtient ainsi les 4 entités par maille.

A chacune des répliques ainsi obtenues on applique en plus un ensemble de translations d'un vecteur \vec{R} du réseau défini par $\vec{R} = h\vec{a} + k\vec{b} + l\vec{c}$, où h , k et l sont des entiers relatifs. En ne conservant que les protéines les plus proches de la protéine centrale, 12 structures sont retenues, ce qui correspond à la coordinance d'une entité dans une structure compacte contenant 4 entités par maille.

A partir des structures de la protéine centrale et de ses 12 voisines nous avons cherché des paires d'atomes, l'un appartenant à la protéine centrale et l'autre à une protéine voisine, dont la distance est inférieure à 3.5 Å ou 5 Å suivant que la chaîne latérale de l'acide aminé considéré est présente ou absente dans la structure RX. La distance de 3.5 Å a été choisie car elle correspond à la longueur d'une liaison hydrogène. Dans la Cerulean, les chaînes latérales de 5 acides aminés ne sont pas résolues : Glu132, Asp133, Lys156, Gln157 et Lys158. Cependant, la position de l'atome $C\beta$ qui suit l'atome $C\alpha$ dans la chaîne latérale est résolue, l'orientation de la chaîne latérale est donc connue et nous avons choisi d'inclure ces acides aminés dans la recherche des contacts entre les protéines en utilisant un rayon de coupure de 5 Å. Pour la protéine centrale, nous avons limité la recherche aux atomes des acides aminés de la position 140 à 176. Cette séquence contient les brins 6 et 7, une longue boucle souple qui prolonge le brin 6 et qui traverse la protéine, ainsi que les coudes entre les brins 6 et 7 et entre les brins 7 et 8. Plusieurs contacts sont observés et la liste des paires d'atomes étant en contact est donnée dans le tableau II.5.

Les interactions les plus fortes sont celles entre les paires d'acides aminés chargés : Arg168-Glu132 et Asp173-Lys126. Etant donné la position de Arg168 sur le brin 7, une forte attraction entre Arg168 et Glu132 peut probablement avoir pour conséquence de perturber le réseau de liaisons hydrogène

protéine centrale	protéine voisine	distances	opération de symétrie
OE1 (Glu142)	OD2 (Asp197)	3.2	$(-x - a, -y, z - c)$
CZ (Arg168)	CB (Glu132) ^a	4.5	$(x - a, -y - b, -z)$
OD2 (Asp173)	NZ (Lys126)	2.7	
OG (Ser175)	N(Gly24)	3.3	

^aLa chaîne latérale de Glu132 n'est que partiellement résolue dans la structure RX

Tableau II.5 – Contacts entre protéines dans le cristal de la Cerulean (structure RX code 2Q57) entre des atomes de la protéine centrale (colonne 1) et des atomes d'une protéine voisine dans le cristal (colonne 2) obtenue par l'opération de symétrie donnée en colonne 4. La distance entre les atomes (colonne 3) est donnée en Å. Les noms des atomes sont ceux indiqués dans la structure RX.

entre les brins 6 et 7 et accentuer l'ouverture entre ces brins (figure II.12). La probabilité que les autres contacts aient un effet sur l'ouverture entre les brins 6 et 7 est plus faible. Une interaction forte existe entre les acides aminés chargés Asp173 et Lys126, cependant Asp173 est positionné sur le coude entre les brins 7 et 8 et cette interaction n'aura qu'un effet indirect sur le *cleft*. Les deux autres paires d'acides aminés, Ser175-Gly24 et Glu142-Asp197 impliquent des interactions plus faibles, par l'intermédiaire de liaisons hydrogène (d'après les calculs réalisés avec le serveur H++, Glu142 est sous sa forme acide, neutre à pH 5). De plus, Ser175 et Glu142 ne sont pas positionnés directement sur les brins 6 et 7 et ils n'auront donc qu'un effet indirect sur l'ouverture entre ces brins (figure II.12).

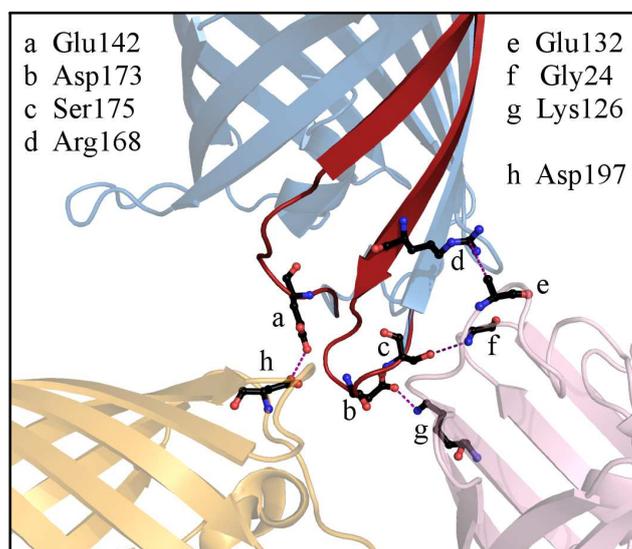


Figure II.12 – Vue des interactions entre des atomes d'une protéine centrale (en bleu) et deux protéines voisines dans la région (en rouge) des brins 6 et 7 de la protéine centrale. La protéine jaune est obtenue par l'opération de symétrie $(-x - a, -y, z - c)$ et la protéine rose est obtenue par l'opération de symétrie $(x - a, -y - b, -z)$. Les acides aminés représentés sont ceux dont au moins un atome est interaction avec un atome de la région rouge de la protéine centrale (tableau II.5).

Revenons maintenant à l'interaction entre Arg168 et Glu132. La chaîne latérale de Glu132 n'est pas complètement résolue dans la structure RX. La difficulté pour résoudre la structure peut être liée au fait que cette partie de la protéine est flexible. Dans le cristal, la chaîne latérale de Glu132 adopterait plusieurs conformations équiprobables ce qui rend les maxima de la densité électronique difficiles à identifier. En raison de ces fluctuations, l'interaction entre Arg168 et Glu132 ne serait donc pas permanente. Cependant, l'extrémité de Arg168 (le groupe NH_2^+) et l'atome $C\beta$ de Glu132 (atome CB dans le tableau II.5) sont orientés de telle sorte qu'une forte interaction entre les chaînes latérales de ces deux acides aminés est possible pour diverses conformations de Glu132. La structure RX obtenue est donc une moyenne entre des structures dans lesquelles l'interaction entre Glu132 et Arg168 existe et d'autres où elle n'existe pas. Dans les configurations où elle existe effectivement, la proximité des chaînes latérales de ces acides aminés conduirait à une interaction suffisante pour rompre partiellement le réseau de liaisons hydrogène entre les brins 6 et 7.

protéine centrale	Molécule d'eau intermédiaire	distance 1 ^a	protéine voisine	distance 2 ^b
OD2 (Asp173)			NZ (Lys126)	3.0
OG (Ser175)			OD1 (Asn23)	3.1
O (Gly174)	Wat379	3.2	N (Asp129)	2.9
OG (Ser175)	Wat335	3.1	O (Gly127)	2.8
			O (Asp21)	3.1
			OD1 (Asp21)	2.5

^aDistance entre un atome de la protéine centrale et une molécule d'eau intermédiaire.

^bDistance entre un atome d'une protéine voisine et une molécule d'eau intermédiaire ou la protéine centrale.

Tableau II.6 – Plus courtes distances dans le cristal de la GFP, en Å, entre des atomes de la protéine centrale (colonne 1) et des atomes d'une protéine voisine dans le cristal (colonne 4) par l'intermédiaire de molécules d'eau cristallographiques (colonne 2). La protéine voisine est obtenue par l'opération de symétrie $(x, -y + b, -z + c)$. Les noms des atomes sont ceux indiqués dans la structure RX de la GFP, entrée 1EMA de la *Protein Data Bank*.

Pour comparaison, nous avons effectué une analyse identique sur la structure de la GFP (entrée 1EMA de la *Protein Data Bank*) qui cristallise dans le même réseau que la Cerulean. Les résultats obtenus sont regroupés dans le tableau II.6. Plusieurs différences sont à noter entre les résultats obtenus pour la Cerulean et ceux obtenus pour la GFP. Tout d'abord, une seule protéine voisine, obtenue par l'opération de symétrie $(x, -y + b, -z + c)$, présente des contacts avec la protéine centrale dans la région autour des brins 6 et 7 (acides aminés de la position 140 à 176, zone rouge de la figure II.13). D'autre part, seulement deux interactions directes entre les atomes de la protéine centrale et les atomes d'une protéine voisine existent (Asp173-Lys126 et Ser175-Asn23). Les autres interactions impliquent une molécule d'eau intermédiaire. Enfin, aucune des interactions citées dans le tableau II.6 ne fait intervenir des acides aminés appartenant aux brins 6 et 7. Les acides aminés de la protéine centrale mis en jeu (Asp173, Gly174 et Ser175) se situent sur le coude entre les brins 7 et 8. Il est donc peu probable que ces interactions, plus faibles que dans la Cerulean et non

positionnées directement sur les brins 6 et 7 aient une influence sur l'ouverture entre ces brins. Ceci est cohérent avec l'ouverture moindre observée expérimentalement dans la structure RX de la GFP.

Il semble donc que, dans le cas de la Cerulean, l'empilement cristallin conduise à des interactions dans la région du *cleft* qui contribuent à l'ouvrir le *cleft* en phase solide. Ces interactions disparaissent en phase aqueuse et le *cleft* se referme en partie. Dans le cas de la GFP de telles interactions ne sont pas présentes dans le cristal et aucune modification du *cleft* n'est observée lorsque l'on passe en phase aqueuse. Nous en déduisons qu'il est très probable que la différence entre la structure RX de la Cerulean et la structure moyenne que nous obtenons par simulation, est due à un effet d'empilement dans le cristal.

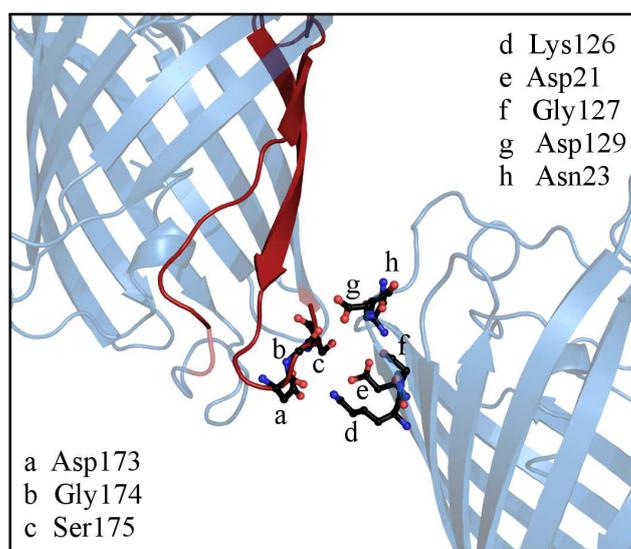


Figure II.13 – Vue des interactions inter-protéines dans le cristal de la GFP(structure RX code 1EMA). La protéine centrale est à gauche, la protéine de droite est obtenue par l'opération de symétrie (x , $-y + b$, $-z + c$). La région représentée en rouge représente les acides aminés de la protéine centrale inclus dans la recherche des contacts avec les protéines voisines. Les acides aminés représentés sont ceux dont au moins un atome est interaction avec un atome de la région rouge de la protéine centrale (tableau II.5), soit directement, soit par l'intermédiaire d'une molécule d'eau.

II.2.5 Structure et dynamique à pH 8

Nous avons fait deux simulations de la Cerulean à pH 8, en utilisant la distribution de protons décrite au paragraphe II.2.1.3. Dans l'une des simulations le chromophore est en conformation *cis* (pH8-*cis*) et dans l'autre il est en conformation *trans* (pH8-*trans*). Les deux conformations du chromophore sont montrées dans la figure I.6 page 20. Dans ces simulations, Asp148 est sous sa forme basique, il porte donc une charge négative. Nous analysons ici la position de la chaîne latérale de Asp148 et le réseau de liaisons hydrogène au voisinage de cet acide aminé et du chromophore pour chacune de ces conformations.

II.2.5.1 Simulation à pH 8 avec le chromophore en conformation *cis*

Pour cette simulation, nous avons pris comme coordonnées initiales celles de la structure RX à pH 5. La stabilité de la structure du squelette protéique a été vérifiée. Les valeurs des RMSd des atomes sont similaires à celles obtenues pour les simulations à pH 5. Le résultat principal de la simulation à pH 8 est que Asp148, qui est maintenant chargé négativement, oriente sa chaîne latérale vers l'extérieur de la protéine, dans le solvant, et ne forme plus de liaison hydrogène avec l'indole du chromophore. La dynamique globale de la protéine est similaire à celle décrite au paragraphe II.2.4, notamment en ce qui concerne le rapprochement des brins 6 et 7.

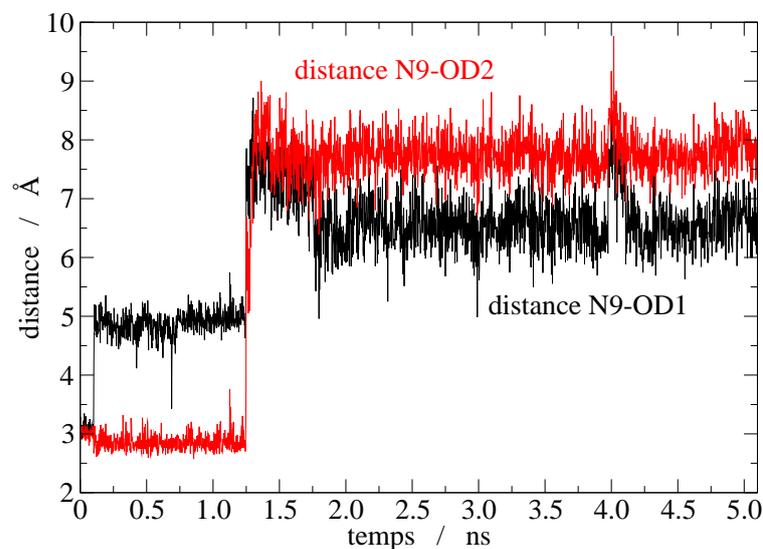


Figure II.14 – Trajectoires de la distance entre les atomes d'oxygène du groupe COO^- de Asp148 et l'atome N9 du chromophore de la Cerulean sur l'ensemble de la simulation à pH 8 avec le chromophore en conformation *cis*. La simulation de la protéine sans contrainte commence à $t=0.1$ ns.

La figure II.14 présente les trajectoires des distances entre les deux atomes d'oxygène (OD1 et OD2 du groupement COO^-) de Asp148 et l'atome d'azote N9 de l'indole du chromophore sur l'ensemble de la simulation. Les 100 premières picosecondes correspondent à la mise à l'équilibre de la température. Le mouvement de la chaîne latérale de Asp148 se déroule en deux étapes. Tout d'abord, lorsque les contraintes sur la protéine sont retirées (à $t=0.1$ ns) la distance entre l'atome

N9 et l'atome OD1 augmente et se stabilise autour de 5 Å, alors que la distance entre l'atome N9 et l'atome OD2 reste stable aux alentours de 2.9 Å. Comme dans les simulations à pH 5, on atteint cette géométrie par une combinaison des torsions α et θ (figure II.6 page 63). Une liaison hydrogène existe entre l'atome OD2 et l'atome N9 entre $t=0.1$ ns et $t=1.2$ ns. Les distances sont similaires à celles obtenues dans les simulations 1 à 3, à pH 5 pour les distances entre l'atome N9 et les atomes OD et ODH ^a, (tableau II.3 page 65). La position de la chaîne latérale de Asp148 est alors similaire à la conformation d'équilibre adoptée dans les simulations 1 à 3. Après une nanoseconde, le groupe COO⁻ de Asp148 se déplace à l'extérieur de la protéine et stabilise sa charge négative par de multiples liaisons hydrogène avec des molécules de solvant. Les deux distances entre N9 et les atomes OD1 et OD2 augmentent brusquement et se stabilisent respectivement autour de 6.5 Å et 7.1 Å. Cette situation est stable durant les 3 nanosecondes suivantes.

	β	α	θ
référence [37]	/	166	-53
pH5- <i>cis</i>	0 (11)	3 (22)	-54 (11)
pH8- <i>cis</i>	/	75 (8)	-144 (22)
pH8- <i>trans</i>	/	101 (6)	-66 (9)

Tableau II.7 – Valeur moyenne et déviation standard (entre parenthèse) des torsions (en degrés) qui permettent de caractériser les conformations de la chaîne latérale de Asp148 obtenues pour les simulation pH5-*cis* (simulation 1), pH8-*cis* (sur les trois dernières nanosecondes) et pH8-*trans*. Les noms des atomes et les torsions sont définis sur la figure II.6 page 63. Les données de la ligne *référence* sont extraites de la structure RX de la Cerulean.

Les valeurs des coordonnées internes permettant de caractériser les conformations de Asp148 sont rassemblées dans le tableau II.7. Les valeurs de θ sont positives et sont en accord avec la position du groupe COO⁻ à l'extérieur de la protéine.

La figure II.15 présente le réseau de liaisons hydrogène analysé sur les 3 dernières nanosecondes de la simulation avec les critères donnés au paragraphe B.2.5 page 173. La fonction carboxylate COO⁻ de Asp148 forme des liaisons hydrogène avec le groupement NH du squelette protéique des acides aminés Asn149 et Lys166 et avec des molécules du solvant. L'indole du chromophore ne trouve alors aucun partenaire pour former une liaison hydrogène. En ce qui concerne le cycle imidazolinone, le réseau de liaisons hydrogène est semblable à celui déjà observé à pH 5 (figure II.9 page 68).

Le nombre de liaisons hydrogène que forme le groupe COO⁻ de Asp148 (figure II.15) est supérieur au nombre maximal attendu. En effet, dans le cas d'un groupe COO⁻, le nombre total de doublets non liant est de 5 et on s'attend donc à ce que ce groupe forme au maximum 5 liaisons hydrogène. Or la somme des pourcentages de liaisons hydrogène obtenue au cours de la simulation, et qui impliquent Asp148, est de 559%. Le groupe COO⁻ formerait donc en permanence plus de 5 liaisons hydrogène. C'est en fait un *artefact* qui provient du critère géométrique utilisé pour tester l'existence

^aEn considérant que l'atome d'oxygène qui porte le proton dans la structure RX est celui proche du carbonyle du squelette protéique de Asp148, l'atome OD1 peut être identifié à OD et l'atome OD2 à ODH.

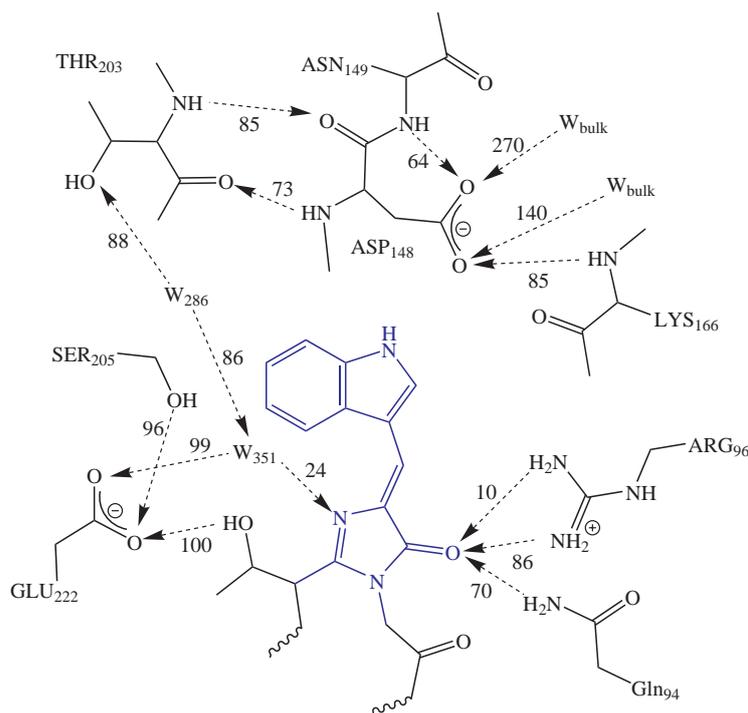


Figure II.15 – Réseau de liaisons hydrogène dans l’environnement du chromophore obtenu à partir de la simulation à pH 8 avec le chromophore en conformation *cis*. Les flèches indiquent la présence d’une liaison hydrogène et sont orientées du donneur vers l’accepteur d’hydrogène. Les nombres associés 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.

d’une liaison hydrogène. Cependant, bien que surestimés, ces pourcentages d’existence des liaisons hydrogène indiquent que tout au long de la simulation le groupe COO^- forment plusieurs liaisons hydrogène avec des molécules d’eau du solvant.

L’orientation du groupement COO^- entre les brins 6 et 7 et vers le solvant est donc stabilisée par un réseau de liaison hydrogène plus dense que lorsqu’il est vers l’intérieur et interagit avec la fonction NH de l’indole. Afin de préciser cette préférence conformationnelle à pH 8 nous avons calculé la variation d’énergie libre partielle le long de la torsion θ (défini sur la figure II.6 page 63) par *Umbrella Sampling*, (paragraphe I.4 page 35).

Le profil d’énergie libre obtenu est présenté sur la figure II.16. Les configurations pour lesquelles les valeurs de l’angle θ sont négatives sont caractéristiques d’une position de la chaîne latérale de Asp148 à l’intérieur de la protéine et les valeurs de l’angle θ positives sont caractéristiques d’une position de la chaîne latérale à l’extérieur de la protéine.

D’après le profil d’énergie libre, il existe deux conformations d’équilibre de la chaîne latérale, l’une à l’intérieur de la protéine pour θ voisin de -50° , l’autre à l’extérieur pour θ voisin de 70° . La conformation plaçant la chaîne latérale de Asp148 à l’extérieur est plus stable d’environ $8 \text{ kcal} \cdot \text{mol}^{-1}$, probablement en raison d’un facteur entropique plus favorable et de la formation de nombreuses liaisons hydrogène avec le solvant. Pour atteindre cette conformation à partir de la géométrie ini-

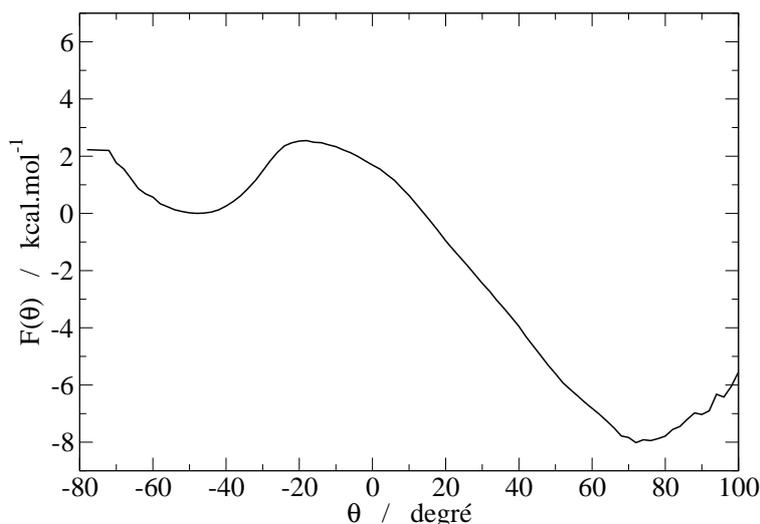


Figure II.16 – Profil d'énergie libre en fonction de la torsion θ autour de la liaison CA-CB de Asp148 à pH 8 (angle dièdre C-CA-CB-CG).

tiale, une faible barrière d'environ $2 \text{ kcal} \cdot \text{mol}^{-1}$ doit être franchie. Elle est probablement due au fait qu'il est nécessaire de rompre la liaison hydrogène entre Asp148 et le groupe NH du chromophore. L'existence de cette barrière est cohérente avec le fait que le changement de conformation s'opère à l'échelle de temps d'une nanoseconde.

Ces résultats complètent ceux donnés par le réseau de liaisons hydrogène obtenu à pH 8 pour confirmer que lorsque le pH augmente de pH 5 à pH 8 la chaîne latérale de Asp148 sort de la protéine et s'oriente vers le solvant. Le chromophore en conformation *cis* est alors isolé au sein de la cavité protéique, et sa partie indole n'est liée par aucune liaison hydrogène avec des acides aminés proches.

II.2.5.2 Simulation à pH 8 avec le chromophore en conformation *trans*

Pour cette simulation, les coordonnées initiales sont obtenues à partir d'une configuration de la simulation à pH8-*cis*, dans laquelle nous avons fait tourner le chromophore pour le placer en conformation *trans* (paragraphe II.2.2 page 59). La structure de la protéine est stable tout le long de la simulation. La figure II.17 présente l'évolution dans le temps du RMSd.

Les valeurs du RMSd de l'ensemble des $C\alpha$ des acides aminés formant un feuillet β ou une hélice α fluctuent entre 0.8 et 1.0 Å, et les valeurs du RMSd de l'ensemble des $C\alpha$ fluctuent entre 1.2 et 1.4 Å. Ces ordres de grandeurs sont similaires à ceux observés figure II.7 pour la simulation à pH 5.

Comme dans la simulation pH8-*cis*, la valeur moyenne de θ au cours de la simulation pH8-*trans* est positive et caractéristique de la position de la chaîne latérale de Asp148 à l'extérieur de la protéine (tableau II.7). Le réseau de liaisons hydrogène obtenu à partir de cette simulation est présenté sur la figure II.18.

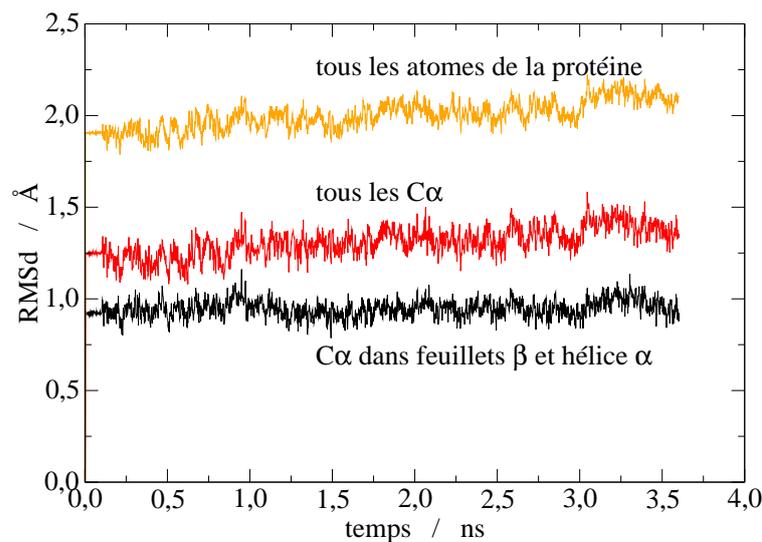


Figure II.17 – Déplacement carré moyen, en fonction du temps, des C_{α} des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β (en noir), de l'ensemble des C_{α} (en rouge) et de l'ensemble des atomes de la protéine (en orange) dans la simulation pH8-*trans*.

Le groupement COO^- de Asp148 forme des liaisons hydrogènes avec Asn149 et des molécules du solvant (comme dans la simulation pH8-*cis*) et avec Arg168 (au lieu de Lys166 dans la simulation pH8-*cis*). Arg168 est situé sur le brin 7 et sa chaîne latérale se replie le long des brins 6 et 7. Cette position de la chaîne latérale de Arg168 avait déjà été observée dans le cas de la GFP [24]. Le réseau de liaisons hydrogène concernant l'indole du chromophore correspond à une situation intermédiaire entre les deux réseaux de liaisons hydrogène obtenus pour les conformations A et B de la ECFP [69]. Le cycle indole du chromophore est en liaison hydrogène soit avec le groupe COO^- de Glu222 (comme dans la ECFP conformation B), soit avec Ser205 (comme dans la ECFP conformation A). Le reste du réseau de liaisons hydrogène est similaire à ceux obtenus pour les autres simulations (figure II.9 page 68 et II.15 page 81) et plus généralement pour les protéines de la famille de la GFP.

Contrairement à la situation observée dans la simulation à pH 8 avec le chromophore en conformation *cis*, l'indole du chromophore forme des liaisons hydrogène lorsque le chromophore est en conformation *trans*. Du point de vue du chromophore, à pH 8, lorsque Asp148 est sous sa forme basique et à l'extérieur de la protéine, la conformation *trans* semble donc énergétiquement plus favorable.

La structure de la Cerulean est inconnue expérimentalement à pH 8. Cependant, la structure RX des deux conformations de la ECFP a été obtenue à pH 8 et présente le chromophore en conformation *trans* [67]. Nous avons comparé la structure de la Cerulean issue de notre simulation pH8-*trans* avec celle de la ECFP qui existe sous deux conformations : A et B (figure II.1 page 51). Dans la ECFP l'acide aminé en position 148 est une histidine qui change d'orientation suivant la conformation A ou B de la protéine. La superposition des structures montre que la structure issue de nos simulations à pH 8 est plus proche de la structure de la conformation B que de celle de la conformation A (figure

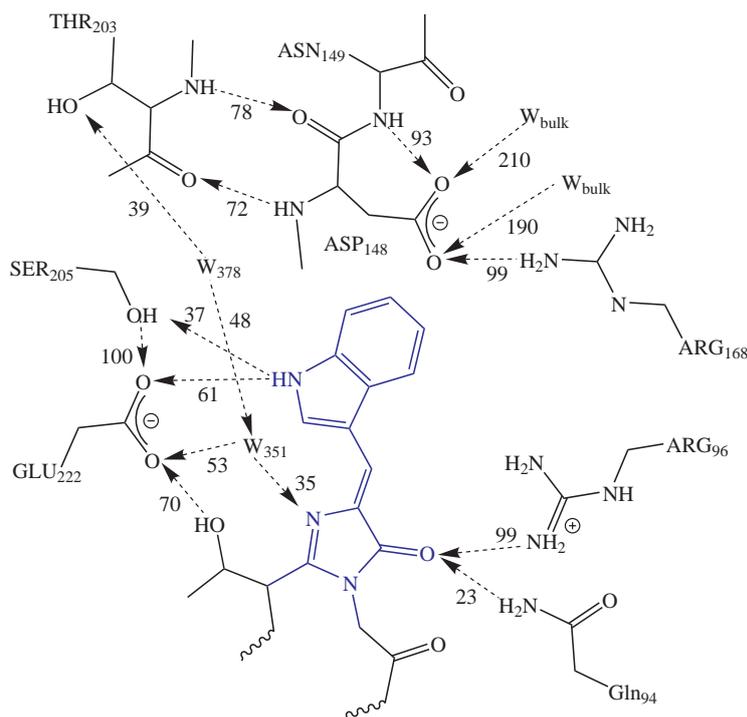


Figure II.18 – Réseau de liaisons hydrogène dans l’environnement du chromophore obtenu à partir de la simulation à pH 8 avec le chromophore en conformation *trans*. Les flèches indiquent la présence d’une liaison hydrogène et sont orientées du donneur vers l’accepteur d’hydrogène. Les nombres associés 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.

II.19). Notamment, le brin qui porte l’acide aminé 148 a une structure proche dans nos simulations et dans la conformation B de la ECFP. De plus l’orientation de la chaîne latérale de Asp148 dans nos simulations est similaire à celle de His148 dans la conformation B de la ECFP. Une telle position avait été supposée par Malo *et al* [37].

Les résultats des simulations à pH 8, le calcul des profils d’énergie libre partielle en fonction de la torsion θ , ainsi que la comparaison de la structure de la Cerulean dans ces simulations avec celle de la ECFP, semblent indiquer que la structure de la Cerulean à pH 8 est proche de celle de la conformation B de la ECFP, avec la chaîne latérale de Asp148 orientée vers l’extérieur de la protéine et le chromophore en conformation *trans*. Cependant, pour confirmer ces hypothèses, il est nécessaire de vérifier si cette structure conduit au décalage du spectre d’absorption de 11-15 nm vers le rouge qui a été observé expérimentalement quand on passe de pH5 à H8.

II.3 Spectre d’absorption en fonction du pH

Le décalage du spectre d’absorption sous l’effet d’un changement de pH peut avoir deux origines [37] : la modification de l’environnement du chromophore dû à la variation du pH (changement de l’état

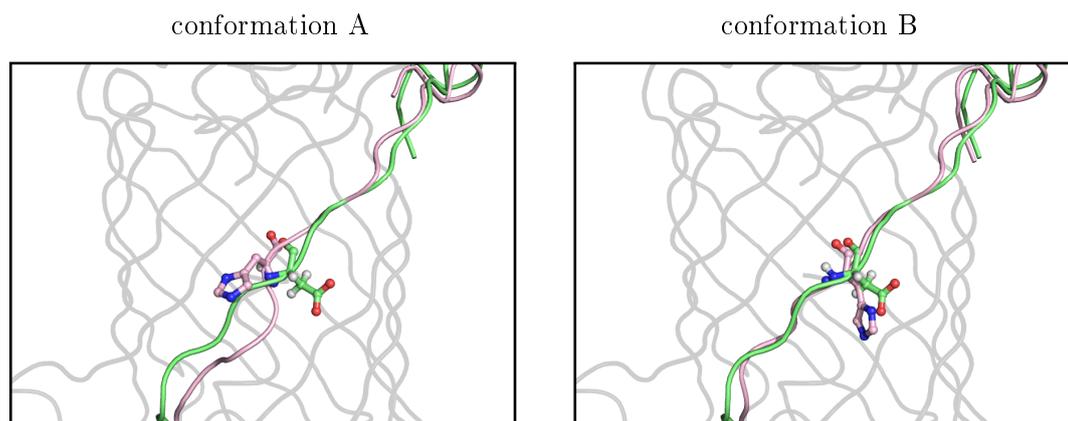


Figure II.19 – Superposition des structures de la Cerulean au cours de la simulation pH8-*trans* et de la structure RX de la ECFP dans sa conformation A, ou dans sa conformation B. Le brin 6 qui porte His148 dans la ECFP et Asp148 dans la Cerulean est mis en évidence. Il est représenté en vert dans la Cerulean et en rose dans la ECFP. Les acides aminés Asp148 et His148 sont représentés respectivement en vert et en rose.

de protonation et éventuellement de la conformation des acides aminés voisins) ou/et l'isomérisation du chromophore.

Le but de cette partie est d'évaluer la contribution de ces deux effets séparément, d'évaluer les incertitudes provenant de la méthode de calcul et d'analyser la nature des transitions électroniques dans les deux conformations du chromophore.

Cependant, l'effet électrostatique lié à la modification de la distribution des protons dans la protéine sous l'effet du changement de pH est pris en compte approximativement dans nos simulations. En effet, pour modéliser le pH dans nos simulations nous avons choisi un micro-état de protonation en supposant qu'il est représentatif du pH considéré. Ceci est une approximation importante compte tenu du fait qu'un nombre important de micro-états de protonation peuvent être quasi équiprobables à un pH donné. Une méthode sera proposée pour compter le nombre de micro-états nécessaires pour représenter un pH et pour calculer l'effet de ces micro-états de protonation sur le spectre d'absorption de la Cerulean.

II.3.1 Calcul d'un spectre d'absorption par simulation moléculaire

Les simulations de dynamique moléculaire utilisent la mécanique classique. Elles ne donnent donc pas d'information sur les énergies de transition électronique du chromophore. Ce paragraphe présente la méthode utilisée pour calculer le spectre d'absorption [69, 79, 80].

II.3.1.1 Choix d'un système quantique

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, le choix de ce modèle de chromophore n'est pas trivial car le chromophore de la Cerulean est relié de manière covalente à la protéine. Il faut donc définir les liaisons qui déterminent les bords du chromophore. Lors d'une étude antérieure, plusieurs modèles de chromophore ont été testés afin de décrire au mieux les énergies de transition électronique du chromophore, en limitant le nombre d'atomes pour conserver des temps de calcul acceptables [69]. Ces modèles contiennent tous les cycles indole et imidazolinone du chromophore et diffèrent par la liaison, dans les parties issues de Thr65 et Gly67, à partir de laquelle s'arrête le chromophore. Il résulte de cette étude que le modèle présenté à la figure I.6 page 20, qui est le même que celui utilisé pour construire le champ de force, présente le meilleur compromis entre la taille du système et la précision des énergies de transition obtenues. Ce modèle pourra donc être utilisé pour calculer les transitions électroniques du chromophore.

II.3.1.2 Calcul d'un spectre d'absorption dans la protéine

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 (équation I.1 page 11). 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 probabilité 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 λ_c est alors donnée par la probabilité d'obtenir une configuration dans laquelle la longueur d'onde d'absorption est égale à λ_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 procéderons 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, par la méthode TDDFT 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.

$$S(\lambda) = \sum_{c=1}^{N_{conf}} \sum_{i=1}^{N_{etat}} f_{c,i} \delta(\lambda - \lambda_c) \quad (\text{II.2})$$

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

$$S(\lambda) = \sum_{c=1}^{N_{conf}} \sum_{i=1}^{N_{etat}} f_{c,i} \exp\left(-\frac{(\lambda - \lambda_{c,i})^2}{2\sigma^2}\right) \quad (\text{II.3})$$

où σ 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 discuté en détail au paragraphe II.3.3.1. Le nombre N_{etat} d'états électroniques excités à prendre en compte correspond aux nombres 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 dans 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.

II.3.2 Spectres d'absorption de la Cerulean

Nous avons calculé le spectre d'absorption de la Cerulean dans les trois situations étudiées jusqu'à présent, pH5-*cis*, pH8-*cis* et pH8-*trans*. Les spectres associés aux trois situations ont été calculés avec la méthode décrite au paragraphe II.3.1 en utilisant environ 800 configurations dans chaque cas. Les trois spectres ainsi obtenus sont présentés figure II.20.

Le principal résultat concerne la variation de la position λ_{max} du maximum du spectre. Quand le pH est abaissé de pH 8 à pH 5 et que seule la distribution des protons dans la protéine est modifiée, le chromophore étant en conformation *cis* aux deux pH, la valeur de λ_{max} se décale de 4 nm vers le rouge. Par contre, lorsque le pH est abaissé de pH 8 à pH 5 et que le chromophore passe de la conformation *trans* à la conformation *cis*, la valeur de λ_{max} se décale de 15 nm vers le bleu ce qui est en bon accord avec le décalage expérimental (décalage de 11-15 nm vers le bleu, figure II.20).

On peut conclure de ces résultats que le changement de l'environnement électrostatique du chromophore ne permet pas à lui seul d'expliquer le décalage de 11-15 nm vers le bleu observé expéri-

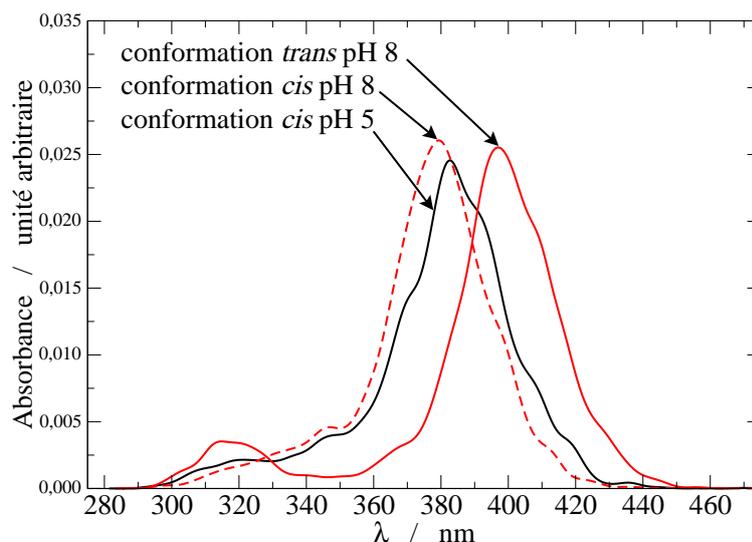


Figure II.20 – Spectres d'absorption de la Cerulean, à pH 5 avec le chromophore en conformation *cis*, pH5-*cis* (noir), à pH 8 avec le chromophore en conformation *cis*, pH8-*cis* (pointillés rouges) et à pH 8 avec le chromophore en conformation *trans*, pH8-*trans* (trait plein rouge).

mentalement (figure II.2 page 53) et que le décalage expérimental est très probablement un signe de l'isomérisation du chromophore en réponse au changement du pH.

Cependant plusieurs points sont à discuter. Tout d'abord les spectres d'absorption expérimentaux de la Cerulean (tout comme ceux de la ECFP) présentent un double maximum [35, 36, 37] qui n'est pas reproduit par nos calculs. Cet double maximum est généralement attribué à une structure vibrationnelle du spectre d'absorption, que nos calculs ne peuvent pas reproduire.

D'autre part, il y a un décalage important entre les spectres calculés et observés : la valeur de λ_{max} sur le spectre obtenu par nos calculs à pH 5 avec le chromophore en conformation *cis* vaut 382 nm alors qu'elle vaut 424 nm expérimentalement (figure II.2 page 53). Un décalage similaire a déjà été constaté plusieurs fois dans d'autres études employant la méthode TDDFT pour calculer des spectres d'absorption [80, 69] et son ordre de grandeur est de nature à mettre en doute la conclusion annoncée sur l'origine du décalage lié à l'abaissement du pH. C'est pourquoi nous avons entrepris une étude plus systématique de l'incertitude provenant de la méthode de calcul utilisée.

II.3.3 Facteurs influençant la forme et la position du spectre d'absorption théorique

Nous étudions ici comment le choix de la fonctionnelle de la densité et du nombre de configurations utilisées pour calculer le spectre d'absorption influe sur sa position et sur son décalage en fonction du pH.

II.3.3.1 Effet de la taille de l'échantillon

Afin de réduire l'incertitude sur le spectre d'absorption calculé, liée au nombre de configurations utilisées pour construire ce spectre, on peut simplement envisager d'augmenter la taille de l'échantillon de configurations utilisées. Cependant la taille de l'échantillon est forcément limitée à un certain niveau par le temps de calcul nécessaire et en l'absence d'une évaluation spécifique de l'incertitude, même le calcul le plus long possible ne garantit pas que le résultat soit significatif. Le point crucial est donc d'évaluer l'incertitude du mieux possible ce que nous allons faire ici en utilisant la méthode du *bootstrap* (paragraphe I.6 page 45).

Dans un premier temps la méthode du *bootstrap* a été appliquée sur l'échantillon des configurations extraites de la simulation à pH 5 avec le chromophore en conformation *cis* (860 configurations), afin de tester les paramètres employés. Nous avons donc un échantillon primitif de 860 configurations et les longueurs d'onde associées (ainsi que les forces d'oscillateur). Puis nous avons ré-échantillonné cet ensemble de géométries par la méthode du *bootstrap* ce qui donne N_e nouveaux échantillons (ou échantillon de *bootstrap*) de 860 configurations chacun. Nous avons calculé les spectres correspondant à ces N_e échantillons, les valeurs de λ_{max} de chaque spectre, ainsi que la valeur moyenne et les fluctuations des N_e valeurs de λ_{max} .

La figure II.21 a), présente l'évolution de la valeur moyenne et des fluctuations des longueurs d'onde λ_{max} en fonction de N_e . La convergence est atteinte rapidement pour la valeur moyenne : elle se stabilise à 10% près à partir d'environ $N_e = 250$ échantillons. L'écart type nécessite un plus grand nombre d'échantillons et se stabilise à 10% près à partir de $N_e = 1000$. L'utilisation de 2000 échantillons semble donc tout à fait suffisante pour que la valeur moyenne et l'écart type de λ_{max} , aient convergé et qu'une incertitude puisse être calculée.

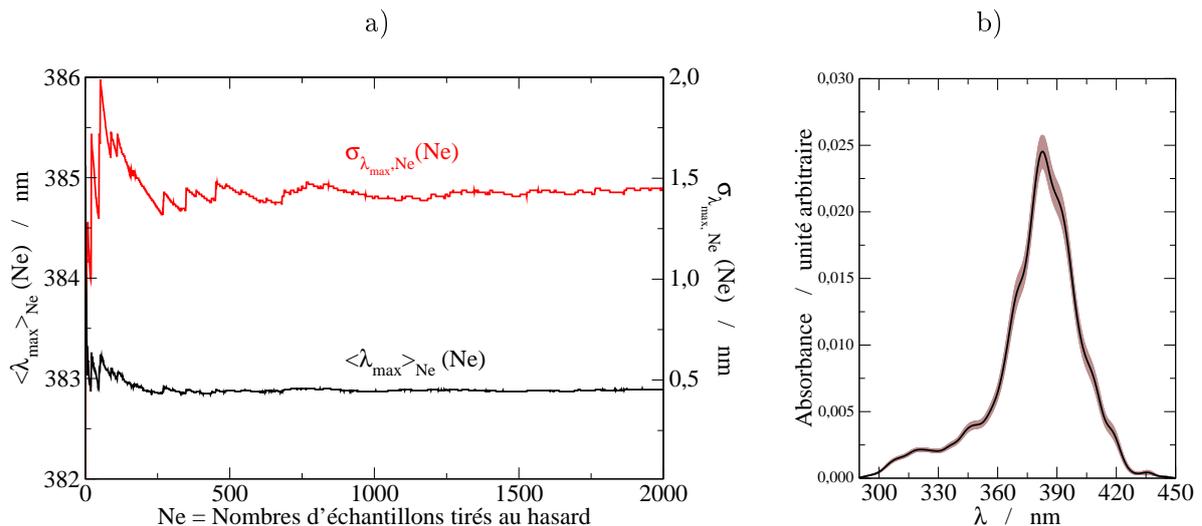


Figure II.21 – a) Variation en fonction du nombre N_e d'échantillons choisis au hasard par la méthode du *bootstrap*, de la valeur moyenne (en noir, échelle à gauche) et de l'écart type (en rouge, échelle à droite) des valeurs de λ_{max} calculés sur les N_e échantillons. b) Spectre d'absorption moyen (en noir) et fluctuations du spectre d'absorption (en marron).

La figure II.21 b) présente le spectre moyen obtenu à partir des $N_e = 2000$ échantillons et les fluctuations autour de ce spectre moyen. Ces fluctuations représentent l'incertitude statistique due au nombre limité de configurations utilisées dans l'échantillon primitif (soit 860 géométries).

En ce qui concerne l'incertitude sur λ_{max} , notée $u(\lambda_{max})$, elle est calculée par l'équation I.31 qui s'écrit ici

$$u(\lambda_{max}) = \sqrt{\left(\lambda_{max}^{spectre} - \langle \lambda_{max} \rangle_{bootstrap}\right)^2 + \sigma_{\lambda_{max}}^2}$$

où $\langle \lambda_{max} \rangle_{bootstrap}$ et $\sigma_{\lambda_{max}}$ sont la moyenne et l'écart type de λ_{max} obtenus à partir des spectres calculés par le ré-échantillonnage du *bootstrap* et $\lambda_{max}^{spectre}$ est la valeur de λ_{max} extraite du spectre de base construit sur l'échantillon primitif. En utilisant les résultats obtenus par la méthode du *bootstrap* avec $N_e = 2000$ (valeur pour laquelle on a vu que la moyenne et l'écart type λ_{max} ont convergé) on obtient $u(\lambda_{max}) = 1.5$ nm. Cette incertitude porte sur la qualité de l'échantillon utilisé pour calculer le spectre. Son ordre de grandeur est cohérent avec les fluctuations du spectre d'absorption observé sur la figure II.21 b).

Il est intéressant de voir comment cette incertitude varie en fonction de la taille de l'échantillon primitif afin de pouvoir fixer une limite inférieure à la taille de cet échantillon. Pour préciser la taille minimale de l'échantillon primitif nous avons calculé l'incertitude $u(\lambda_{max})$ sur une série d'échantillons primitif de taille N_{sub} croissante. Les échantillons de cette série ont été obtenus en sélectionnant N_{sub} configurations réparties uniformément et au hasard parmi les 860 géométries de l'échantillon primitif. La figure II.22 présente la convergence de l'incertitude $u(\lambda_{max}, N_{sub})$ en fonction de N_{sub} .

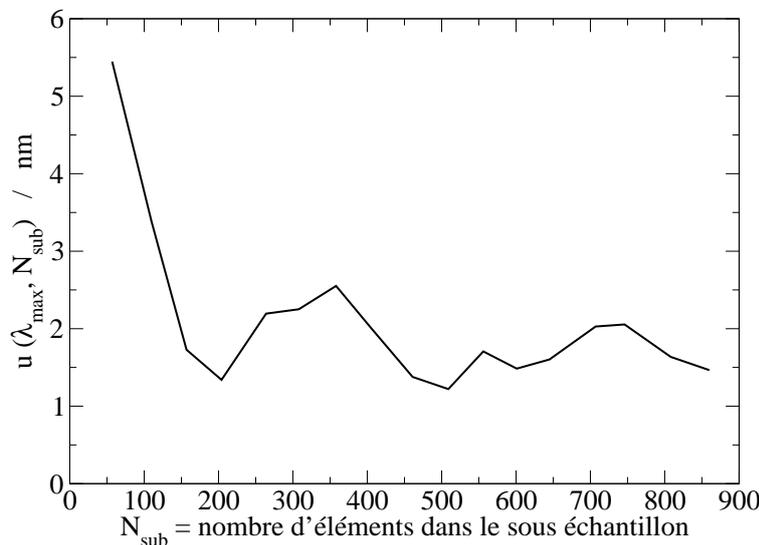


Figure II.22 – Variation de l'incertitude sur λ_{max} en fonction du nombre N_{sub} d'éléments dans l'échantillon primitif.

L'incertitude sur λ_{max} converge rapidement en fonction de N_{sub} . Un ensemble contenant 200 configurations semble en fait suffisant pour obtenir une valeur acceptable du maximum d'absorption avec une incertitude statistique inférieure à 2 nm. Le fait que l'incertitude statistique converge vers

une valeur non nulle est probablement lié à la durée de la simulation (5 ns ici). En effet, pour que l'incertitude statistique soit nulle il faudrait que le nombre de configurations soit infini mais également que ces configurations aient été extraites d'une simulation d'une durée infinie. Dans ce cas la position de λ_{max} serait indépendante des conditions initiales et l'incertitude statistique serait nulle. Il resterait cependant à évaluer l'effet des incertitudes sur le champ de force et la dynamique moléculaire utilisés pour calculer le spectre d'absorption et λ_{max} .

L'incertitude expérimentale sur λ_{max} est de l'ordre du nanomètre. Il est donc inutile de chercher à obtenir une incertitude statistique inférieure à celle obtenue ici. L'échantillon de configurations que nous avons utilisé pour calculer les spectres d'absorption présentés au paragraphe II.3.2 (utilisant environ 800 configurations) et extrait de simulations d'une durée de 5 ns, est donc satisfaisant.

II.3.3.2 Effet de la fonctionnelle sur le calcul du spectre d'absorption

Nous allons maintenant utiliser la méthode du *bootstrap* pour obtenir une mesure de l'incertitude due à la fonctionnelle de la densité. Pour cela, nous avons choisi un ensemble de 8 fonctionnelles parmi les plus utilisées et recouvrant les principaux types existants :

- approximation LDA (*Local Density Approximation*) : SVWN (Fonctionnelle d'échange de Slater [81] et fonctionnelle de corrélation de Vosko, Wilk et Nusair [82]).
- approximation GGA (*Generalised Gradient Approximation*) : PBE (Fonctionnelle d'échange et de corrélation de Perdew, Burke et Ernzerhof [83]), BLYP (fonctionnelle d'échange de Becke [84] et fonctionnelle de corrélation de Lee, Yang et Parr [85]) et BP86 (fonctionnelle d'échange de Becke [84] et fonctionnelle de corrélation de Perdew 1986 [86]).
- approximation meta-GGA (un terme dépendant du laplacien de la densité est ajouté à la fonctionnelle) : TPSS (fonctionnelle d'échange et corrélation de Tao, Perdew, Staroverov et Scuseria [87]).
- fonctionnelles hybrides (contenant une partie d'échange Hartree-Fock) : B3LYP [88] (construite à partir de BLYP), TPSSh [89] (construite à partir de TPSS) et PBE0 [90] (construite à partir de PBE).

La figure II.23 présente les spectres calculés avec chaque fonctionnelle de la liste précédente sur un échantillon contenant 200 configurations réparties de manière homogène le long de la simulation à pH 5 avec le chromophore en conformation *cis*. D'après l'étude présentée au paragraphe précédent (paragraphe II.3.3.1) un échantillon de 200 configurations est suffisant pour avoir une incertitude statistique de l'ordre de l'incertitude expérimentale.

Les spectres de la figure II.23 illustrent la grande dispersion des positions des spectres d'absorption obtenus dans la littérature pour ce type de molécule [72, 73, 79, 91]. Ils recouvrent une large gamme de longueurs d'onde comprise entre 300 et 510 nm. La longueur d'onde du maximum d'absorption varie de 372 à 432 nm, soit dans un intervalle [-10 nm ; 50 nm] autour de la valeur du maximum d'absorption obtenu avec la fonctionnelle B3LYP. Cet intervalle est comparable à l'intervalle [-6 nm ; 37 nm] obtenu pour le chromophore de la ECFP dans le vide ou en phase aqueuse avec un ensemble similaire de fonctionnelles de la densité [91]. Il semble donc que les fluctuations sur la valeur de

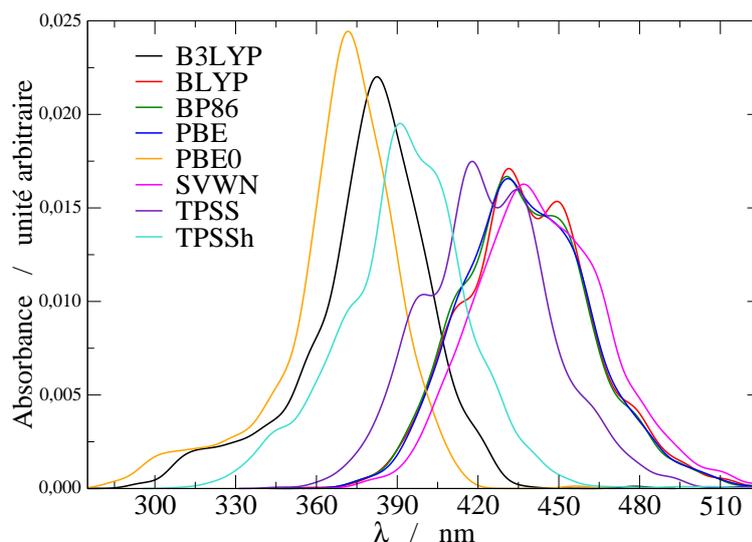


Figure II.23 – Spectres d'absorption calculés avec les fonctionnelles B3LYP, BLYP, BP86, PBE, PBE0, SVWN, TPSS et TPSSH à partir de 200 configurations extraites de la simulation à pH 5 avec le chromophore en conformation *cis*.

la longueur d'onde correspondant au maximum du spectre d'absorption obtenue par TDDFT soit de l'ordre de plusieurs dizaines de nanomètres. Or les décalages du spectre d'absorption que nous cherchons à représenter sont de l'ordre de 11-15 nm et sont donc significativement plus faibles. Compte tenu de ces fluctuations des résultats en fonction de la fonctionnelle de la densité utilisée, il n'est donc pas possible de considérer la méthode TDDFT comme étant suffisamment précise pour faire confiance aux résultats obtenus sur la position des spectres.

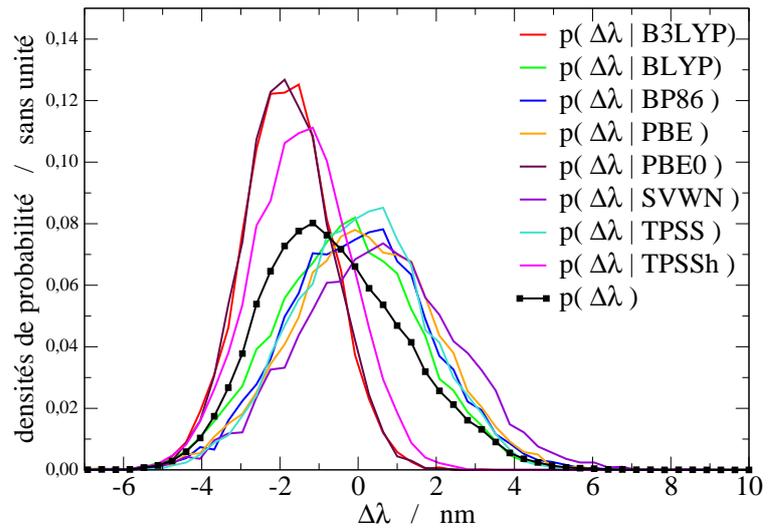
On remarque aussi que les spectres d'absorption obtenus avec des fonctionnelles hybrides sont plutôt étroits et centrés aux alentours de 380 nm, alors que ceux obtenus par les fonctionnelles pures sont plutôt larges et centrés aux alentours de 440 nm.

II.3.3.3 Calcul de l'incertitude sur le décalage du spectre d'absorption sous l'effet du changement de pH

Contrairement à la position absolue du spectre d'absorption, la variation de λ_{max} sous l'effet du changement de pH est nettement plus stable en fonction de la fonctionnelle utilisée. Le décalage du maximum du spectre d'absorption sera noté $\Delta\lambda_{max}$ et sera calculé comme $\Delta\lambda_{max} = \lambda_{max}^{pH8} - \lambda_{max}^{pH5}$. Afin d'évaluer l'incertitude sur $\Delta\lambda_{max}$, nous avons calculé, par la méthode du *bootstrap*, les distributions de $\Delta\lambda_{max}$, en fonction de la fonctionnelle utilisée II.24. Ces distributions sont notées $p(\Delta\lambda_{max}|DFT)$, où *DFT* désigne la fonctionnelle utilisée. Compte tenu de la taille limitée de l'échantillon, la position du maximum du spectre d'absorption est mal définie et dépend de la façon dont le spectre est construit (utilisation d'un histogramme ou somme de gaussiennes de différentes largeurs). Pour simplifier l'analyse des résultats nous avons supposé que le maximum du spectre est égal à la moyenne des longueurs d'onde pondérées par la force d'oscillateur. Cette hypothèse consiste à supposer que les spectres d'absorption sont symétriques.

Les deux effets du changement de pH présentés au paragraphe II.3.2 ont été étudiés : l'effet électrostatique (lié à la modification de la distribution des protons dans la protéine, le chromophore restant dans sa conformation *cis*) et l'effet de l'isomérisation du chromophore. La figure II.24 a) présente les distributions, pour chaque fonctionnelle, du décalage du spectre d'absorption sous l'effet du changement de la distribution de protons (c'est à dire la différence entre les spectres d'absorption correspondant à pH5-*cis* et pH8-*cis*). La figure II.24 b) présente les distributions, pour chaque fonctionnelle, du décalage du spectre d'absorption sous l'effet de l'isomérisation du chromophore (c'est à dire la différence entre les spectres d'absorption correspondant à pH5-*cis* et pH8-*trans*).

a) distributions du décalage du maximum du spectre d'absorption lorsque le pH passe de pH 5 à pH 8 et que le chromophore ne change pas de conformation.
Passage de pH5-*cis* à pH8-*cis*



b) distributions du décalage du maximum du spectre d'absorption lorsque le pH passe de pH 5 à pH 8 et que le chromophore change de conformation.
Passage de pH5-*cis* à pH8-*trans*

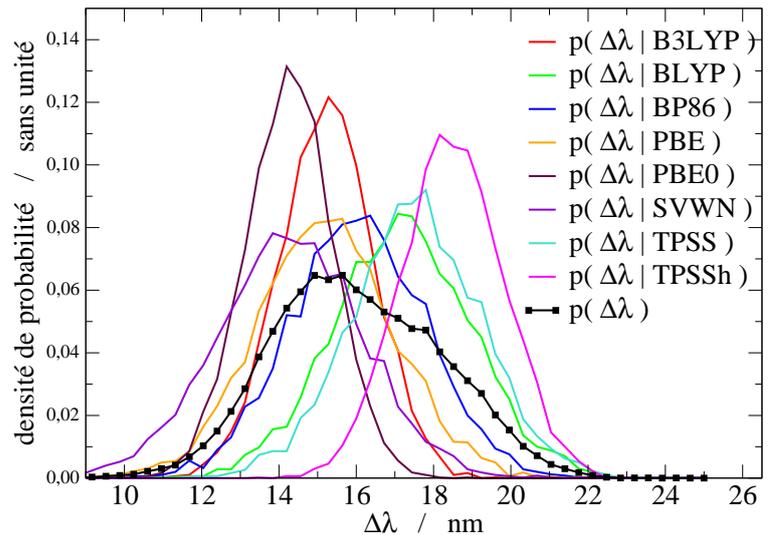


Figure II.24 – Distribution $p(\Delta\lambda_{max}|DFT)$ du décalage du maximum du spectre d'absorption en fonction de la fonctionnelle utilisée a) lorsque le pH varie de pH 5 à pH 8 le chromophore restant en conformation *cis* b) lorsque le pH varie de pH 5 à pH 8 et le chromophore passe de sa conformation *cis* à sa conformation *trans*. La distribution noire avec des carrés est la moyenne des distributions ; elle est obtenue par l'équation II.4.

Les distributions $p(\Delta\lambda_{max}|DFT)$ sont centrées autour de zéro dans le cas de l'effet électrostatique (figure II.24 a). Elles sont centrées autour de 16 nm dans le cas de l'isomérisation du chromophore (figure II.24 b). Ces observations confirment, quelque soit la fonctionnelle, que la contribution majeure au décalage du spectre d'absorption lorsque le pH passe de pH 5 à pH 8 est bien l'isomérisation du chromophore. Les distributions $p(\Delta\lambda_{max}|DFT)$ sont plus étroites dans le cas de fonctionnelles hybrides et sont plus larges pour les fonctionnelles pures. Cette observation est directement liée aux similitudes également observées sur les spectres d'absorption pour des fonctionnelles de même type. Dans le cas de l'effet électrostatique d'absorption, les distributions de $\Delta\lambda_{max}$ sont également centrées sur les mêmes valeurs pour des fonctionnelles de même type. Par contre, dans le cas de l'isomérisation du chromophore les positions des maxima des distributions $p(\Delta\lambda_{max}|DFT)$ sont quelconques.

Afin de calculer une incertitude sur le décalage du spectre d'absorption qui dépendent à la fois de l'échantillon utilisé et de la fonctionnelle utilisée, il faut rassembler les distributions $p(\Delta\lambda_{max}|DFT)$ obtenues pour les différentes fonctionnelles et construire la distribution $p(\Delta\lambda)$. Pour ce faire on utilise l'équation suivante

$$p(\Delta\lambda_{max}) = \frac{\sum_{i=1}^8 p(\Delta\lambda_{max}|DFT_i) p(DFT_i)}{\sum_{i=1}^8 p(DFT_i)}$$

où $p(DFT_i)$ quantifie la qualité des fonctionnelles. Ce facteur peut être extrait à partir de la comparaison des résultats expérimentaux et théoriques obtenus par Timerghazin *et al* [91]. Nous avons commencé ce travail mais il n'est pas encore terminé, et nous considérerons que la valeur de $p(DFT_i)$ est la même quelque soit la fonctionnelle utilisée. La distribution $p(\Delta\lambda)$ est donc simplement construite selon l'équation

$$p(\Delta\lambda_{max}) = \frac{1}{8} \sum_{i=1}^8 p(\Delta\lambda_{max}|DFT_i) \quad (\text{II.4})$$

Les distributions de $p(\Delta\lambda)$ obtenues sont présentées sur les figures II.24 a) et b) en noir avec des carrés. Ces distributions sont centrées sur $\Delta\lambda_{max} = -0.8$ nm dans le cas de l'effet électrostatique, avec une largeur $\sigma_{\Delta\lambda_{max}} = 1.8$ nm, et sur $\Delta\lambda_{max} = 15.9$ nm, avec une largeur $\sigma_{\Delta\lambda_{max}} = 2.2$ nm, dans le cas de l'effet de l'isomérisation du chromophore. Ces résultats sont proches de ceux obtenues avec la fonctionnelle B3LYP, présentés au paragraphe II.3.2. En effet, avec cette fonctionnelle, lorsque le pH passe de pH 5 à pH 8, le décalage du spectre d'absorption est faible dans le cas de l'effet électrostatique ($\Delta\lambda_{max}(B3LYP) = -4$ nm) et il vaut 15 nm lorsque le chromophore passe de sa conformation *cis* à sa conformation *trans*.

A partir des distributions $p(\Delta\lambda)$ nous disposons de tous les éléments nécessaires pour calculer l'incertitude $u(\Delta\lambda_{max})$ qui est donnée par l'équation I.31 et s'écrit dans ce cas précis :

$$u(\Delta\lambda_{max}) = \sqrt{(\langle \Delta\lambda_{max} \rangle_{bootstrap} - \langle \lambda_{max} \rangle_{DFT})^2 + \sigma_{\Delta\lambda_{max}}^2}$$

où $\langle \Delta\lambda_{max} \rangle_{bootstrap}$ et $\sigma_{\Delta\lambda_{max}}$ sont la moyenne et l'écart type associés à la distribution $p(\Delta\lambda_{max})$ (figures II.24 a et b) et $\langle \lambda_{max} \rangle_{DFT}$ est la moyenne des décalages des maxima des spectres d'absorption obtenus à partir des 8 fonctionnelles utilisés et des échantillons primitifs (figure II.23). Les incertitudes obtenues sont rassemblées dans le tableau II.8.

	$\Delta\lambda_{max}$ (nm)	$u(\Delta\lambda_{max})$ (nm)
effet électrostatique	-0.7	1.8
isomérisation	15.9	2.2

Tableau II.8 – Valeur moyenne de $\Delta\lambda_{max}$ et incertitude $u(\Delta\lambda_{max})$ compte tenu de l'effet de la taille de l'échantillon utilisé et de la fonctionnelle de la densité utilisée.

Les incertitudes obtenues (de l'ordre de 2 nm) sont faibles devant le décalage du spectre d'absorption observé lors de l'isomérisation du chromophore (de l'ordre de 15 nm). Les calculs de spectres d'absorption effectués avec la méthode TDDFT sont donc suffisamment précis pour représenter un décalage du maximum du spectre d'absorption de cet ordre de grandeur. Par contre, le décalage de 4 nm vers le rouge observé avec B3LYP pour l'effet électrostatique semble ne pas être significatif car il est du même ordre de grandeur que l'incertitude sur le maximum du spectre. On peut donc conclure de façon fiable que l'isomérisation du chromophore induit un décalage du maximum spectre d'absorption de l'ordre de 15 nm et que le changement de la distribution des protons dans la protéine a un effet faible, de l'ordre de l'incertitude expérimentale, sur la position de ce spectre.

II.3.4 Transitions électroniques impliquées dans les spectres d'absorption

Pour construire le spectre d'absorption, nous avons inclus dans le calcul les quatre premières transitions électroniques. Nous avons analysé ces transitions électroniques à partir des calculs TDDFT effectués au niveau B3LYP/6-31G* dans les trois cas présentés précédemment (pH5-*cis*, pH8-*cis* et pH8-*trans*). Les principales orbitales mises en jeu dans le cas du chromophore en conformation *cis* sont présentées sur la figure II.25.

Les orbitales HOMO et LUMO sont des orbitales de type π délocalisées sur l'ensemble du système π du chromophore. L'orbitale moléculaire 95 est de type σ . Elle est principalement développée sur les paires libres de l'azote N13 et du carbonyle de l'imidazolinone, et sur une partie du système σ de ce même cycle. L'orbitale moléculaire 96 est une orbitale π exclusivement développée sur le cycle indole. Dans un certains nombre de configurations on observe un mélange des orbitales 95 et 96. Les orbitales moléculaires obtenues dans la conformation *trans* sont similaires mais le mélange entre les orbitales 95 et 96 n'est pas observé.

Quelque soit le cas considéré (pH ou conformation du chromophore), l'opérateur d'échelle (paragraphe D.3.1 page 183) qui construit l'état excité à partir de l'état fondamental est majoritairement développé sur la transition mono-électronique correspondant à la promotion d'un électron de l'orbitale HOMO vers l'orbitale LUMO. La transition électronique du premier état excité est donc de type $\pi\pi^*$ et c'est celle qui présente le plus souvent la plus grande force d'oscillateur (tableau II.9).

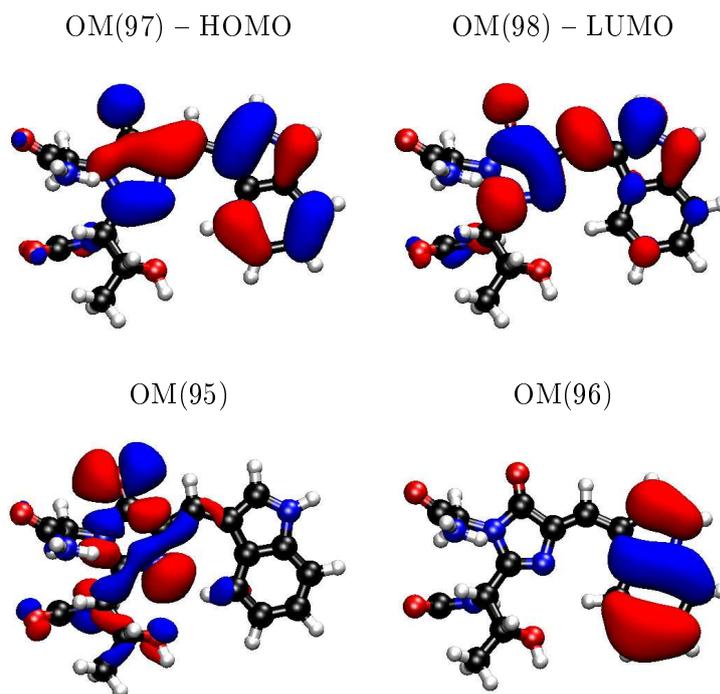


Figure II.25 – Orbitale moléculaire du chromophore de la Cerulean en conformation *cis* obtenues au niveau B3LYP/6-31G*. La valeur de l'orbitale sur l'isosurface tracée est 0.02.

Bien que l'orbitale LUMO soit un peu moins développée sur la partie indole du chromophore, (figure II.25), le transfert de charge lors de la promotion d'un électron de l'orbitale HOMO vers l'orbitale LUMO est faible. Or, le transfert HOMO-LUMO étant prépondérant dans le premier état excité, le transfert de charge lors du passage de l'état fondamental à l'état excité est donc faible, ce qui explique le fait que le spectre d'absorption ne dépend que faiblement du champ électrostatique créé par la protéine.

transition	pH5- <i>cis</i>	pH8- <i>cis</i>	pH8- <i>trans</i>
1	0.41 (0.18)	0.31 (0.21)	0.52 (0.15)
2	0.12 (0.17)	0.23 (0.22)	0.08 (0.11)
3	0.08 (0.08)	0.06 (0.07)	0.03 (0.05)
4	0.03 (0.04)	0.02 (0.03)	0.06 (0.06)

Tableau II.9 – Valeur moyenne et fluctuations (entre parenthèses) de la force d'oscillateur associée au deux premiers états excités en fonction des différentes situations envisagées concernant le pH et la conformation du chromophore.

La force d'oscillateur des transitions 3 et 4 est faible quelque soit la simulation. Dans les simulations pH5-*cis* et pH8-*cis*, où le chromophore est en conformation *cis*, la force d'oscillateur de la première transition électronique est plus faible que dans la simulation pH8-*trans*, où le chromophore est en conformation *trans*. La force d'oscillateur de la deuxième transition est assez importante dans les cas

pH5-*cis* et pH8-*cis* alors qu'elle est faible dans le cas pH8-*trans* (tableau II.9). Ceci est lié au poids dans l'opérateur d'échelle des deux premiers état électroniques des transitions mono-électroniques entre les orbitales occupées 95, 96 et l'orbitale vacante 98. Ces transitions sont respectivement de type $n\pi^*$ et à transfert de charges et elles ont une faible force d'oscillateur. En conformation *trans*, leur poids est négligeable dans l'opérateur d'échelle du premier état excité et majoritaire dans l'opérateur d'échelle du second état excité, ce qui explique que la force d'oscillateur de la première transition soit élevée et que celle de la deuxième soit faible. En conformation *cis*, les transitions mono-électroniques faisant intervenir les orbitales 95 et 96 contribuent faiblement à l'opérateur d'échelle du premier état excité et elles contribuent à poids égal avec la transition mono-électronique HOMO-LUMO, à l'opérateur d'échelle du deuxième état excité. Ce mélange entre des transitions de type $\pi\pi^*$ et une transition de type $n\pi^*$ est rendu possible grâce au mélange entre les orbitales 95 et 96. Il abaisse la force d'oscillateur de la première transition et augmente la force d'oscillateur de la deuxième. Dans la conformation *cis*, la deuxième transition contribue donc de manière importante au spectre d'absorption.

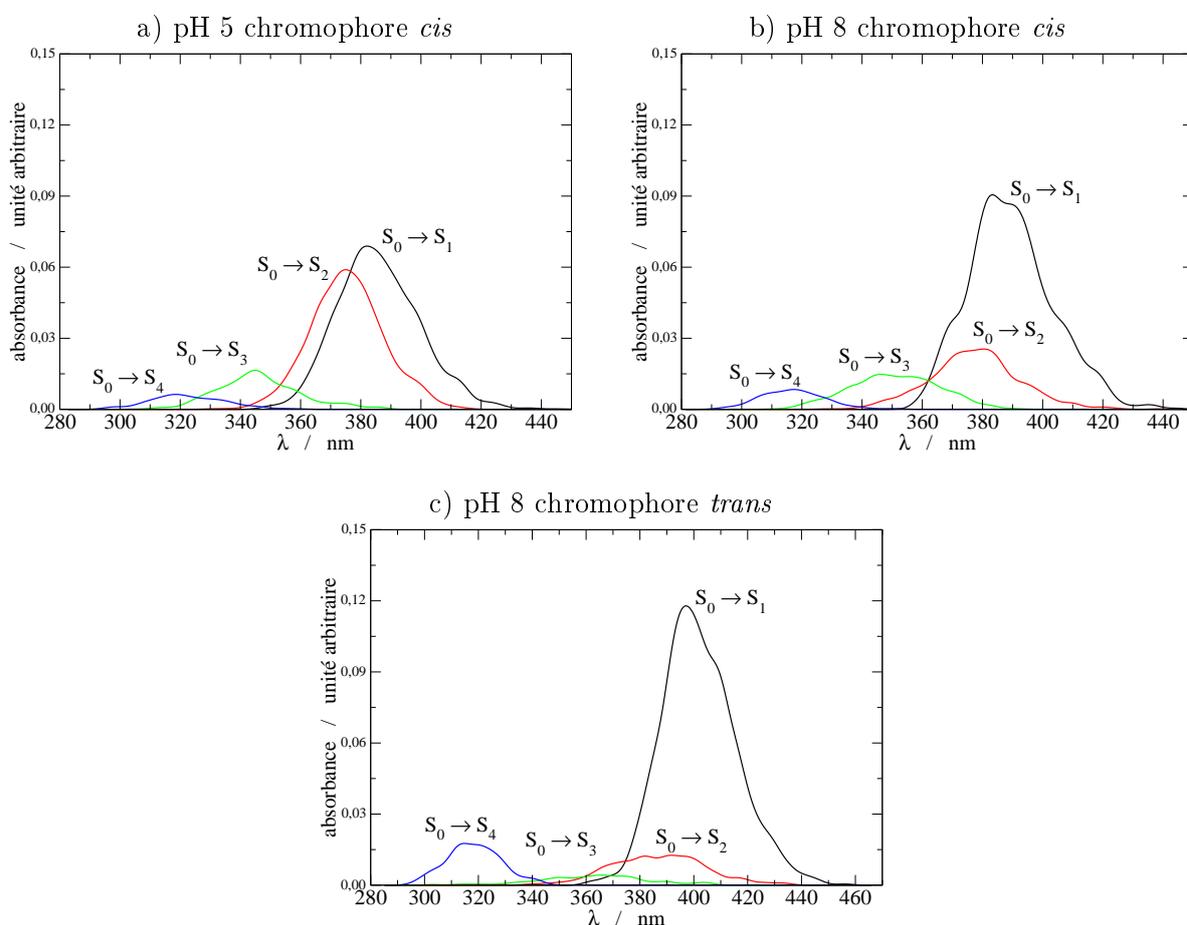


Figure II.26 – Spectres d'absorption a) pour la simulation pH5-*cis* b) pour la simulation pH8-*cis* et c) pour la simulation pH8-*trans* où la contribution à l'absorbance de chaque état excité est explicitée : en noir le premier état excité, en rouge le deuxième état excité, en vert le troisième état excité et en bleu le quatrième état excité.

La figure II.26 présente les spectres d'absorption où la contribution de chaque état excité a été ex-

plicitée. Les contributions à l'absorbance de chaque état excité aux spectres d'absorption montrent clairement que dans la conformation *cis* le deuxième état excité contribue de façon non négligeable au spectre d'absorption et que cette contribution dépend du pH et donc dans ce cas de la distribution de protons (le chromophore restant en conformation *cis*). Lorsque le pH passe de 8 à 5, la contribution du deuxième état excité augmente et le spectre d'absorption se décale vers le bleu, comme observé sur la figure II.20. Les contributions des troisièmes et quatrième états excités sont faibles et indépendantes du pH. Dans la simulation pH8-*trans* où le chromophore est en conformation *trans*, la bande principale centrée sur 400 nm est exclusivement due à l'absorption du premier état excité. Ceci est dû au fait que la transition électronique est principalement développée sur la transition mono-électronique de type $\pi\pi^*$ de l'orbitale HOMO vers l'orbitale LUMO et ne se mélange pas avec la transition $n\pi^*$ comme en conformation *cis*. Cette absence de mélange fait que le deuxième et le troisième état excité ont une absorbance nulle. Le quatrième état excité est responsable de la seconde bande d'absorption vers 320 nm. Cependant, ces observations ne permettent pas d'expliquer le décalage de 15 nm du spectre d'absorption lorsque le chromophore passe de la conformation *cis* à la conformation *trans*.

D'un point de vue qualitatif ce décalage du spectre d'absorption peut s'expliquer en considérant, dans chacune des conformations du chromophore, les positions relatives des cycles indole et imidazolinone. Ils ont plus proches l'un de l'autre dans la conformation *cis* que dans la conformation *trans* (figure II.1 page 51). Le recouvrement entre le système π de l'indole et celui de l'imidazolinone est donc plus important dans la conformation *cis*. Dans une approche de type Hückel étendue, cette augmentation de l'interaction entre les deux parties du système π est synonyme d'une augmentation (en valeur absolue) des termes hors diagonaux de l'hamiltonien, ce qui a pour effet d'augmenter les écarts entre les énergies des orbitales moléculaires du système π . Les énergies des orbitales moléculaires devraient donc être plus espacées dans la conformation *cis* que dans la conformation *trans* ce qui conduirait à un décalage vers le rouge lors d'une isomérisation de la conformation *cis* vers la conformation *trans*. Les énergies des orbitales HOMO et LUMO ont été extraites des calculs TDDFT (au niveau B3LYP/6-31G*) utilisés pour construire les spectres d'absorption. Les distributions de la différence d'énergie entre l'orbitale HOMO et l'orbitale LUMO dans les trois situations pH5-*cis*, pH8-*cis* et pH8-*trans*, sont présentées sur la figure II.27.

Comme attendu, la différence d'énergie entre les deux orbitales est plus faible dans la conformation *trans* que dans la conformation *cis*. On constate aussi qu'elle est peu affectée par le changement de pH sans isomérisation. Les positions relatives des trois distributions sont semblables aux positions relatives des trois spectres d'absorption (compte tenu du fait que la longueur d'onde varie en sens inverse de la différence d'énergie). Le fait qu'il y ait une correspondance directe entre, d'une part, la différence d'énergie entre l'orbitale HOMO et l'orbitale LUMO et, d'autre part, les spectres d'absorption, est relié au fait que la transition électronique vers le premier état excité consiste essentiellement en la promotion d'un électron de l'orbitale HOMO vers l'orbitale LUMO. Ces observations sont donc en faveur d'une justification du décalage du spectre d'absorption vers le bleu par une perturbation du système π du chromophore, lorsqu'il passe de la conformation *trans* à la conformation *cis*.

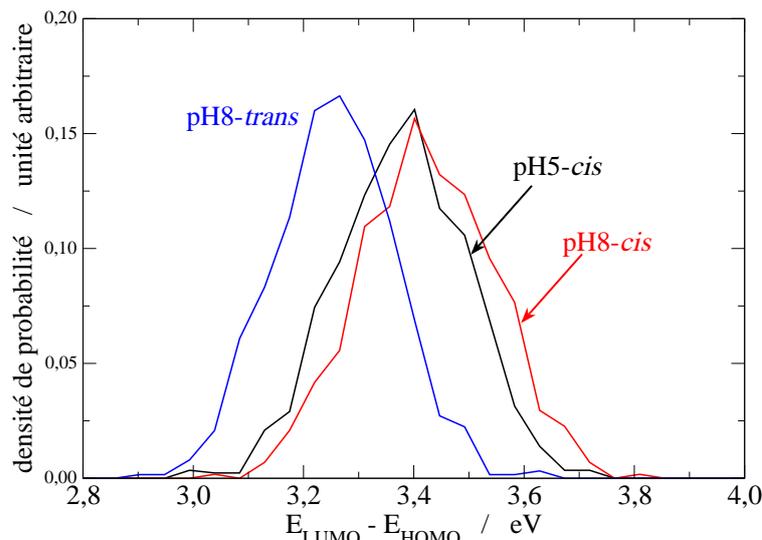


Figure II.27 – Densités de probabilité de la différence d'énergie entre l'orbitale HOMO et l'orbitale LUMO du chromophore construites à partir des calculs TDDFT utilisés pour faire les spectres d'absorption, à pH 5 lorsque le chromophore est en conformation *cis* (en noir), à pH 8 lorsque le chromophore est en conformation *cis* (en rouge) et à pH 8 lorsque le chromophore est en conformation *trans*.

II.4 Vers une description plus complète de l'effet du pH

Dans ce chapitre nous avons étudié la structure de la Cerulean à pH 5 et à pH 8. Au paragraphe II.2.1 nous avons décrit comment nous avons modélisé le pH dans nos simulations. Pour chaque pH nous avons calculé, avec le serveur H++, le micro-état de protonation le plus probable au pH considéré (pH 5 ou pH 8) et nous avons choisi de représenter la distribution des protons dans la protéine par ce micro-état de protonation. Or nous avons également fait remarquer qu'à un pH donné, les probabilités associées à plusieurs micro-états de protonation différents sont très proches. L'utilisation d'un seul micro-état de protonation pour décrire la distribution des protons dans une protéine est donc une approximation importante. Une prise en compte plus globale de l'effet du pH sur une propriété donnée implique d'évaluer la moyenne de cette propriété sur une distribution de micro-états de protonation. Ceci est *a priori* difficile, voir impossible à réaliser compte tenu du nombre de micro-états et donc de calculs à effectuer.

Nous proposons ici une méthode permettant de modéliser d'une manière plus globale l'effet du pH sur un spectre d'absorption. Le développement de cette méthode est encore en cours et seulement quelques résultats préliminaires sont présentés.

A partir de l'expression de la probabilité d'un micro-état de protonation, on peut obtenir par une simulation de type Monte-Carlo la distribution des micro-états de protonation [92]. Pour tenir compte de ces micro-états dans une configuration donnée, notre approche consiste à estimer les fluctuations de la longueur d'onde d'absorption en fonction des variations du champ électrostatique ressenti par le chromophore, au moyen d'une expression analytique prédéterminée. Dans une première approche on se placera à une géométrie fixe qui est la géométrie de la structure RX de la Cerulean.

II.4.1 Simulation Monte Carlo dans l'espace des micro-états de protonation

Dans une protéine du type de la Cerulean, 77 acides aminés présentent une fonction acido-basique. Chaque acide aminé peut se trouver sous sa forme acide ou basique ce qui fait donc 2^{77} (soit environ $1.5 \cdot 10^{23}$) micro-états de protonation possibles, un micro-état de protonation étant défini par la donnée de l'état de protonation de chacun des N acides aminés acido-basiques de la protéine. On peut caractériser chaque micro-état de protonation par un vecteur $X = (x_1, x_2, \dots, x_N)$ de dimension N , x_i prenant la valeur 1 si l'acide aminé i est sous sa forme acide et la valeur 0 si il est sous sa forme basique. A chaque micro-état de protonation est associé une énergie libre, $\Delta\mathcal{F}(X)$, calculée par rapport à un micro-état de protonation de référence dans lequel tous les acides aminés acido-basiques sont neutres qui s'exprime selon (paragraphe I.5 page 42)

$$\Delta\mathcal{F}(X) = -k_B T \sum_i x_i \ln 10 (pK_{intr,i} - pH) + \frac{1}{2} \sum_{i,j} (x_i + q_i^0) (x_j + q_j^0) W_{ij} \quad (\text{II.5})$$

Les valeurs de $pK_{intr,i}$ et W_{ij} sont données par le serveur de calcul H++ pour l'ensemble des acides aminés acido-basiques de la protéine et on peut donc calculer $\Delta\mathcal{F}$ pour n'importe quel micro-état de protonation. Mais les ordinateurs actuels ne permettent pas de calculer la valeur de $\Delta\mathcal{F}$ pour l'ensemble des micro-états de protonation de la protéine^a.

Afin d'obtenir les micro-états de protonation les plus probables nous avons mis en oeuvre une simulation Monte Carlo, dans l'espace des micro-états de protonation. Dans cette simulation, les coordonnées que l'on considère sont les x_i caractérisant l'état de protonation de l'acide aminé i . A chaque pas de la simulation un acide aminé est choisi au hasard et son état de protonation est inversé. La probabilité avec laquelle on accepte ce changement d'état de protonation est déduite du calcul de la variation de $\Delta\mathcal{F}$ due à ce changement d'état de protonation.

La figure II.28 présente l'évolution en fonction du nombre de pas Monte Carlo du $\Delta\mathcal{F}$ des micro-états de protonation et de la moyenne de $\Delta\mathcal{F}$ obtenus pour la structure RX de la Cerulean à pH 7 et à 300 K. La moyenne du $\Delta\mathcal{F}$ converge après environ 10000 pas Monte-Carlo. L'ensemble des 10000 valeurs de $\Delta\mathcal{F}$ est alors une répartition d'équilibre que nous allons étudier.

II.4.2 Densité de probabilité des micro-états de protonation

Nous avons réalisé des simulations Monte-Carlo pour plusieurs valeurs du pH allant de pH 1 à pH 14 par pas de 1. Pour chaque simulation, nous avons obtenu une distribution des valeurs de $\Delta\mathcal{F}$ comme nous venons de le décrire. Trois de ces distributions, pour pH 5, pH 7 et pH 10 sont présentées sur la figure II.29.

On remarque que, suivant le pH, la distribution des micro-états de protonation a une allure très différente. A pH 5, la distribution des micro-états de protonation présente une forte densité d'état

^aPar un calcul simple, en considérant qu'un ordinateur réalise 10^{10} opérations par seconde et qu'une seule opération est nécessaire pour calculer le $\Delta\mathcal{F}$ associé à un micro-état de protonation, plus de $3 \cdot 10^5$ années seraient nécessaires pour réaliser ces calculs.

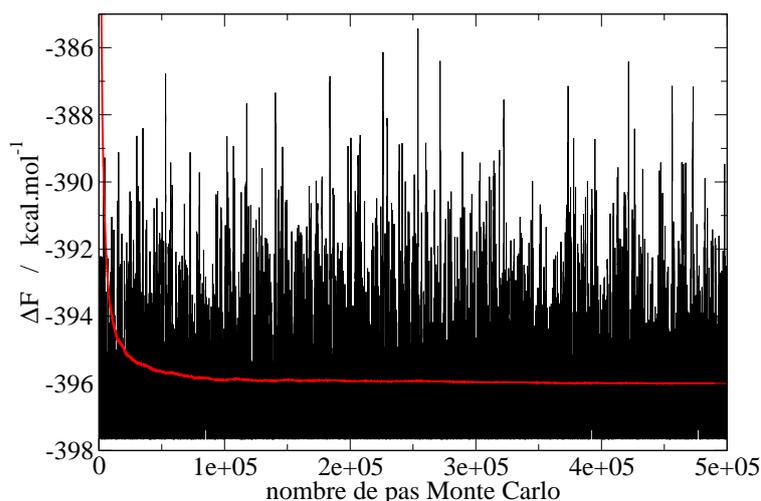


Figure II.28 – Différence d'énergie libre, $\Delta\mathcal{F}$, à pH 7, entre l'état échantillonné par Monte Carlo et l'état de référence (dans cet état, tous les acides aminés acido-basiques sont neutres), en fonction du nombre de pas Monte Carlo (en noir). Valeur moyenne de $\Delta\mathcal{F}$ en fonction du nombre de pas Monte Carlo (en rouge).

entre 0 et 12 kcal·mol⁻¹ alors qu'elle est beaucoup plus faible à pH 7. A pH 10 la situation est intermédiaire entre celle observée à pH 5 et pH 7. Pour quantifier le caractère plus ou moins concentré de cette densité, nous avons calculé en fonction du pH le nombre N_{mep} de micro-états de protonation qui doivent être retenus pour que la probabilité que la protéine adopte l'un d'entre eux à température ambiante soit égale à 0.8. Ce nombre est obtenu en intégrant la densité de probabilité de $\Delta\mathcal{F}$ obtenue à chaque pH. Les résultats sont présentés sur la figure II.30.

On remarque que le nombre de micro-états de protonation N_{mep} est très variable suivant le pH. Il est très élevé, plusieurs dizaines de milliers, pour des pH compris entre pH 3 et pH 6, il est de quelques milliers pour des pH compris entre pH 10 et pH 13, et de quelques dizaines pour les pH compris entre pH 7 et pH 9.

Les valeurs des pKa en phase aqueuse des acides aminés acido-basiques sont également reportées en abscisse sur la figure II.30 (en rouge) avec en ordonnée le nombre de fois où cet acide aminé apparaît dans la séquence de la Cerulean. On remarque que les valeurs de N_{mep} sont fortement corrélées au nombres d'acides aminés acido-basiques dont le pKa se situe dans la même gamme de pH. Cette corrélation est probablement due au fait que les interactions entre acides aminés dans la Cerulean sont faibles et les $pK_{\frac{1}{2}}$ dans la protéine sont alors peu différents des pKa en solution. Donc pour des valeurs de pH proches des valeurs de pKa des acides aminés en solution un grand nombre d'acides aminés ont une probabilité proche de 1/2 d'être sous leur forme acide ou basique. Le changement de l'état de protonation de tels acides aminés aura donc peu d'influence sur la valeur du $\Delta\mathcal{F}$ des micro-états de protonation. Ainsi plus le nombre d'acides aminés pour lesquels les formes acide et basique sont équiprobables sera grand, plus le nombre de micro-états de protonation ayant des valeurs d'énergie libre proches sera grand. Le fait que les $pK_{\frac{1}{2}}$ dans la Cerulean soient peu différents des pKa en solution est dû au fait que la majorité des acides aminés acido-basiques ont leur chaîne

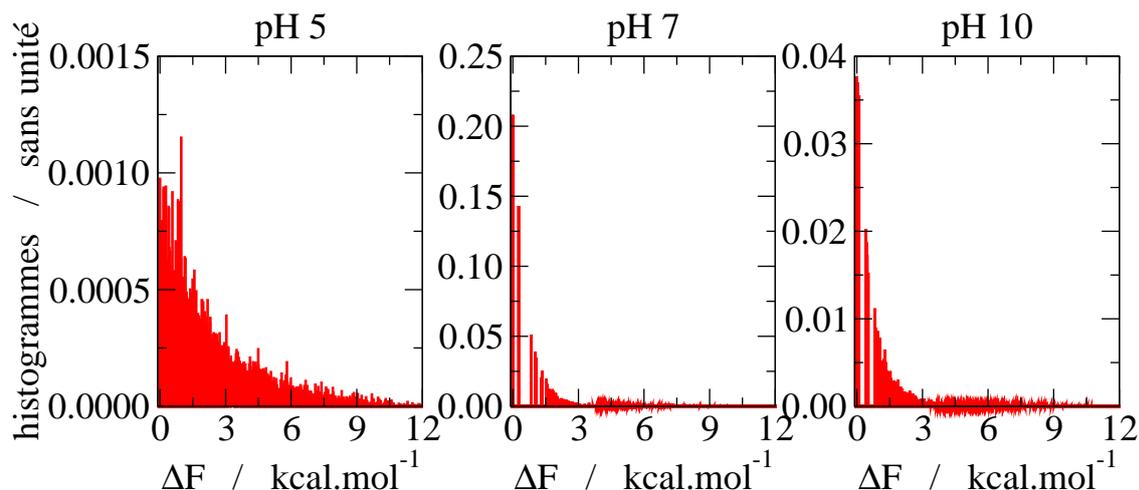


Figure II.29 – Histogrammes représentant la probabilité des micro-états de protonation en fonction de leur énergie libre associée à différents pH. La valeur $\Delta\mathcal{F}=0 \text{ kcal}\cdot\text{mol}^{-1}$ a été attribuée au micro-état le plus probable.

latérale orientée vers le solvant.

Ces simulations Monte Carlo nous permettent donc d'identifier les micro-états de protonation les plus probables sans avoir à réduire le nombre d'acides aminés acido-basiques inclus dans le calcul. On peut également en extraire le nombre de micro-états de protonation qui doivent être pris en compte pour décrire correctement la situation d'équilibre. Il faut maintenant trouver un moyen d'évaluer les fluctuations dues à toutes ces distributions de protons sur les observables qui nous intéressent.

II.4.3 Effet des micro-états de protonation sur une observable

D'un micro-état de protonation à l'autre, la géométrie étant fixe, seule la position des protons acides change dans la protéine. Cette redistribution des protons a pour effet de modifier le champ électrostatique dans la protéine et notamment au niveau du chromophore. C'est l'effet de ce changement du champ électrostatique sur une longueur d'onde que nous devons maintenant calculer.

Pour les pH compris entre pH 6 et pH 9, le nombre de micro-états de protonation à considérer étant inférieur à 100, on peut envisager de faire un calcul quantique des longueurs d'onde d'absorption pour l'ensemble des micro-états. Par contre cette solution n'est plus envisageable si le nombre de micro-états de protonation dépasse plusieurs milliers. Une autre approche doit donc être considérée et deux possibilités sont envisagées.

La première approche consiste à calibrer les longueurs d'onde d'absorption du chromophore pour un ensemble de champs électrostatiques extérieurs, par un ensemble de calculs Gaussian en présence de champs électrostatiques connus.

La deuxième approche consiste à utiliser un hamiltonien effectif de dimension deux dont les vecteurs de base représentent l'état fondamental et l'état excité. On peut ensuite calibrer les éléments de

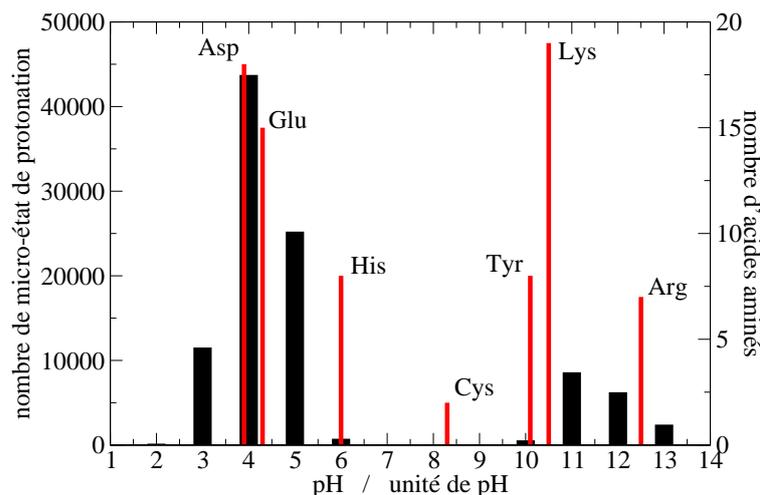


Figure II.30 – En noir, échelle de gauche, nombre de micro-états de protonation nécessaires pour obtenir un échantillon de micro-états de protonation tel que la probabilité que le micro-état de protonation de la protéine appartienne à cet échantillon soit de 0.8. En rouge, valeur du pKa (en abscisse) des acides aminés acido-basiques et, en ordonnée (échelle de droite) leur nombre dans la Cerulean.

cette matrice en fonction du potentiel électrostatique créé par un champ électrostatique extérieur. Une façon simple de faire cela est d'utiliser des charges effectives associées à chaque élément de la matrice. Il serait alors aisé de calculer l'effet d'un micro-état de protonation ou d'un changement de micro-état, c'est à dire du potentiel électrostatique, sur les éléments de cette matrice. La longueur d'onde d'absorption serait alors donnée par la diagonalisation de cette matrice une fois que les éléments auront été calculés pour un micro-état de protonation donné.

II.5 Conclusion du chapitre

Dans ce chapitre, nous avons présenté une étude de la Cerulean par des simulations de dynamique moléculaire et des calculs de chimie quantique. Nous avons réalisé plusieurs simulations de dynamique moléculaire à pH 5 avec le chromophore en conformation *cis* et à pH 8 avec le chromophore en conformation *cis* et *trans*. Ces calculs nous ont permis de vérifier plusieurs des hypothèses faites par Malo *et al* pour expliquer la dépendance en pH du spectre d'absorption.

La chaîne latérale de l'acide aspartique en position 148 s'oriente à l'intérieur de la protéine lorsque celui-ci est sous sa forme acide et forme une liaison hydrogène stable avec le cycle indole du chromophore en conformation *cis*. Cependant, la liaison hydrogène intra-acide aminé proposée par les auteurs de la structure RX à pH 5, entre l'un des atomes d'oxygène de la fonction acide carboxylique de Asp148 et l'oxygène du carbonyle de ce même acide aminé, n'a jamais été observée dans nos simulations. La chaîne latérale de Asp148 adopte une conformation qui lui permet de faire une liaison hydrogène avec la lysine en position 166.

Nous avons pu observer que la formation de cette liaison hydrogène avec Lys166 est liée à une fermeture partielle du *cleft* entre les brins 6 et 7 du tonneau de feuillet β . En étudiant les contacts

entre protéines voisines dans la structure cristalline de la Cerulean nous avons observé que, dans le cristal de la Cerulean, des contacts entre protéines voisines pourraient être responsable d'une ouverture partielle du *cleft*. Ces interactions disparaissant en solution, le *cleft* adopterait alors une structure plus proche de celle de la GFP dans le cristal ou en phase aqueuse.

Lorsque Asp148 est sous sa forme basique, anionique, sa chaîne latérale s'oriente vers l'extérieur de la protéine. La liaison hydrogène avec le cycle indole du chromophore est alors rompue. Dans cette situation, la conformation *trans* du chromophore est préférable car elle lui permet de retrouver de nouveaux partenaires pour former un réseau dense de liaisons hydrogène.

Nous avons calculé le spectre d'absorption de la Cerulean à pH 5 en conformation *cis* et à pH 8 en conformation *cis* et *trans*. Cela nous a permis de séparer l'effet du changement du micro-état de protonation de la protéine qui est un effet électrostatique, de l'effet de l'isomérisation du chromophore. L'effet du changement de champ électrostatique sur le spectre d'absorption du chromophore est faible, alors que l'isomérisation du chromophore résulte en un décalage du spectre très semblable à celui observé expérimentalement lors du passage de pH 8 à pH 5. Nos résultats confortent ainsi nettement 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 pH 5 et pH 8. Nous avons débuté une étude plus rigoureuse et systématique des effets de pH sur le spectre d'absorption. Au delà des propriétés d'absorption de la Cerulean, cette approche pourrait être appliquée à l'effet du pH sur toute propriété sensible aux variations de potentiel électrostatique.

Chapitre III

Désactivation de la fluorescence de la *Green Fluorescent Protein*

Sommaire

III.1 Introduction	106
III.1.1 Etat de protonation du chromophore de la GFP	106
III.1.2 Etudes expérimentales du chromophore en solution	107
III.1.3 Etudes quantiques sur l'état excité anionique du chromophore	107
III.1.4 Mouvements du chromophore dans la protéine	109
III.1.5 Processus de désactivation de la fluorescence étudiés dans ce travail	111
III.2 Méthodologie	111
III.3 Résultats de dynamique moléculaire	113
III.3.1 Simulations du chromophore de la GFP en phase aqueuse	113
III.3.2 Simulation du chromophore de la GFP dans la protéine sans contrainte	118
III.3.3 Profil d'énergie libre partielle en fonction de φ	122
III.4 Calcul du déclin de la fluorescence	134
III.4.1 Modèle cinétique de l'état excité fluorescent	135
III.4.2 Echantillonnage d'évènement rare par dynamique Brownienne	137
III.4.3 Intégration des équations cinétiques	145
III.5 Conclusions et perspectives	153

Le but de ce chapitre est de présenter une méthode permettant de calculer le déclin de la fluorescence et le rendement quantique de fluorescence d'une protéine en prenant en compte des mécanismes de désactivation non radiatifs de l'état électronique excité associés à une déformation du chromophore.

Cette méthode a été appliquée à l'étude de la désactivation de la fluorescence suite à un mouvement de torsion d'angle φ dans la GFP. Nous avons choisi ce système car il a été largement étudié expérimentalement et par des méthodes numériques. De nombreux résultats sont donc à notre disposition pour paramétrer les méthodes utilisées ou comparer les résultats obtenus avec les données expérimentales. Après avoir présenté les résultats connus sur les propriétés photophysiques de la GFP et de son chromophore, la méthode proposée sera présentée.

III.1 Introduction

Le temps de vie de la GFP sauvage est d'environ 3 ns et son rendement quantique de fluorescence est élevé (0.8) [93]. La fluorescence de la GFP et de son chromophore, seul en solution plus ou moins visqueuse ont souvent été étudiées expérimentalement. On trouve également de nombreuses études de l'état électronique excité du chromophore par des méthodes de chimie quantique. Le but de cette partie est d'en extraire une vue d'ensemble sur la dynamique et les processus à l'oeuvre dans l'état excité fluorescent du chromophore isolé en phase gaz, en solution ou dans la protéine GFP.

III.1.1 Etat de protonation du chromophore de la GFP

La fluorescence de la GFP sauvage est en fait une superposition de la fluorescence de deux formes du chromophore : une forme neutre et une forme anionique dans laquelle le phénol est sous sa forme basique, c'est à dire phénolate [7]. De nombreuses études montrent que la déprotonation du phénol est la conséquence d'un transfert de proton le long d'une chaîne de liaisons hydrogène impliquant une molécule d'eau, une sérine (Ser205) et un acide glutamique (Glu222) qui est l'accepteur final du proton [94,95,96].

Le mutant S65T (code 1EMA de la PDB [18]) conduit à stabiliser fortement la forme anionique du chromophore [97]. La substitution de la sérine en position 65 par une thréonine correspond à l'ajout d'un groupement méthyle. Brejc *et al* [97] ont comparé les structures RX de la GFP sauvage et du mutant S65T et concluent que l'ajout d'un méthyle induit une gêne stérique conduisant à la rupture de la liaison hydrogène entre le groupement hydroxyle de Ser65 (accepteur de doublet) et le groupement carboxylate de Glu222 (donneur de doublet) au profit du groupement carbonyle de Val61 (donneur de doublet remplaçant Glu222). Cela stabilise la forme neutre de Glu222 obtenue après transfert du proton provenant du phénol du chromophore. Glu222 forme alors une liaison hydrogène en tant qu'accepteur de doublet avec le groupement hydroxyle de Thr65. Dans ce travail, nous avons étudié ce mutant, dans lequel le chromophore est toujours anionique à l'état électronique fondamental et excité.

III.1.2 Etudes expérimentales du chromophore en solution

Les caractéristiques de la fluorescence du chromophore de la GFP en solution, neutre ou anionique, sont radicalement différentes de celles observées dans la protéine.

Le rendement quantique de fluorescence en solution est mille fois plus petit que celui qui est observé dans la protéine. Les déclin de fluorescence sont correctement modélisés par une somme de deux exponentielles dont les constantes de temps sont de l'ordre ou inférieures à la picoseconde [98, 99, 100, 101, 102] (0.21 ps et 1.1 ps pour le chromophore anionique dans l'eau [99] ; 0.8 ps et 4.2 ps dans l'éthanol [98]).

Kummer *et al* [98] ont mesuré le déclin de fluorescence du chromophore de la GFP en solution en faisant varier la viscosité de plusieurs ordres de grandeur, notamment en faisant varier la température. Ils en déduisent que le déclin de fluorescence est fortement dépendant de la viscosité du solvant. Ils trouvent que dans une solution de glycerol à 150 K, le déclin de fluorescence est de l'ordre de la nanoseconde, comme dans la protéine. Ils en concluent que le mécanisme de désactivation de la fluorescence qui conduit à un déclin de la fluorescence de l'ordre de la picoseconde en solution a lieu par l'intermédiaire d'une déformation interne du chromophore nécessitant un volume important ^a. Ceci est cohérent avec l'effet inhibant de la protéine sur ce mécanisme.

Mandal *et al* [99] ont mesuré les déclin de fluorescence du chromophore dans des solvants faiblement visqueux et trouvent que le déclin de la fluorescence dépend faiblement de la viscosité. Ils proposent un modèle avec deux états électroniques (état fondamental et état excité) et deux coordonnées pour décrire le déclin de fluorescence (une relaxation vibrationnelle et une coordonnée menant à une intersection conique). Ils en déduisent que la coordonnée qui mène à l'intersection conique correspond probablement à un mouvement nécessitant un faible volume, probablement une rotation concertée autour des deux liaisons du pont. L'effet inhibant de la protéine serait alors dû à une modification de la surface d'énergie potentielle de l'état excité ou à une relaxation vibrationnelle différente à partir de la région Franck Condon.

Les travaux expérimentaux suggèrent donc que le mécanisme de désactivation de la fluorescence implique une torsion simple autour d'une des liaisons du pont reliant les deux cycles aromatiques du chromophore, ou un mouvement de torsion concerté autour de ces deux liaisons.

III.1.3 Etudes quantiques sur l'état excité anionique du chromophore

L'état excité du chromophore de la GFP a été largement étudié par des approches théoriques, employant des méthodes semi-empiriques [103, 104, 105] et des calculs *ab initio* en phase gaz [48, 106, 49, 107] ou en solution [108].

Dans son état électronique fondamental le chromophore est plan et l'indice de liaison entre les atomes C11-C12 est plus élevé que celui entre les atomes C12-C7 (figure I.6 page 20). La torsion τ

^aDans ce cas, le volume désigne la plus petite zone de l'espace pouvant contenir l'ensemble des atomes du chromophore tout au long de sa déformation. Plus le nombre d'atomes impliqués par la déformation est important plus ce volume sera grand. Par exemple, la rotation d'un méthyle nécessite un volume beaucoup plus faible que celle d'un cycle phénolate autour de son axe de symétrie.

est donc plus difficile que la torsion φ . Lorsqu'on optimise la géométrie du premier état électronique excité en partant de la géométrie de l'état électronique fondamental (géométrie Franck-Condon, FC), quelque soit le niveau de calcul utilisé, on obtient une géométrie plane dans laquelle l'ordre des indices des liaisons entre les deux cycles est inversé. L'état électronique excité est de type $\pi\pi^*$ avec un transfert d'électron principalement de la HOMO vers la LUMO [106]. Les orbitales obtenues à partir d'un calcul DFT au niveau B3LYP/6-31G* sont présentées sur la figure I.4. Cet état présente un faible transfert de charge par rapport à l'état fondamental. C'est dans cette géométrie que le chromophore est fluorescent, nous l'appellerons FS ("fluorescent state").

En utilisant des méthodes multiconfigurationnelles de haute précision, Martin *et al* [49], Toniolo *et al* [48], Altoe *et al* [108], Olsen *et al* [107] 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 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 $\tau=0^\circ$, nous la noterons par la suite $TW\varphi$. L'autre correspond à $\tau=\pm 90^\circ$ et $\varphi=0^\circ$, nous l'appellerons $TW\tau$. Ces deux géométries sont présentées sur la figure III.1. D'après les résultats de Martin *et al*, $TW\varphi$ se situe environ $7 \text{ kcal}\cdot\text{mol}^{-1}$ en dessous de la géométrie FS et $TW\tau$ se situe $9 \text{ kcal}\cdot\text{mol}^{-1}$ en dessous de la géométrie FS. Les chemins de réaction entre FS et ces deux géométries présentent de faibles barrières d'activation de l'ordre de $1 \text{ à } 2 \text{ kcal}\cdot\text{mol}^{-1}$. Dans ces deux géométries twistées, il existe un fort transfert de charge entre l'état fondamental et l'état excité. Dans le cas de $TW\varphi$, la charge négative est localisée sur l'imidazolinone dans l'état électronique excité, alors qu'elle est localisée sur le phénolate dans l'état électronique fondamental. Dans le cas de $TW\tau$, la charge négative est localisée sur le phénolate dans l'état excité alors qu'elle est localisée sur l'imidazolinone dans l'état fondamental.

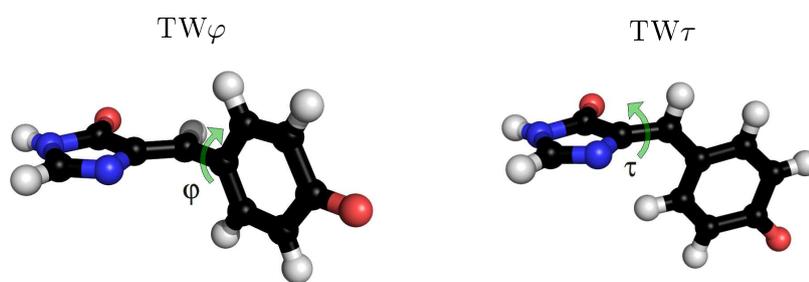


Figure III.1 – Aperçu des géométries $TW\varphi$ (à gauche) et $TW\tau$ (à droite).

Martin *et al* [49] proposent deux voies possibles de désactivation non radiative de l'état électronique excité en phase gaz (figure III.2). D'une part, dans les régions où φ et τ prennent des valeurs comprises entre 60° et 120° l'état fondamental et l'état excité fluorescent sont proches en énergie et une désactivation non radiative par conversion interne est donc envisageable. Par ailleurs, ils ont identifié une intersection conique située $5 \text{ kcal}\cdot\text{mol}^{-1}$ au dessus de la géométrie plane FS qui correspond à une géométrie dans laquelle les torsions autour des liaisons du pont sont très importantes ($\varphi = -105^\circ$ et $\tau = 95^\circ$). Cette intersection conique n'est pas accessible directement à partir de la géométrie FS, mais à partir des géométries $TW\varphi$ ou $TW\tau$ par un chemin que les

auteurs n'ont pas identifié.

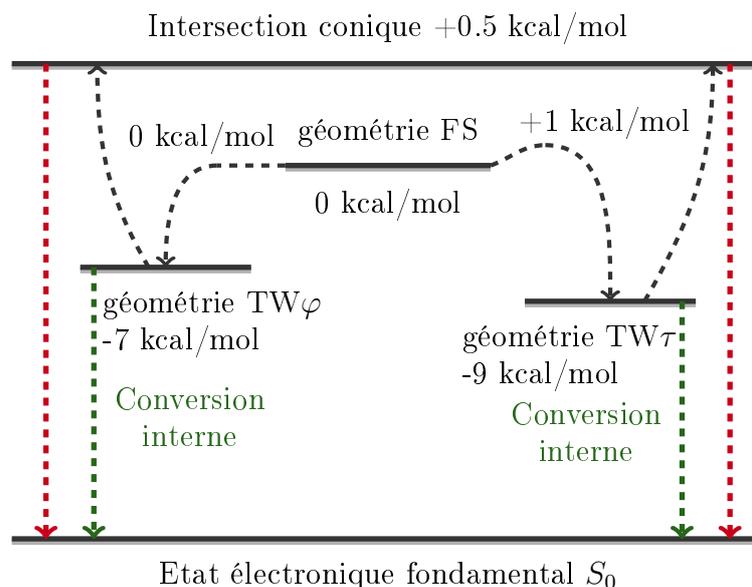


Figure III.2 – Schéma proposé par Martin *et al* pour représenter la surface d'énergie potentielle de l'état excité fluorescent du chromophore de la GFP. Les flèches rouges et vertes en pointillé représentent des transitions électronique non radiatives. Les flèches noires en pointillé représentent les chemins sur la surface d'énergie potentielle de l'état électronique excité envisagés par Martin *et al*.

Toniolo *et al* [48] ont mis en oeuvre une dynamique semi-empirique du chromophore dans l'état excité, dans le vide ou en solution aqueuse. Ils ont montré que le retour à l'état fondamental est plus rapide en phase aqueuse que dans le vide et qu'il est associé à un mouvement concerté autour des deux liaisons reliant les deux cycles aromatiques. Des calculs en présence de solvant ont également été effectués par Altoe *et al* [108] en utilisant un modèle de solvant implicite. Comme dans les études précédentes dans le vide, ils trouvent que dans l'état excité, les géométries TW φ et TW τ sont plus basses en énergie que la géométrie plane FS, de respectivement 5 kcal·mol⁻¹ et 13 kcal·mol⁻¹. Dans ces géométries l'état électronique excité présente le même caractère de transfert de charge que dans le vide. Le chemin conduisant à la géométrie TW φ est sans barrière, mais dans cette géométrie, l'état fondamental et l'état fluorescent sont trop éloignés pour qu'une conversion interne soit possible. A l'inverse, le chemin conduisant à la géométrie TW τ présente une faible barrière et, dans cette géométrie, l'état électronique fondamental et l'état électronique excité sont proches (6 kcal·mol⁻¹) ce qui est compatible avec une désactivation non radiative. Compte tenu de ces résultats, Altoe *et al* proposent que le mécanisme principal de désactivation de la fluorescence pour le chromophore en solution est associé au passage par la géométrie TW τ .

III.1.4 Mouvements du chromophore dans la protéine

Des études qualitatives de dynamique moléculaire de nombreux mutants avec le chromophore à l'état électronique fondamental, montrent que la cavité protéique dans laquelle se trouve le chromophore

est suffisamment grande pour autoriser des mouvements de torsion [109, 110]. Les angles φ et τ peuvent prendre des valeurs allant jusqu'à respectivement 40° et 30° (en valeur absolue). Le calcul du volume nécessaire à l'isomérisation selon φ ou τ montre que l'isomérisation selon φ nécessite un volume plus faible, ce qui explique une amplitude de variation plus importante sur cet angle [111, 112].

Une déformation particulière concernant les angles φ et τ est la torsion concertée de ces angles [112]. Ce mouvement est habituellement nommé Hula-Twist et a été initialement introduit par Liu *et al* [113, 114, 115] lors de l'étude de l'isomérisation du rétinol (le chromophore des protéines de la famille de la rhodopsine) qui intervient dans le mécanisme de la vision. Il s'agit d'un mouvement de torsion concerté autour de deux liaisons consécutives d'un système π [103]. Seul l'atome central sort du plan initial tandis que les deux atomes latéraux sont translatés dans ce plan.

Dans le cas du chromophore de la GFP, ce mouvement peut se produire si les deux cycles ne restent pas coplanaire (mouvement de "papillon"), ou bien, si les deux atomes de carbone CA2 et CG2 se pyramidalisent. Cette dernière déformation s'accompagne d'une forte élévation de l'énergie potentielle et elle n'est donc possible en pratique que pour des déformations de faibles amplitudes. La déformation de type "papillon" est inhibée par la protéine parce qu'elle implique soit un déplacement de l'atome d'oxygène du phénolate soit un basculement de l'atome du pont reliant les deux cycles aromatiques et de l'imidazolinone. Or, le déplacement de l'atome d'oxygène du phénolate est limité car celui-ci est maintenu par trois liaisons hydrogène qui stabilisent la charge négative du phénolate [97] (figure III.3 b). De plus, le cycle imidazolinone est reliée au reste de la protéine par deux liaisons covalentes et une liaison hydrogène et donc son basculement est improbable. Le hula-twist ne semble donc pas envisageable dans la protéine mais seulement en phase liquide. Il pourrait conduire le chromophore dans une région de sa surface d'énergie potentielle où se trouvent des intersections coniques [49]. Cette isomérisation pourrait donc être un mécanisme de désactivation de la fluorescence et être responsable des déclin de fluorescence ultra-rapides observés en solution [48, 99]. Cependant les calculs de chimie quantique publiés et résumés précédemment n'ont pas pu mettre en évidence un chemin d'énergie minimum correspondant à ce type de mouvement.

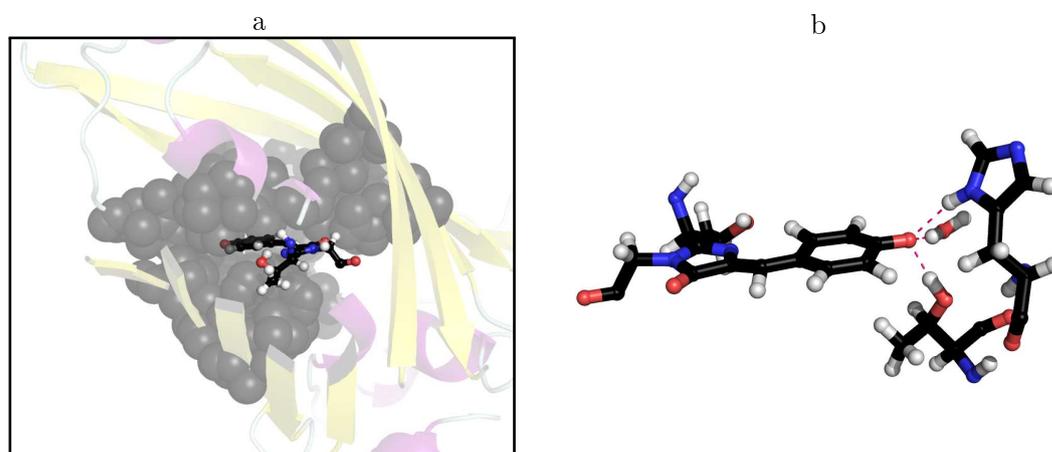


Figure III.3 – Aperçu de l'encombrement stérique du chromophore de la GFP, figure a, et des liaisons hydrogène limitant le mouvement de l'oxygène du phénolate, figure b.

Dans ce travail, nous nous sommes restreints à l'étude du chemin menant à $TW\varphi$, c'est à dire au mouvement de torsion φ . En effet, la torsion τ nécessite un volume plus important que la torsion φ (figure III.1). Le chromophore étant enfermé à l'intérieur de la protéine (figure III.3 a) la gêne stérique occasionnée par la torsion τ sera certainement très importante et la probabilité qu'un tel mouvement ait lieu est beaucoup moins grande. Par ailleurs, la torsion τ nécessite un déplacement important de l'oxygène du phénolate qui est maintenu par trois liaisons hydrogène (figure III.3 b) alors que la torsion φ laisse l'oxygène du phénolate en place.

III.1.5 Processus de désactivation de la fluorescence étudiés dans ce travail

Nous avons choisi d'étudier l'effet de la protéine sur le chemin de réaction menant à la géométrie $TW\varphi$ dans l'état électronique excité. Le long de ce chemin de réaction, seule la torsion φ varie, les coordonnées τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$ restant à leurs valeurs dans la géométrie plane (figure III.1 a). En se basant sur le schéma proposé par Martin *et al* [49] (figure III.2), nous avons considéré deux processus de désactivation de la fluorescence liés à la coordonnées de réaction φ :

- une désactivation de l'état excité fluorescent par conversion interne pour des géométries dans lesquelles l'angle de torsion φ prend des valeurs dans l'intervalle $\mathcal{D}=[60^\circ ; 120^\circ]$.
- une désactivation de l'état excité fluorescent par l'intermédiaire de l'intersection conique identifiée par Martin *et al* qui peut être atteinte à partir de la géométrie $TW\varphi$ dans laquelle φ vaut 90° .

Pour ces deux processus nous avons évalué leur effet sur le déclin de la fluorescence et sur le rendement quantique de fluorescence. Compte tenu du fait que la GFP est une protéine qui présente un rendement quantique de fluorescence élevé, il est évident que les processus de désactivation de la fluorescence en solution sont nettement défavorisés dans la protéine. L'objectif de cette étude est donc de savoir s'ils sont compétitifs avec la désactivation radiative, d'évaluer leur constante de vitesse et de tester la mise en oeuvre de notre méthode.

III.2 Méthodologie

Les différents modèles qui permettent de décrire la cinétique des processus physico-chimiques sont basés sur une séparation des échelles de temps. En effet, l'échelle de temps associée à un processus physico-chimique est généralement beaucoup plus longue que celle associée à la dynamique du système à l'équilibre. Ces processus sont donc qualifiés d'évènements rares. On définit généralement une coordonnée (ou un ensemble de coordonnées) appelée coordonnée de réaction, à laquelle est associé le mouvement permettant d'atteindre l'état final à partir de l'état initial, et qui varie à une échelle de temps longue par rapport à la dynamique du système. Sans cette séparation des échelles de temps la notion de constante de vitesse est inexistante. Il existe deux possibilités qui aboutissent à une séparation des échelles de temps : soit l'état final et l'état initial sont séparés par une barrière d'énergie potentielle supérieure à l'énergie thermique, auquel cas la probabilité de franchir la barrière est faible et le passage de l'état final à l'état initial sera long, soit la barrière

d'énergie potentielle est faible mais le passage de l'état initial à l'état final implique une déformation importante de la géométrie du système.

C'est dans le cadre de cette séparation des échelles de temps qu'ont été développés les différents modèles de calcul de constantes de vitesse [116]. La plupart se placent dans l'hypothèse d'un mouvement Brownien qui permet de décrire la façon dont le système quitte l'état initial pour atteindre l'état final. Dans ce cas, la coordonnée de réaction est traitée explicitement alors que les autres coordonnées sont traitées de manière effective. Par simulation, les constantes de vitesse sont obtenues à partir de données statistiques sur le passage du système au niveau de l'état de transition. Ce type de calcul a été initié par la théorie de Kramers [116] qui cherche à calculer la vitesse avec laquelle une particule s'échappe d'un puit de potentiel.

La durée de vie moyenne de la GFP sauvage est d'environ 3 ns [7] et son rendement quantique de fluorescence est de 0.8 [117]. On peut donc en déduire que le temps de vie de fluorescence (temps de vie de l'état excité si le rendement quantique de fluorescence était égal à 1) est d'environ 3.8 ns (équation I.4 page 13). L'ordre de grandeur des temps caractéristiques des mécanismes de désactivation de la fluorescence étudié ici est donc d'une dizaine de nanosecondes (15 ns). Cette échelle de temps est la même que celle des simulations de dynamique moléculaire tous atomes qui, sur des systèmes de la taille de la GFP, sont actuellement au maximum de l'ordre de quelques dizaines de nanosecondes. Le mécanisme que l'on cherche à étudier est donc un événement rare à l'échelle d'une simulation de dynamique moléculaire et se place dans le cadre décrit plus haut du calcul des constantes de vitesse.

Un échantillonnage des processus associés aux deux mécanismes de la désactivation non radiative de la fluorescence est nécessaire pour calculer une constante de vitesse. Cependant, dans notre cas il ne peut pas être fait directement par simulation moléculaire tous atomes. En effet compte tenu des échelles de temps, les processus cherchés ne sont jamais observés et la statistique obtenue est nulle. Nous avons donc mis en oeuvre une méthode combinant plusieurs types de simulations à diverses échelles de temps qui nous ont permis d'évaluer les constantes de vitesses des deux mécanismes de désactivation de la fluorescence décrit plus haut. Dans ce qui suit, nous allons présenter cette méthode. Elle comporte les trois étapes suivantes :

1. La dynamique du mouvement de torsion φ au sein de la protéine est étudiée au travers du calcul du profil d'énergie libre partielle en fonction de φ par des simulations de dynamique moléculaire biaisées.
2. une dynamique Brownienne de la coordonnée φ utilisant le profil d'énergie libre obtenu au 1) nous permettra d'obtenir un très grand nombre (1 million) de trajectoires de la coordonnée φ sur un intervalle de temps de plusieurs dizaine de nanosecondes (50 ns).
3. les résultats de la dynamique Brownienne seront utilisés dans un modèle cinétique permettant de calculer le déclin de la population de l'état excité et le rendement quantique.

III.3 Résultats de dynamique moléculaire

Afin de valider le champ de force paramétrisé pour le chromophore anionique de la GFP dans son état électronique excité (paragraphe I.3.4 page 33), nous l'avons tout d'abord étudié en phase aqueuse. Ces simulations nous ont également permis de tester et de mettre au point les simulations de dynamique moléculaire biaisées utilisées pour calculer le profil d'énergie libre partielle par la méthode du *umbrella sampling*. Nous avons ensuite calculé le profil d'énergie libre partielle en fonction de la coordonnée φ dans la protéine. L'analyse des trajectoires obtenues au cours de cette étude, nous a permis d'analyser l'effet de la protéine sur la dynamique du chromophore dans son état électronique excité et d'identifier les acides aminés qui jouent un rôle majeur.

III.3.1 Simulations du chromophore de la GFP en phase aqueuse

En phase aqueuse, nous avons simulé l'entité chromophore telle qu'elle a été définie lors de la paramétrisation du champ de force (paragraphe I.3.1 page 20) dans une boîte d'eau à une température de 300 K et une pression de 1 atm. Nous avons mis en oeuvre différents types de simulations : une simulation à l'équilibre de 1 ns pour caractériser la structure et la dynamique du chromophore en solution, des simulations hors équilibre de 10 ps pour caractériser le retour à la géométrie d'équilibre et des simulations biaisées pour calculer le profil d'énergie libre.

a) Simulation en phase aqueuse à l'équilibre

Nous avons réalisé une simulation de dynamique moléculaire d'une nanoseconde du chromophore anionique de la GFP dans son premier état électronique excité, en phase aqueuse. A partir de la géométrie plane du chromophore dans la structure RX de la GFP, nous avons minimisé l'énergie de la boîte de simulation puis le système a été mis à l'équilibre en deux étapes. Tout d'abord, une simulation dans l'ensemble (N,V,T) pour équilibrer la température, suivie d'une simulation (N,P,T) pour équilibrer la pression. L'utilisation de contraintes, tel que cela a été fait dans le cas de la Cerulean ne présente pas d'intérêt dans ce cas. Le substrat étant une simple molécule des déformations artificielles de sa structure ne sont pas à craindre. A la suite de ces différentes étapes de préparation du système, nous avons fait une simulation de dynamique moléculaire dans l'ensemble (N,P,T) à une température de 300 K et une pression de 1 atm.

La figure III.4 présente l'histogramme des valeurs de φ obtenues au cours de cette simulation ainsi que les valeurs moyennes et les écarts types des coordonnées internes reliées aux déformations du chromophore impliquant le pont qui relie les deux cycles aromatiques. Les valeurs d'équilibre obtenues au cours de cette simulation montrent que la géométrie d'équilibre en phase aqueuse est la géométrie TW φ . Ceci correspond bien à une des géométries de l'état électronique excité du chromophore de plus basse énergie prédites par les calculs quantiques en phase aqueuse [108]. D'autre part ces résultats sont cohérent avec le champ de force que nous avons utilisé. Les valeurs des coordonnées τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$ restent proches des valeurs imposées dans le champ de force (tableau I.5 page 34). On observe donc bien la dynamique de la torsion φ correspondant à la rotation du phénolate autour de son axe CG2-CZ, le reste du chromophore restant plan.

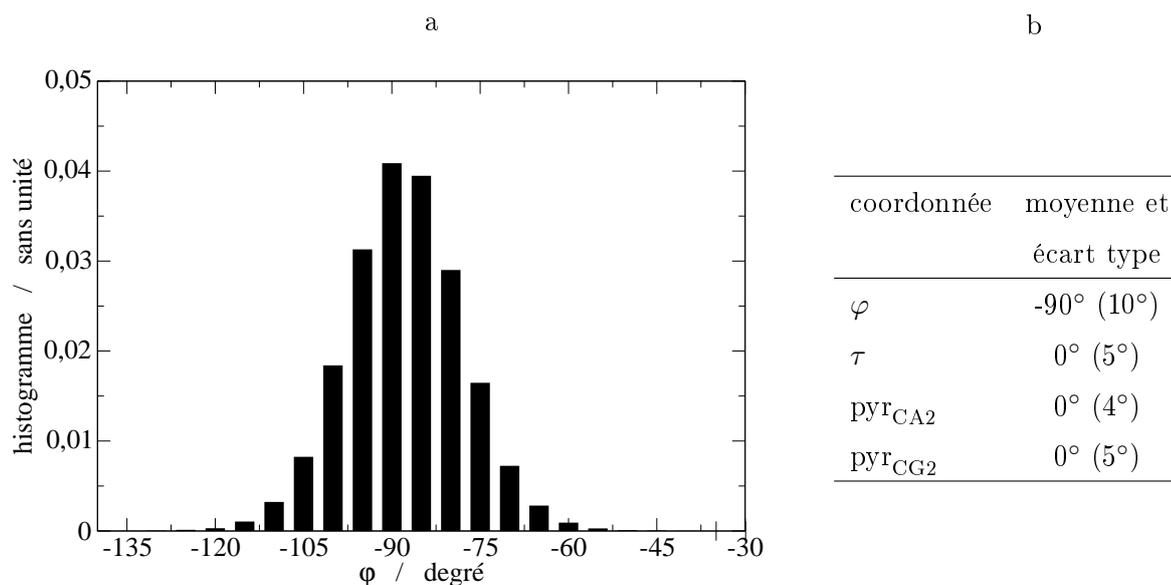


Figure III.4 – a) Histogramme des valeurs de φ obtenues au cours de la simulation de une nanoseconde. b) Valeurs moyennes et écart type des coordonnées φ , τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$ qui concerne le pont reliant les deux cycles aromatiques.

b) Simulation en phase aqueuse hors équilibre

Partant d'une géométrie plane, on obtient donc dans nos simulations une géométrie twistée, $\text{TW}\varphi$. Afin d'évaluer le temps caractéristique associé à cette déformation du chromophore nous avons réalisé un ensemble de trajectoires hors équilibre à partir de configurations du chromophore dans une géométrie plane. Pour obtenir ces configurations, nous avons fait une simulation d'une nanoseconde en imposant une contrainte harmonique sur la coordonnée φ avec une constante de force de $10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ et centrée sur $\varphi=0^\circ$. De cette simulation nous avons extrait un échantillon de 50 configurations dans lesquelles le chromophore est plan. La moyenne des valeurs de φ sur cet échantillon est 2° et l'écart type est 7° . Cet échantillon est représentatif de la région Franck Condon du chromophore. En effet, dans l'état fondamental la géométrie plane du chromophore correspond au minimum d'énergie potentielle [48,49]. De plus dans les simulations de dynamique moléculaires de la GFP avec le chromophore dans son état électronique fondamental la moyenne et l'écart type de la coordonnée φ sont similaires à ce que nous avons pour cet échantillon de configurations ($\langle \varphi \rangle = 2.1^\circ \pm 7.7^\circ$ [24] et $\langle \varphi \rangle = -2.2^\circ \pm 7.7^\circ$ [75]).

Nous avons ensuite réaliser 1500 trajectoires de 10 ps chacune, sans contraintes sur la torsion φ , en faisant varier les coordonnées et les vitesses initiales. La figure III.5 présente la trajectoire moyenne, notée $\langle \varphi \rangle(t)$ et l'écart type sur les trajectoires, noté $\sigma_\varphi(t)$, de la coordonnée φ sur l'ensemble des trajectoires obtenues. L'écart type présente un maximum aux alentours de $t=1.3 \text{ ps}$. A $t=10 \text{ ps}$, l'ensemble des trajectoires ont atteint une situation d'équilibre qui est celle décrite au paragraphe précédent et correspond à la géométrie $\text{TW}\varphi$. Les valeurs de $\langle \varphi \rangle$ et de σ_φ à $t=10 \text{ ps}$ sont similaires à celles obtenues au cours de la simulation à l'équilibre (figure III.4 b). Compte tenue de la symétrie du cycle phénolate, nous avons considéré la valeur absolue de φ . En partant de la géométrie plane 30% des trajectoires se stabilisent autour de la valeur $\varphi=90^\circ$ et 70% se stabilisent autour de la valeur $\varphi=-90^\circ$.

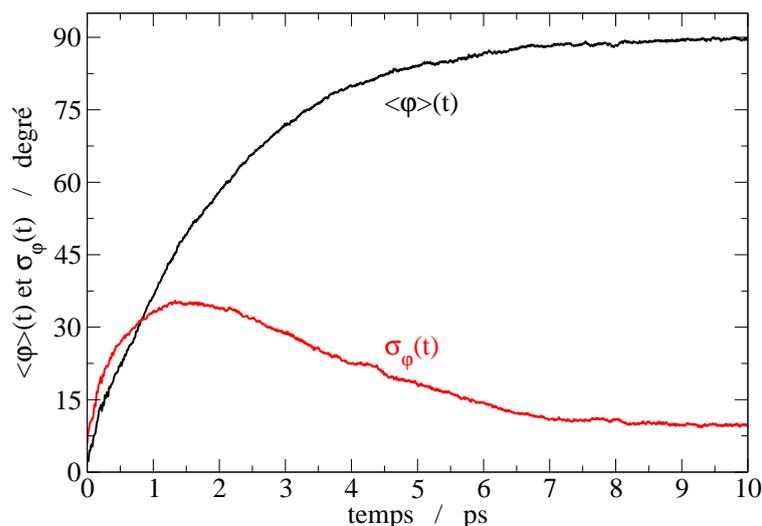


Figure III.5 – Trajectoire moyenne, $\langle \varphi \rangle$ en noir, et trajectoire de l'écart type, σ_φ en rouge, de la valeur absolue de la torsion φ en fonction du temps sur les 1500 trajectoires hors équilibre.

Le passage de la géométrie plane à la géométrie TW_φ se fait de manière soudaine. La variation progressive de la courbe $\langle \varphi \rangle(t)$ est dû au fait que les trajectoires de la coordonnée φ oscillent un certain temps au voisinage de la géométrie plane et que ce temps dépend fortement de la trajectoire. Ce phénomène est également à l'origine du maximum de la courbe $\sigma_\varphi(t)$ car, à temps court, deux types de trajectoires coexistent : celles qui ont déjà atteint des valeurs proches de $\varphi=90^\circ$ et celles qui oscillent toujours autour de la géométrie plane.

Cet instant, pour lequel les trajectoires atteignent la géométrie TW_φ , est intéressant pour notre étude car il peut être relié au déclin de la fluorescence. Expérimentalement, le déclin de fluorescence est mesuré en collectant le nombre de photons reçus en fonction du temps, à partir d'un instant $t=0$ où un ensemble de molécules est porté à l'état excité. Cet instant $t=0$ est généralement l'instant de déclenchement de la source laser. Etant donné que nous avons utilisé comme point de départ de nos simulations un échantillon de géométries planes, représentatif de la région Franck Condon, le temps $t=0$ des simulations peut être associé à l'instant où un ensemble de chromophores est porté à l'état excité. Chaque trajectoire représente alors l'évolution d'un chromophore. Or, pour le chromophore de la GFP en solution aqueuse, la géométrie plane est fluorescente alors que la géométrie TW_φ est non fluorescente. En supposant que si le chromophore atteint la géométrie TW_φ il retourne sans émission de photon à l'état fondamental, le temps moyen à partir duquel le chromophore atteint la géométrie TW_φ est donc une approximation grossière de la durée de vie de l'état excité [116, p 290]. Pour calculer le temps moyen pour lequel le chromophore atteint la géométrie TW_φ nous avons calculé la distribution des temps de premier passage $p_1(t)$ sur l'ensemble des trajectoires. Cette distribution est obtenue en faisant l'histogramme, sur l'ensemble des trajectoires, des temps pour lesquels la trajectoire passe pour la première fois par la valeur 90° . Elle est présentée sur la figure III.6.

La distribution des temps de premier passage est nulle à $t=0$, puis elle croît rapidement. Elle présente un maximum aux alentours de 1.3 ps puis elle décroît et s'annule aux alentours de 8 ps. La

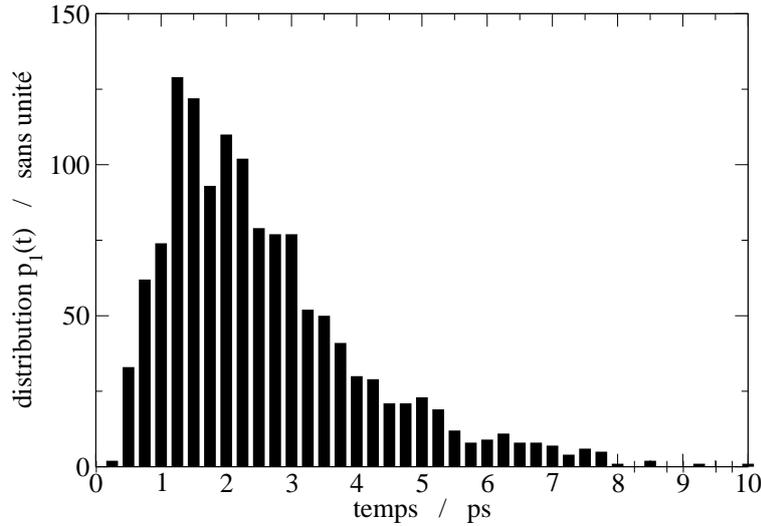


Figure III.6 – Distribution des temps de premier passage à la géométrie $TW\varphi$ sur l'ensemble des trajectoires.

valeur moyenne est de 2.5 ps. L'obtention d'une telle distribution nécessite un nombre important de trajectoires, notamment pour les temps longs. C'est pour cette raison que nous avons eu besoin de faire un grand nombre de trajectoires. Le fait que le maximum de cette distribution ne soit pas en zéro est dû au fait qu'il existe un temps minimum pour que le système atteigne la géométrie $TW\varphi$ à partir de la géométrie plane. L'ordre de grandeur du maximum et de la valeur moyenne est de quelques picosecondes et est donc en bon accord avec les temps de vie expérimentaux du chromophore en solution [99].

En résumé, le champ de force que nous avons mis au point pour l'état électronique excité du chromophore de la GFP conduit à une géométrie d'équilibre en phase aqueuse similaire à la géométrie $TW\varphi$. Le temps nécessaire au système pour atteindre cette géométrie est du même ordre de grandeur que les mesures expérimentales du temps de vie, du chromophore en solution aqueuse. Le champ de force utilisé nous permet donc de représenter correctement la dynamique du chromophore.

c) Calcul de profil d'énergie libre partielle

Nous avons calculé le profil d'énergie libre partielle en fonction de φ par la méthode du *Umbrella Sampling*. Nous avons utilisé 23 simulations biaisées du chromophore de la GFP dans son état électronique excité en phase aqueuse d'une durée de 1 ns chacune (après 100 ps de mise à l'équilibre). Le potentiel de biais utilisé, V^b , comporte deux termes quadratiques :

$$V^b = k(\varphi_1 - \varphi_0)^2 + k(\varphi_2 - \varphi_0 - 180^\circ)^2 \quad (\text{III.1})$$

où φ_1 et φ_2 sont associés aux angles dièdres CA2-CB2-CG2-CD1 et CA2-CB2-CG2-CD2 (figure I.6 page 20) qui permettent tous deux de mesurer l'angle de torsion φ . La constante de force k vaut $10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$. φ_0 prend une valeur différente dans chaque simulation biaisée et définit la zone de l'espace des configurations que l'on souhaite échantillonner. Les valeurs de φ_0 que nous

avons utilisées sont comprises entre -200° et 20° et espacées de 10° ce qui nous donne 23 valeurs correspondant aux 23 simulations biaisées effectuées. Le fait d'ajouter un potentiel de biais sur deux angles dièdres décrivant la torsion φ permet d'imposer un couple de force symétrique sur le phénolate et évite ainsi des déformations non physique de ce cycle.

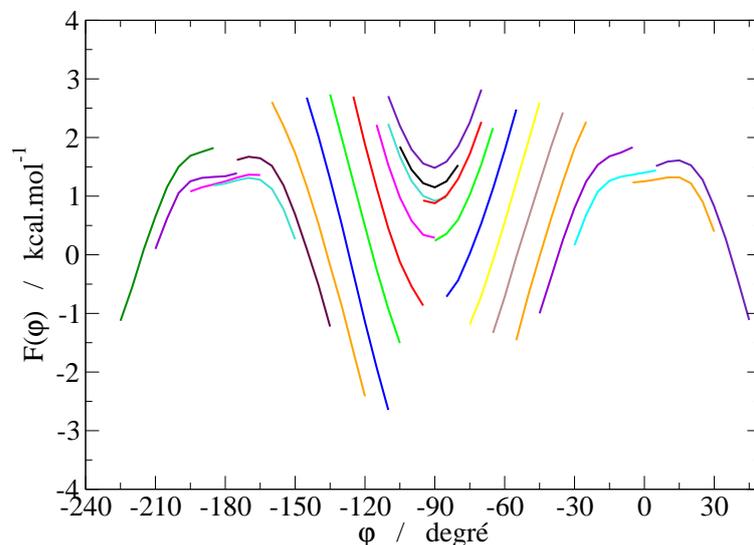


Figure III.7 – Courbes $\mathcal{F}_i(\varphi)$ d'énergie libre partielle de la coordonnée φ obtenues pour chaque simulation biaisée. Pour chaque simulation biaisée, seules les valeurs de φ pour lesquelles l'histogramme des valeurs de φ est supérieur à 5% de son maximum ont été reportées sur la figure.

La figure III.7, présente les fragments d'énergie libre partielle obtenus à partir de chacune des simulations biaisées, noté $\mathcal{F}_i(\varphi)$, et de la simulation sans contrainte du chromophore en phase aqueuse. Chacune des courbes $\mathcal{F}_i(\varphi)$ est obtenues à partir de l'équation I.19 en utilisant l'histogramme des valeurs de φ dans chaque simulation biaisée et est définie à une constante près.

On remarque que les courbes $\mathcal{F}_i(\varphi)$ obtenues se recouvrent et sont parallèles deux à deux dans une gamme de valeurs communes de φ . Pour deux simulations biaisées consécutives on a donc bien échantillonné une zone commune de l'espace des configurations. Un écart de 10° entre deux valeurs de φ_0 consécutives avec une constante de force de $10 \text{ kcal} \cdot \text{mol}^{-1}$ nous permet donc de raccorder les différentes courbes $\mathcal{F}_i(\varphi)$ et obtenir le profil d'énergie libre partielle complet qui est présenté sur la figure III.8.

Le profil d'énergie libre obtenu présente un minimum pour $\varphi = -90^\circ$ et deux maxima en $\varphi = 0^\circ$ et -180° . De plus, elle est légèrement dissymétrique. Les deux maxima sont situés $7.5 \text{ kcal} \cdot \text{mol}^{-1}$ et $7.8 \text{ kcal} \cdot \text{mol}^{-1}$ au dessus de la géométrie de plus basse énergie. Le minimum d'énergie libre correspond bien à la géométrie twistée $\text{TW}\varphi$ alors que les deux maxima représentent la géométrie plane ce qui est bien en accord avec les observations faites jusqu'à présent pour ces géométries dans l'état électronique excité du chromophore de la GFP. Au paragraphe I.4, différentes façons permettant de raccorder entre elles les courbes d'énergie libre ont été présentées. Dans le cas du chromophore en solution aqueuse l'utilisation de l'une ou l'autre des méthodes présentées conduit aux mêmes résultats. On observe qu'au voisinage de la géométrie plane le profil d'énergie libre

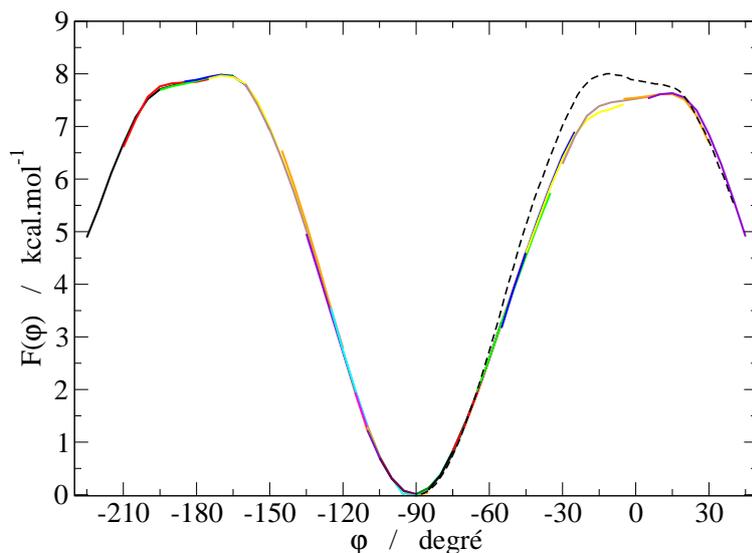


Figure III.8 – Profil d'énergie libre partielle en fonction de la coordonnée φ en phase aqueuse. La courbe noire en pointillé est obtenue par symétrie par rapport à l'axe passant par $\varphi = -90^\circ$.

présente une pente faible conduisant préférentiellement aux valeurs plus négatives de φ . Cette pente justifie le fait que la majorité des trajectoires hors équilibre aient atteint une géométrie d'équilibre proche de $\varphi = -90^\circ$. Elle est probablement due au fait que la rotation du phénol autour de son axe n'est pas équivalente suivant le sens de rotation. Cette dissymétrie, peut provenir d'une part, du fait que le chromophore n'est pas symétrique et, d'autre part, du fait que les charges effectives sur le phénol ne sont pas exactement symétriques par rapport à l'axe du phénolate.

Le profil d'énergie libre présente également un faible décalage entre 0° et -180° (voir la courbe en pointillés sur la figure III.8). Ce décalage est probablement la conséquence de deux effets. Tout d'abord, les courbes $\mathcal{F}_i(\varphi)$ de la figure III.7 ne sont pas parfaitement parallèles deux à deux. Il en résulte une incertitude sur les constantes utilisées pour positionner les différentes courbes les unes par rapport aux autres qui s'ajoute au fur et à mesure que l'on s'éloigne de la géométrie d'équilibre ($\varphi = -90^\circ$). D'autre part, nous avons fait l'hypothèse que les charges effectives sont indépendantes de la géométrie du chromophore et nous avons utilisé celles obtenues pour $\varphi = 0^\circ$ pour toutes les valeurs de φ . Il en résulte que ces charges ne sont pas adaptées pour la géométrie $\varphi = 180^\circ$ car le chromophore n'est pas parfaitement symétrique par rapport à l'axe du phénolate. Cependant, cette différence d'énergie libre n'est que de $0.3 \text{ kcal} \cdot \text{mol}^{-1}$ et est inférieure à la précision que l'on peut attendre de ce type de calculs.

III.3.2 Simulation du chromophore de la GFP dans la protéine sans contrainte

La dynamique du chromophore dans le premier état électronique excité et à l'intérieur de la protéine a été étudiée en réalisant une dynamique moléculaire de la GFP avec le champ de force du chromophore présenté au paragraphe I.3.4. Cette simulation est décrite et analysée ci-dessous.

III.3.2.1 Coordonnées de départ

La structure RX de la GFP dans son état excité n'est pas disponible. Bien qu'il soit possible de réaliser des expériences de diffraction de rayons X résolues en temps elles sont à l'heure actuelle encore difficile à mettre en oeuvre [6, Chapitre 13]. Cependant, comme le temps caractéristique de la fluorescence de la GFP n'est que d'environ 3 ns, il est raisonnable de penser que la structure de la protéine est peu modifiée lorsqu'on passe à l'état excité. Comme nous l'avons vu au paragraphe III.1.1 (page 106), le chromophore de la GFP existe sous deux formes : une forme neutre et une forme anionique [97]. Pour simplifier notre étude, nous avons étudié le mutant S65T de la GFP dans lequel la forme anionique du chromophore est fortement stabilisée et nous avons négligé l'existence d'une forme neutre. Nous avons donc utilisé la structure RX de ce mutant (entrée 1EMA [18] de la *Protein Data Bank*) comme point de départ de nos simulations.

Nous avons choisi de placer l'ensemble des acides aminés acido-basiques dans leur état standard de protonation à pH neutre ^a sauf le chromophore qui est sous sa forme basique et l'acide aminé Glu222 sous sa forme acide. Glu222 est généralement considéré comme étant l'accepteur final de la chaîne de proton initiée par la perte du proton du phénol du chromophore. La position δ ou ϵ du proton sur les histidines a été choisie de la même façon que ce qui avait déjà été fait dans la référence [24].

III.3.2.2 Dynamique et environnement du chromophore

A partir des coordonnées initiales présentées au paragraphe précédent, nous avons minimisé l'énergie potentielle du système et nous avons mis à l'équilibre la température et la pression du système. Puis nous avons réalisé une simulation de la GFP avec le chromophore dans son état électronique excité de 6 ns dans l'ensemble (N,P,T) à une température de 300 K et une pression de 1 atm.

La figure III.9 présente l'évolution en fonction du temps du RMSd pour trois groupes d'atomes. La partie de la trajectoire correspondante à la mise à l'équilibre de la température et de la pression n'est pas présentée. Les courbes du RMSd obtenues pour l'ensemble des C_α constituant les feuillets β et l'hélice α , pour l'ensemble des C_α de la protéine et pour l'ensemble des atomes de la protéine sont constantes tout au long de la simulation et oscillent respectivement autour de 0.8 Å, 1.1 Å et 1.7 Å. Ces valeurs sont similaires à celles obtenues lors des simulations de la Cerulean (paragraphe II.2.3.2 et II.2.5.2) ou dans d'autres études réalisées par simulation de dynamique moléculaire de la GFP dans l'état électronique fondamental [24, 75]. La structure est donc stable tout au long de la simulation.

La figure III.10 présente le réseau de liaisons hydrogène calculé sur les 6 ns de la simulation de la GFP avec le chromophore dans son état électronique excité. On observe que l'oxygène du phénolate est impliqué dans 3 liaisons hydrogène stables avec les acides aminés Thr203 et His148 et avec une molécule d'eau. Le réseau de liaison hydrogène est peu différent de celui obtenu dans d'autres études pour le chromophore anionique de la GFP à l'état fondamental [75, 118, 119] et des autres mutants de

^aA pH neutre les acides aspartiques, acides glutamiques et les histidines sont sous leur forme basique tandis que les lysines, les arginines et les tyrosines sont sous leur forme acide (annexe A.1).

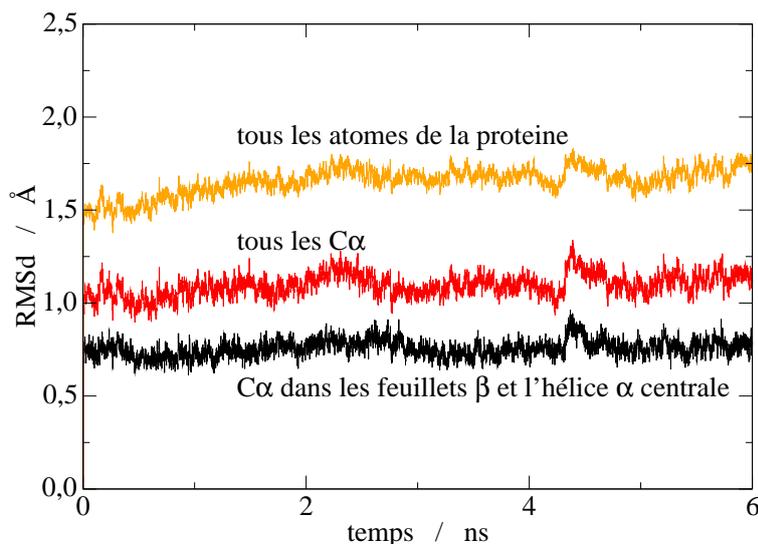


Figure III.9 – Déplacement carré moyen en fonction du temps des C_{α} des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β en noir, de l'ensemble des C_{α} en rouge et de l'ensemble des atomes de la protéine en orange.

la GFP. Notamment, autour du cycle imidazolinone, on retrouve les liaisons hydrogène habituelles avec les acides aminés Arg96, Gln94 et Glu222. La faible différence de réseau de liaisons hydrogène autour du chromophore entre l'état électronique excité et l'état électronique fondamental est en partie due au fait que les charges des atomes du chromophore qui forment des liaisons hydrogène, sont peu modifiées lorsque le chromophore passe de son état électronique fondamental à l'état électronique excité (tableau I.1 page 25).

coordonnée	φ	τ	$\text{pyr}_{\text{CA}2}$	$\text{pyr}_{\text{CG}2}$
valeur moyenne	2°	1°	1°	0°
écart type	10°	5°	4°	4°

Tableau III.1 – Valeur moyenne et écart type des coordonnées φ , τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$ qui concerne le pont reliant les deux cycles aromatiques du chromophore.

Le tableau III.1 présente les valeurs moyennes et les écarts types des coordonnées impliquant les liaisons du pont reliant les deux cycles aromatiques du chromophore. Les valeurs obtenues pour les coordonnées τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$ sont similaires à celles obtenues pour la simulation du chromophore dans son état électronique excité en phase aqueuse (figure III.4 b page 114). Cependant, contrairement à la simulation en phase aqueuse où la géométrie d'équilibre correspondait à la géométrie twistée $\text{TW}\varphi$ ($\langle \varphi \rangle_{\text{aq}} = 90^{\circ}$), on remarque que tout au long de la simulation dans la protéine, la géométrie du chromophore est plane, $\langle \varphi \rangle_{\text{protéine}} = 2^{\circ}$. Bien que le minimum de l'énergie potentielle du chromophore dans son état électronique excité corresponde à la géométrie twistée $\text{TW}\varphi$, les contraintes stériques dues à la protéine imposent donc au chromophore de conserver une géométrie plane. Cette observation est en accord avec le fait que dans la protéine le chromophore est

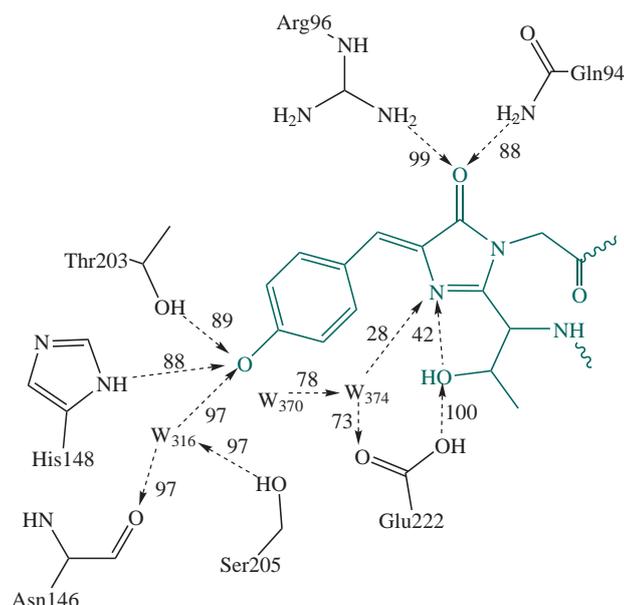


Figure III.10 – Réseau de liaisons hydrogène dans l’environnement du chromophore obtenu à partir de la simulation. Les flèches indiquent la présence d’une liaison hydrogène et sont orientées du donneur vers l’accepteur d’hydrogène. Les nombres associés à chaque liaison hydrogène indiquent en pourcentage le temps pendant lequel la liaison hydrogène existe.

fluorescent. Par ailleurs, les fluctuations de φ sont sensiblement plus grandes dans notre simulation de la GFP avec le chromophore dans l’état électronique excité que celles des simulations de la GFP avec le chromophore à l’état fondamental qui étaient de l’ordre de 7° [24, 75]. Ceci est probablement dû au fait que les courbures de la surface d’énergie potentielle du chromophore au voisinage de la géométrie plane sont différentes dans l’état électronique fondamental et l’état électronique excité.

La figure III.11 présente la trajectoire de la coordonnée φ sur les 6 ns de la simulation et l’histogramme des valeurs de φ sur cette trajectoire. On remarque que l’histogramme est centré sur 0° et que la trajectoire atteint rarement des valeurs supérieures à 30° en valeurs absolues. On observe également que la trajectoire est légèrement dissymétrique par rapport à la droite $\varphi=0^\circ$. Les configurations aux voisinages de la géométrie TW φ ou dans une zone où φ prend des valeurs comprises entre 60° et 120° ne sont jamais ou que rarement échantillonnées. Les mouvements de torsion φ de grande amplitude sont des événements rares à l’échelle de notre simulation. Une méthode spécifique doit donc être mise en oeuvre afin de pouvoir l’étudier. Dans cette optique, nous avons mis en oeuvre des simulations de dynamique moléculaire biaisées permettant d’échantillonner des zones de l’espace des configurations dans lesquelles φ prend des valeurs proches de 90° et dans l’intervalle $\mathcal{D}=[60^\circ ; 120^\circ]$. Ces valeurs correspondent aux situations favorables à une désactivation non radiative de l’état excité.

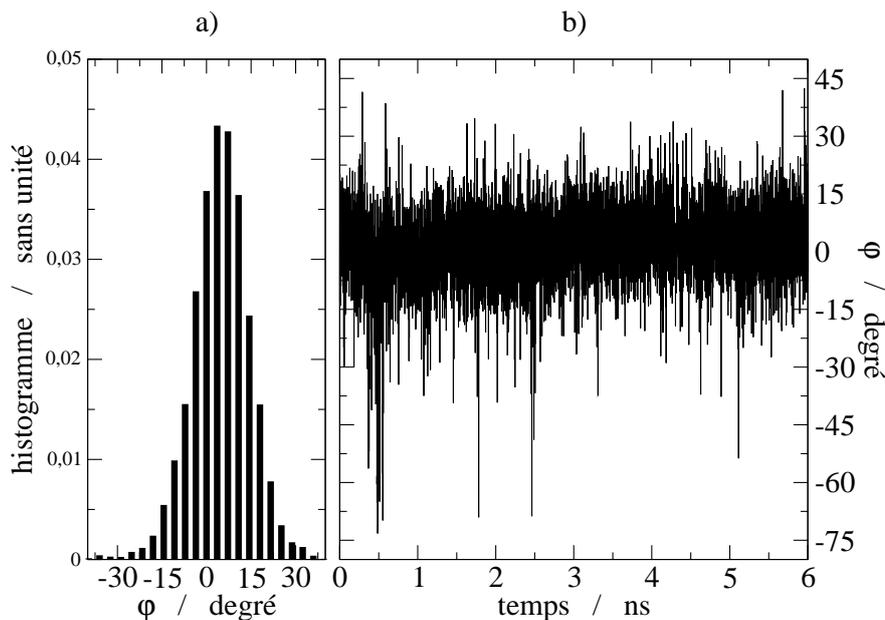


Figure III.11 – a) Histogrammes des valeurs de φ au cours de la simulation. b) Trajectoire de φ sur les 6 ns de la simulation de la GFP avec le chromophore dans son état excité fluorescent.

III.3.3 Profil d'énergie libre partielle en fonction de φ

Le calcul du profil d'énergie libre partielle en fonction de φ est plus complexe dans la protéine qu'en phase aqueuse en raison de la présence d'un environnement du chromophore fortement inhomogène et des mouvements contraints des chaînes latérales des acides aminés à l'intérieur de la protéine. Dans ce paragraphe, sera détaillé le calcul du profil d'énergie libre partielle. Les difficultés liées à ce type de calcul dans les systèmes complexes comme les protéines seront mis en avant.

III.3.3.1 Simulations moléculaires biaisées

Le potentiel de biais utilisé dans les simulations biaisées est le même que celui que nous avons utilisé en phase aqueuse (équation III.1 page 116). Sur l'ensemble des simulations biaisées, ou fenêtres, que nous avons réalisées, nous en avons retenu 15 qui nous ont permis d'échantillonner l'ensemble du domaine de φ allant de -90° à 90° . Le tableau III.2 résume les paramètres du potentiel de biais utilisés (la valeur de la constante de force, k , et la position du minimum du potentiel φ_0), ainsi que la durée de chaque simulation biaisée. Les histogrammes des valeurs de φ obtenus à partir de chacune des simulations biaisées, appelés histogrammes biaisés, sont présentés sur la figure III.12.

Nous avons réalisé les simulations biaisées consécutivement, en choisissant judicieusement une configuration dans la fenêtre réalisée avec une valeur de φ_0 inférieure (en valeur absolue) comme coordonnées de départ pour réaliser une nouvelle fenêtre avec une valeur de φ_0 plus importante. Ce point est important car il garantit que la configuration de la fenêtre i utilisée comme point de départ pour la fenêtre $i + 1$ avec le nouveau potentiel de biais n'est pas une configuration aberrante compte tenu de ce nouveau potentiel. Cela permet d'éviter l'obtention de trajectoires non physiques. En

i	k (kcal · mol ⁻¹ · rad ⁻²)	φ_0	durée	i	k (kcal · mol ⁻¹ · rad ⁻²)	φ_0	durée
1	10.0	-120°	2 ns	10	10.0	20°	500 ps
2	10.0	-110°	2 ns	11	10.0	40°	500 ps
3	10.0	-100°	2 ns	12	10.0	50°	500 ps
4	10.0	-80°	1 ns	13	10.0	60°	1 ns
5	10.0	-70°	2 ns	14	10.0	65°	1 ns
6	20.0	-50°	2 ns	15	10.0	80°	1 ns
7	10.0	-40°	2 ns				
8	20.0	-30°	1 ns				
9	20.0	-20°	1 ns				

Tableau III.2 – Constante de force, k en kcal · mol⁻¹ · rad⁻², et position du minimum, φ_0 du potentiel de biais employé dans les simulations biaisées utilisées pour construire le profil d'énergie libre en fonction de φ dans la GFP. La durée correspond au temps exploité dans chaque simulation.

pratique on utilise le fait que les histogrammes de φ pour deux fenêtres consécutives se recouvrent et on choisit une configuration de départ pour la fenêtre $i+1$ dans cette zone de recouvrement.

Nous avons choisi les valeurs de la constante de force, k , et de la valeur φ_0 du potentiel de biais avec lesquelles la simulation est effectuée de sorte que, d'une part, l'histogramme biaisé de φ ne soit pas centré sur une valeur trop éloignée de la valeur φ_0 du potentiel de biais et, d'autre part, pour que l'échantillonnage des configurations dans deux fenêtres consécutives recouvre une zone commune de l'espace des phases. Les durées des simulations ont également été adaptées pour remplir au mieux ces deux conditions. En pratique, elles sont difficiles à obtenir, la description des problèmes rencontrés fera l'objet des paragraphes suivants.

Les histogrammes biaisés de la coordonnée φ recouvrent bien l'ensemble du domaine de variation de φ (figure III.12). De plus, on observe que les histogrammes de deux fenêtres consécutives se recouvrent deux à deux et que les histogrammes biaisés obtenus pour $\varphi_0 = \pm 20^\circ$ recouvrent l'histogramme de φ obtenu à partir de la simulation sans contraintes. Ces simulations peuvent donc être utilisées pour construire le profil d'énergie libre.

III.3.3.2 Déformations du chromophore et de la protéine

Pour chaque simulation biaisée, nous avons calculé les moyennes et les écarts types des coordonnées φ , τ , pyr_{CA2} et pyr_{CG2} . Leurs variations en fonction du minimum du potentiel de biais, φ_0 , sont présentées sur la figure III.13. Les points pour $\varphi_0=0^\circ$ correspondent aux valeurs obtenues pour la simulation sans contrainte.

On remarque que la moyenne des coordonnées pyr_{CA2} et pyr_{CG2} reste proche de zéro quelque soit la simulation biaisée. Les valeurs de la moyenne de τ sont plus sensibles au potentiel de biais employé mais elles restent toujours inférieure à 10° en valeur absolue. La moyenne de φ dans chaque fenêtre

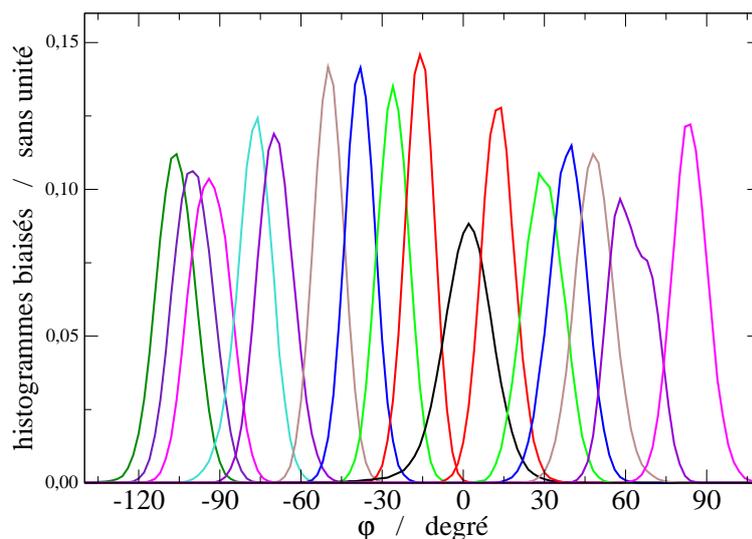


Figure III.12 – Histogrammes des valeurs de φ obtenues au cours des 15 simulations biaisées et de la simulation sans contraintes (en noir).

suit relativement bien la droite $\langle \varphi \rangle = \varphi_0$. Grâce au potentiel de biais employé et aux paramètres du champ de force utilisés pour les coordonnées τ , $\text{pyr}_{\text{CA}2}$ et $\text{pyr}_{\text{CG}2}$, nous avons donc bien imposé une rotation du cycle phénolate autour de son axe CG2-CZ conduisant à une géométrie proche de TW_φ (figure III.1 page 108).

Dans chacune des simulations biaisées, nous avons également mesuré la variation du déplacement carré moyen par rapport à la structure RX de la GFP. La figure III.14 présente la variation du RMSd en fonction de la moyenne des valeurs de φ pour l'ensemble des simulations biaisées que nous avons effectué.

On remarque que le RMSd de l'ensemble des $\text{C}\alpha$ de la protéine et des $\text{C}\alpha$ des acides aminés formant des feuilletts β ou l'hélice α centrale dépend peu de la fenêtre que l'on considère. De plus, les valeurs du RMSd sont peu différentes de celles obtenues au cours de la simulation de la GFP sans contrainte. Le RMSd des atomes des acides aminés proches du chromophore et formant la cage dans laquelle est confinée le chromophore présente une variation plus significative en fonction de la fenêtre de simulation et présente un minimum pour la simulation sans contrainte. De plus, on remarque que les écarts types sont nettement plus faibles pour les valeurs de φ_0 les plus faibles (en valeur absolue). La structure de la protéine ne subit donc pas de déformation importante sous l'effet de la rotation du cycle phénolate. On peut cependant noter des déformations plus importante de la cage entourant le chromophore pour des fortes valeurs de la torsion φ qui traduisent le fait que lorsque le cycle phénolate tourne sur son axe, il heurte certains acides aminés qui doivent se déformer pour que le cycle puisse poursuivre sa rotation.

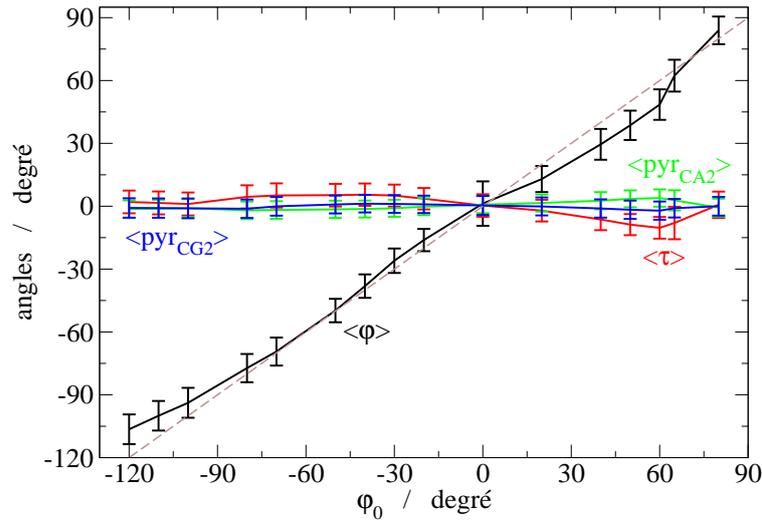


Figure III.13 – Variation de la valeur moyenne des coordonnées φ (en noir), τ (en rouge), $\text{pyr}_{\text{CA}2}$ (en vert) et $\text{pyr}_{\text{CG}2}$ (en bleu) en fonction de la valeur du paramètre φ_0 du potentiel de biais. Les points pour $\varphi_0=0^\circ$ correspondent aux valeurs obtenues pour la simulation sans contraintes. Les barres d’erreur représente les écarts type. Le trait pointillé représente la droite $\langle \varphi \rangle = \varphi_0$.

III.3.3.3 Identification des acides aminés gênant la rotation du phénolate

Afin d’identifier les acides aminés qui gênent la rotation du phénolate autour de son axe à l’intérieur de la protéine, nous avons calculé le couple de force exercé individuellement par chaque acide aminé et par le chromophore lui même sur le cycle phénolate. Le couple de force Γ s’exerçant sur le cycle phénolate est donné par l’équation

$$\Gamma = -\frac{d\mathcal{U}}{d\varphi} \simeq -\frac{\mathcal{U}(\varphi + \delta\varphi) - \mathcal{U}(\varphi)}{\delta\varphi} \quad (\text{III.2})$$

où \mathcal{U} est l’énergie potentielle totale du système. Pour chaque configuration extraite des simulations biaisées listées dans le tableau III.2, nous avons calculé le couple de force exercé par chaque acide aminé sur le phénolate par différences finies en imposant manuellement une faible rotation du cycle phénolate autour de son axe CG2-CZ. Les valeurs moyennes du couple de force des acides aminés ayant une contribution majoritaire au couple de force total s’exerçant sur le phénolate sont présentées dans le tableau III.3.

Le signe du couple de force a une signification importante. Pour des valeurs de φ négatives, une valeur positive du couple de force correspond à un couple de rappel qui contribue donc à faire tendre le chromophore vers une géométrie plane. Inversement pour des valeurs de φ positives, le couple de force de rappel qui tend à contraindre le chromophore dans sa géométrie plane prend des valeurs négatives. Quelque soit la fenêtre que l’on considère le couple de force total est donc un couple de force de rappel car il est positif pour des valeurs négatives de φ et il est négatif pour des valeurs positives de φ . Inversement, le couple de force dû au potentiel intramoléculaire du chromophore n’est pas un couple de rappel et induit des forces en faveur de la rotation du cycle phénolate. Ces observations sont en accord, d’une part, avec le fait que le minimum du potentiel interne du

$\varphi < 0$		$\varphi > 0$	
acide aminé	$\Gamma(\text{kcal} \cdot \text{mol}^{-1} \cdot \text{degre}^{-1})$	acide aminé	$\Gamma(\text{kcal} \cdot \text{mol}^{-1} \cdot \text{degre}^{-1})$
$\varphi_0 = -30^\circ$	$\langle \varphi \rangle = -26.0^\circ$	$\varphi_0 = 40^\circ$	$\langle \varphi \rangle = 29.5^\circ$
Wat374	0.07	Thr62	-0.11
Thr62	0.05	Thr203	-0.06
<i>Chromophore</i>	-0.04	<i>Chromophore</i>	0.04
Wat370	0.01	Wat374	0.02
Thr203	-0.005	Wat370	-0.003
<i>total</i>	<i>0.13</i>	<i>all</i>	<i>-0.16</i>
$\varphi_0 = -50^\circ$	$\langle \varphi \rangle = -49.8^\circ$	$\varphi_0 = 60^\circ$	$\langle \varphi \rangle = 48.5^\circ$
<i>Chromophore</i>	-0.10	Thr62	-0.16
Thr62	0.09	<i>Chromophore</i>	0.11
Thr203	0.02	Thr203	-0.09
Wat370	0.01	Wat370	-0.01
Wat374	0.01	Wat374	0.01
<i>total</i>	<i>0.04</i>	<i>all</i>	<i>-0.19</i>
$\varphi_0 = -80^\circ$	$\langle \varphi \rangle = -77.3^\circ$	$\varphi_0 = 65^\circ$	$\langle \varphi \rangle = 62.3^\circ$
Thr62	0.09	Thr62	-0.15
<i>Chromophore</i>	-0.07	<i>Chromophore</i>	0.12
Thr203	0.03	Thr203	-0.04
Wat370	0.02	Wat370	-0.02
Wat374	0.01	Wat374	-0.01
<i>total</i>	<i>0.05</i>	<i>all</i>	<i>-0.14</i>

Tableau III.3 – Valeurs moyennes des couples Γ ($\text{kcal} \cdot \text{mol}^{-1} \cdot \text{degre}^{-1}$) exercé sur le cycle phénolate par les acides aminés proches du chromophore dans une sélection de simulations biaisées caractérisées par la valeur φ_0 du minimum du potentiel de biais. $\langle \varphi \rangle$ représente la valeur moyenne de φ dans ces simulations. Les lignes intitulées “total” correspondent au couple total exercé par l’ensemble des acides aminés de la protéine et des molécules d’eau cristallographiques. La ligne “chromophore” correspond à la contribution du potentiel intramoléculeaire du chromophore au couple total.

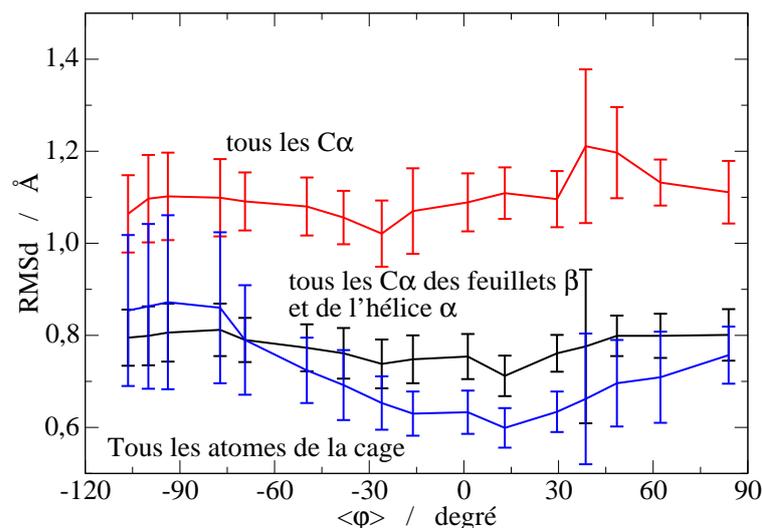


Figure III.14 – Variation de la moyenne du RMSd et de son écart type (barre d'erreur) en fonction de la moyenne de φ dans chaque simulation biaisée pour (i) les $C\alpha$ des acides aminés impliqués dans les feuillets β et l'hélice α centrale (en noir), (ii) l'ensemble des carbone α de la protéine (en rouge) et (iii) l'ensemble des atomes des acides aminés de la cage dans laquelle se trouve le chromophore dans la protéine (en bleu) (acides aminés 61, 62, 63, 68, 69, 94, 96, 145, 148, 150, 165, 167, 183, 203, 205, 222, 224). Les points pour $\varphi_0=0^\circ$ correspondent aux valeurs obtenues pour la simulation sans contrainte.

chromophore correspond à la géométrie twistée $TW\varphi$ dans laquelle φ prend des valeurs proches de 90° et, d'autre part, avec la géométrie d'équilibre plane obtenue pour la simulation dans la protéine sans contrainte. Le couple de force intramoléculaire exercé par le chromophore n'est pas suffisant pour contrer le couple de force créé par les autres acides aminés de la protéine.

On remarque également qu'un petit nombre d'acides aminés joue un rôle majeur dans le couple de force total qui s'exerce sur le phénolate. Ils sont représentés sur la figure III.15. Le type de molécule exerçant un couple de force majoritaire est très différent selon le sens de rotation du phénolate. Pour les valeurs positives de φ , il s'agit essentiellement de deux acides aminés thréonine en position 62 et 203 qui sont situés de part et d'autre du cycle phénolate. Pour les valeurs négatives de φ , petites en valeurs absolues, c'est une molécule d'eau structurale, wat374 qui exerce le couple de force le plus important. Pour des valeurs négatives, plus élevées en valeurs absolues, Thr62 exerce un couple important ainsi qu'une autre molécule d'eau structurale, wat370. Les molécules d'eau concernées sont présentes dans la structure RX et sont confinées à l'intérieur de la protéine. Leur déplacement est donc limité et nécessite de franchir des barrières d'énergie potentielle élevées. Ce point sera développé plus en détail par la suite.

III.3.3.4 Construction du profil d'énergie libre partielle

La figure III.16 présente les courbes d'énergie libre partielle $\mathcal{F}_i(\varphi)$ obtenues dans chaque simulation biaisée (figure a) ainsi que le profil d'énergie libre partielle obtenu une fois qu'elles sont raccordées (figure b). Le troisième graphique (figure c) compare les profils d'énergie libre partielle obtenus

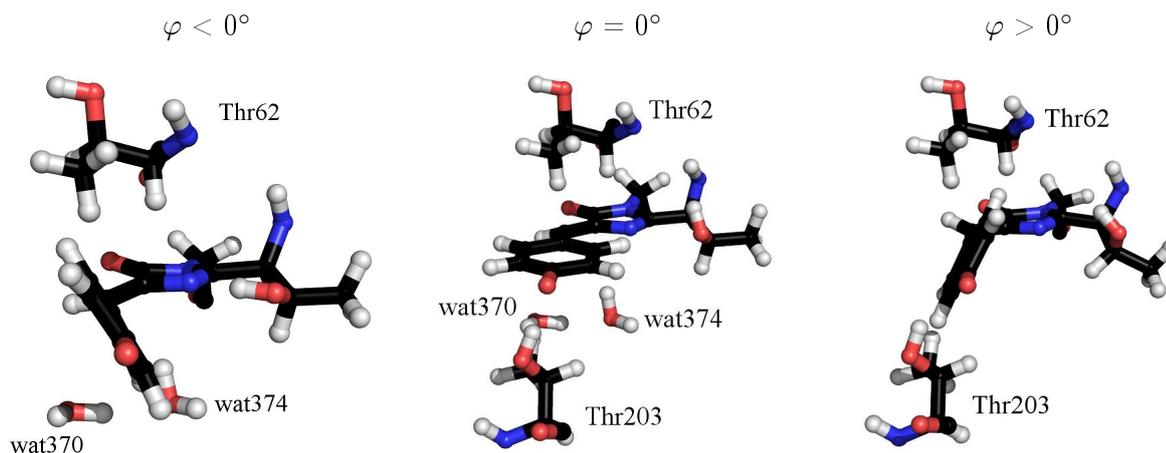


Figure III.15 – Aperçu du chromophore et des acides aminés et des molécules d’eau qui exercent un couple important sur le phénolate, à gauche dans le cas de $\varphi > 0^\circ$ et à droite dans le cas de $\varphi < 0^\circ$.

avec les deux fonctions de poids présentées au paragraphe I.4.3 permettant de raccorder les courbes $\mathcal{F}_i(\varphi)$. Le profil d’énergie libre partielle, en rouge sur la figure III.16 c, est fortement dissymétrique. Il présente un minimum d’énergie libre correspondant à la géométrie plane du chromophore et un plateau dans la zone $\varphi \in [-80^\circ; -40^\circ]$. La géométrie $TW\varphi$ dans laquelle $\varphi = 90^\circ$ se situe 4.2 ou 6.3 kcal·mol⁻¹ au dessus de la géométrie plane suivant que l’on considère respectivement la géométrie $\varphi = -90^\circ$ ou $\varphi = +90^\circ$. On remarque également que l’emploi de l’une ou l’autre des deux méthodes pour raccorder les courbes $\mathcal{F}_i(\varphi)$ a peu d’effet sur l’allure globale du profil d’énergie libre partielle. On vérifie ici qu’à partir d’un jeu de fenêtres données, la suppression nette des extrémités de l’histogramme biaisé de φ ou leurs prises en compte pondérées par la valeur de l’histogramme (méthode proposée par McCammon *et al* [52]) sont équivalentes pour positionner les courbes $\mathcal{F}_i(\varphi)$ les unes par rapport aux autres.

Le profil d’énergie libre montre que la protéine contraint le chromophore à rester plan ce qui est en accord avec le fait que la fluorescence soit intense dans la protéine et avec le signe des couples de force calculés pour les différentes fenêtres. La dissymétrie de la courbe obtenue traduit directement la différence des interactions mises en jeu suivant que le phénolate tourne dans le sens positifs ou négatifs que nous avons identifié à l’aide du calcul des couples de force au paragraphe précédent. C’est une manifestations de l’hétérogénéité de l’environnement du chromophore.

Ce profil d’énergie libre souffre cependant de plusieurs défauts liés aux difficultés de ce type de calcul dans un système complexe tel qu’une protéine. Tout d’abord, on remarque que dans la zone où φ prend des valeurs comprises dans l’intervalle $[-90^\circ; -30^\circ]$ les courbes $\mathcal{F}_i(\varphi)$ ne sont pas parallèles deux à deux. La barrière aux alentours de $\varphi \simeq 70^\circ$ n’est pas clairement définie. En effet, dans cette zone, les courbes $\mathcal{F}_i(\varphi)$ se croisent et cette barrière n’est pas présente en $\varphi \simeq -110^\circ$ alors que le profil d’énergie libre devrait avoir une périodicité de π compte tenu de la symétrie du cycle phénolate. Les origines de ces difficultés et certaines solutions sont présentées dans les paragraphes suivants.

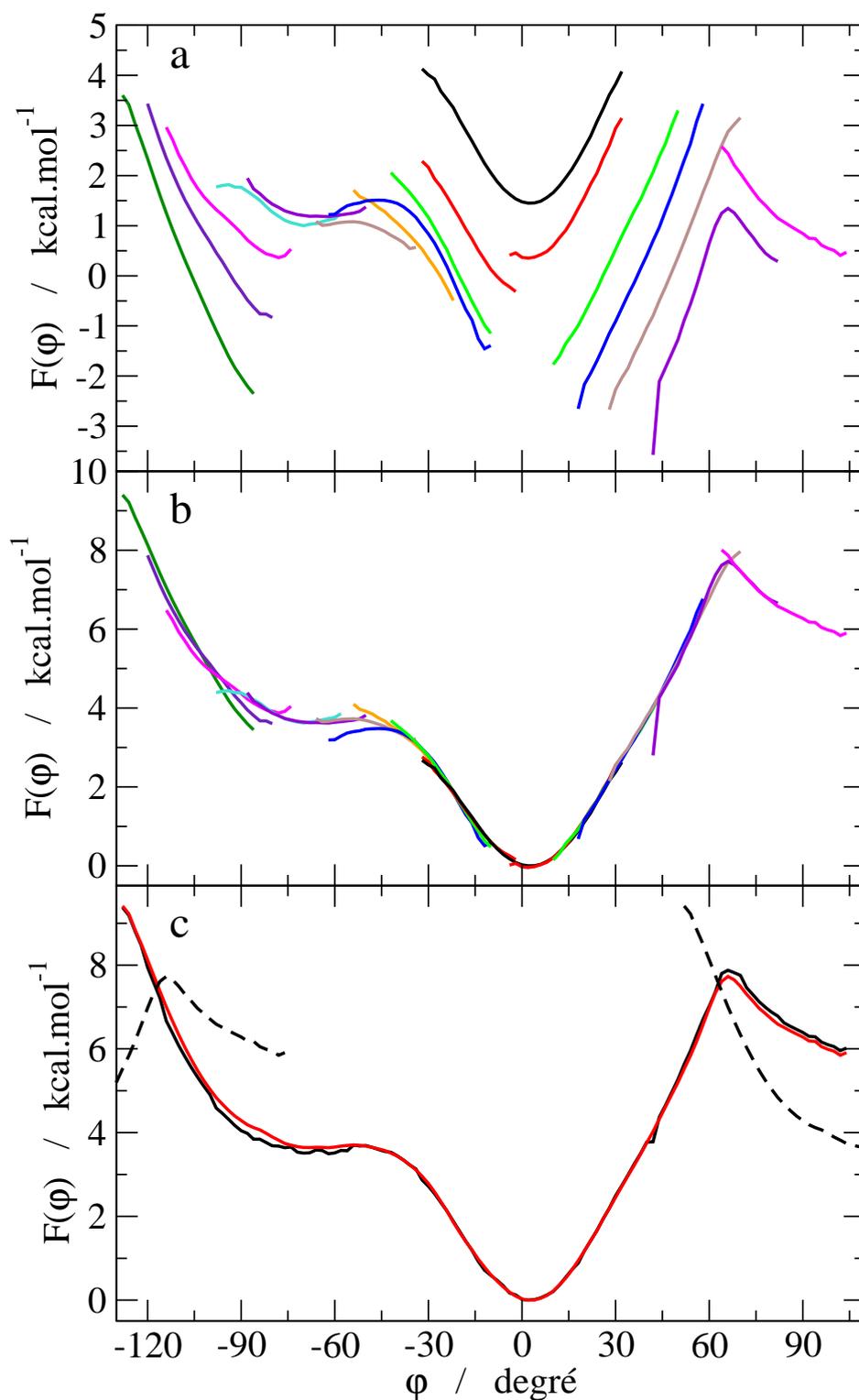


Figure III.16 – a) Courbes d'énergie libre partielle en fonction de φ obtenues pour chaque simulation biaisée listée dans le tableau III.2. b) Profil d'énergie libre complet obtenu après a raccordé les différents morceaux de la figure a). c) Comparaison du profil d'énergie libre partielle obtenu par la méthode standard (en noir) et par la méthode de McCammon (en rouge). Les courbes noires en pointillé correspondent à la courbe rouge translatée de $\pm 180^\circ$.

III.3.3.5 Echantillonnage de configurations autour d'une barrière d'énergie potentielle

Pour des valeurs de φ positives, on a montré que le couple de force qui s'exerce sur le phénolate est majoritairement dû aux deux acides aminés thréonine situés de part et d'autre du chromophore (figure III.15). La barrière située aux alentours de $\varphi \simeq 70^\circ$ correspond au franchissement par le phénolate du méthyle de Thr62. Cependant, ni le chromophore ni Thr62 n'étant fixes, la position précise de la barrière est difficile à localiser. Pour calculer le profil d'énergie libre, la simulation doit explorer une zone commune de l'espace des phases au cours de deux fenêtres consécutives ce qui nous permet alors d'imposer que les énergies libres partielles calculées dans cette zone à partir de deux simulations différentes sont égales. Au voisinage d'une barrière, la continuité de la courbe $\mathcal{F}(\varphi)$ ne peut être assurée que si une ou plusieurs simulations échantillonnent les deux côtés de la barrière au niveau du sommet. La difficulté pour obtenir un tel échantillonnage vient de la complexité de la surface d'énergie potentielle et du faible volume de l'espace des phases associé au sommet de la barrière.

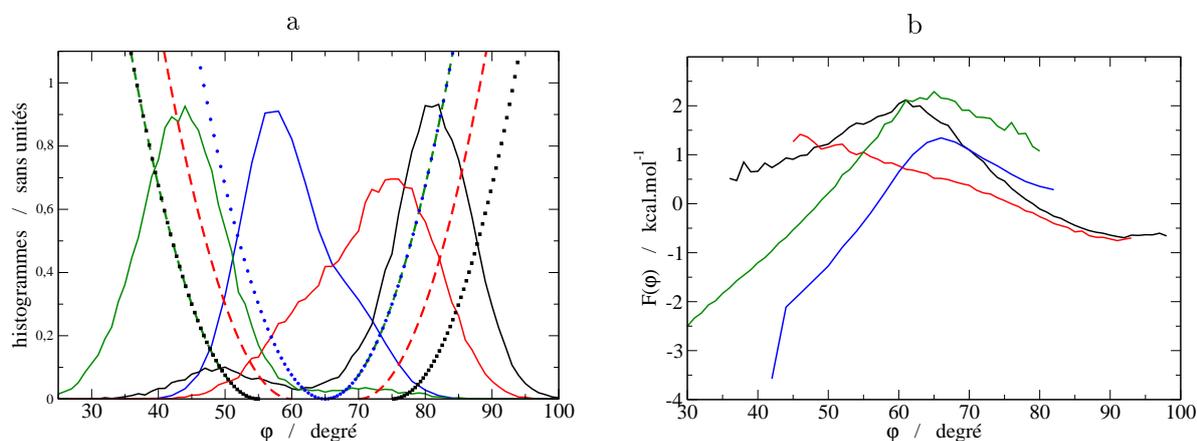


Figure III.17 – a : Histogrammes biaisés des valeurs de φ obtenus pour quatre simulations biaisées (courbes en trait plein). Les courbes en pointillé représentent les potentiels de biais utilisés. La couleur de l'histogramme et du potentiel de biais associé correspond. b : Courbes $\mathcal{F}_i(\varphi)$ obtenues pour les quatre simulations biaisées dont les histogrammes biaisés sont présentés sur la figure a.

Pour illustrer ce point, nous avons comparé plusieurs potentiels de biais en faisant varier la position du minimum et la constante de force, ou en utilisant une forme dans laquelle le potentiel est constant sur une gamme de valeur de φ (en rouge, vert et noir sur la figure III.17). Quelque soit la simulation, la trajectoire de φ se localise d'un côté ou de l'autre de la barrière. On échantillonne alors des configurations meta-stables, le sommet de la barrière n'est jamais échantillonné (figure III.17 b) et il est difficile de le localiser. En particulier, dans la simulation biaisée correspondant à l'histogramme bleu (figure III.17), la trajectoire de φ explore un côté de la barrière au début de la simulation puis l'autre côté dans la suite de la simulation, en passant une seule fois la barrière. Nous avons cependant conservé les résultats de cette simulation car la courbe $\mathcal{F}_i(\varphi)$ est parallèle avec les courbes $\mathcal{F}_i(\varphi)$ obtenues pour des fenêtres situées avant et après la barrière (figure III.17 b). Les deux situations échantillonnées semblent donc correctement représenter le système de part et d'autre de la barrière.

La présence de barrières peut également induire un phénomène d'hystérésis. Du fait de la présence de barrières, le système reste bloqué dans une région de l'espace des phases alors que cette région ne correspond pas au bassin de plus basse énergie. Ce phénomène entraîne une erreur sur la position du maximum de la barrière et se manifeste par l'absence de périodicité dans notre profil d'énergie libre. D'autre part, lorsque le système franchit la barrière, la cage formée par les acides aminés entourant le chromophore se déforme légèrement. Sur la figure III.14 on voit que pour des valeurs proches de $\varphi=50^\circ$, le RMSd de la cage est plus élevé et notamment l'écart type est important ce qui traduit une fluctuation importante de la structure de la cage. Une fois la barrière franchie, le système s'est donc déformé et il ne revient pas à sa structure initiale du fait d'un grand nombre de minimum locaux sur la surface d'énergie potentielle. Il s'en suit un phénomène d'hystérésis qui fait que le profil d'énergie libre partielle obtenu n'est pas périodique. De plus, la coordonnée φ est couplée à d'autres mouvements de la protéine, notamment ceux des acides aminés Thr62 et Thr203 et de molécules d'eau cristallographiques, les échelles de temps propres à ces différents mouvements étant du même ordre de grandeur.

Ainsi, même en contraignant la simulation à explorer un domaine restreint de valeurs de φ , des temps de simulations de quelques nanosecondes ne sont pas suffisants pour que le système explore le goulot d'étranglement qui conduit d'un côté à l'autre de la barrière. Une approche incluant plusieurs contraintes bien choisies permettrait peut-être de décrire correctement la barrière. Une méthode, plus globale comme la métadynamique serait probablement la plus adaptée. Nous n'avons pas eu le temps dans le cadre de ce travail de thèse de la mettre en oeuvre.

III.3.3.6 Bassins d'énergie multiples

Les difficultés rencontrées dans la zone où φ prend des valeurs comprises dans l'intervalle $[-90^\circ; -30^\circ]$ sont dues à la présence des molécules d'eau structurales wat374 et wat370. Ces molécules d'eau sont confinées à l'intérieur de la protéine et elles ont donc un mouvement limité car elles se positionnent dans des bassins d'énergie relativement profonds grâce à la formation de liaisons hydrogène (figure III.10). Cependant, elles ne sont pas reliées de manière covalente au reste de la protéine et à cause des interactions répulsives dues à la rotation du phénolate elles peuvent franchir des barrières d'énergie potentielle pour se positionner dans un bassin d'énergie voisin laissant la place au phénolate. Ce déplacement de wat374 est illustré par la figure III.18. Il est la cause de la présence du plateau observé sur le profil d'énergie libre entre -80° et -40° . En effet, pour des faibles valeurs de φ (en valeur absolue) la molécule d'eau bloque la rotation du phénolate tandis que pour des valeurs de φ entre -80° et -40° cette molécule d'eau s'est éloignée du phénolate qui est donc plus libre. Les couples de forces calculés pour les simulations biaisées avec $\varphi_0 < 0^\circ$ traduisent bien ces observations. Le couple de force exercé par wat374 est élevée pour la fenêtre avec $\varphi_0 = -30^\circ$ et diminue fortement dans la fenêtre avec $\varphi_0 = -50^\circ$. Les mouvements sur la coordonnée φ sont donc fortement corrélés à la position de la molécule d'eau wat374.

On se ramène donc à une situation équivalente à celle rencontrée pour la barrière due à Thr62 mais dans ce cas, le problème est lié à la barrière séparant les deux positions de wat374 et n'est donc pas directement sur la coordonnée φ . Pour que deux simulations biaisées échantillonnent des

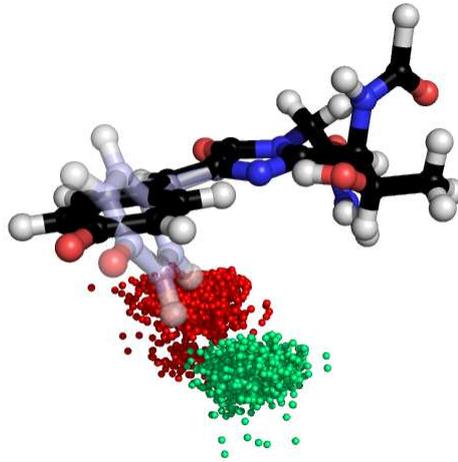


Figure III.18 – Déplacement de la molécule d’eau wat374 dû à la rotation du cycle phénol. Les points rouge et vert correspondent respectivement aux positions de l’atome d’oxygène de wat374 au cours des fenêtres correspondant à $\varphi_0 = -20^\circ$ et $\varphi_0 = -80^\circ$.

configurations dans une zone commune de l’espace des phases il est nécessaire que la molécule d’eau wat374 explore ses deux positions d’équilibre. Il est donc nécessaire d’échantillonner correctement la barrière entre ces deux positions. Or, compte tenu de la profondeur des bassins d’énergie libre dans lesquels se trouvent ces molécules d’eau, l’échantillonnage de plusieurs bassins au cours d’une simulation n’est pas envisageable nous avons alors utilisé un potentiel de biais portant sur la position de wat374. Avec AMBER, on ne peut pas ajouter un potentiel de biais directement sur des coordonnées cartésiennes. Nous avons donc ajouté un potentiel de biais sur la distance entre l’atome de carbone CA2 du chromophore et l’atome d’oxygène de la molécule d’eau wat374, $d_{\text{CA2-W374}}$, afin de contraindre wat374 loin du chromophore. Le potentiel de biais a été choisi à partir de la figure III.18 de sorte que wat374 se place dans la zone correspondant à sa position d’équilibre loin du chromophore (points vert).

Sur la figure III.19, nous avons représenté la corrélation entre les valeurs de φ et les valeurs de la distance $d_{\text{CA2-W374}}$ dans la simulation de la GFP sans contrainte et avec le potentiel de biais suivant

$$V^b(d_{\text{CA2-W374}}) = \begin{cases} 7.0 (d_{\text{CA2-W374}} - 4.5)^2 & \text{si } d_{\text{CA2-W374}} < 4.5 \text{ \AA} \\ 0 & \text{si } 4.5 < d_{\text{CA2-W374}} < 5.5 \text{ \AA} \\ 7.0 (d_{\text{CA2-W374}} - 5.5)^2 & \text{si } d_{\text{CA2-W374}} > 5.5 \text{ \AA} \end{cases} \quad (\text{III.3})$$

On remarque que dans les configurations où $d_{\text{CA2-W374}}$ est inférieure à 4.5 Å, φ prend des valeurs comprises entre -30° et 30° alors que dans les configurations où $d_{\text{CA2-W374}}$ est supérieure à 4.5 Å, φ prend des valeurs pouvant aller jusqu’à -100° . Il suffit donc que la molécule d’eau wat374 s’éloigne du phénolate pour qu’il ait une grande liberté de mouvement. Dans la simulation de la GFP sans contrainte, on observe que la simulation explore quelques fois des configurations dans lesquelles φ prend des valeurs inférieures à -45° (figure III.11 page 122 aux alentours de 0.5 ns et de 2 ns). On

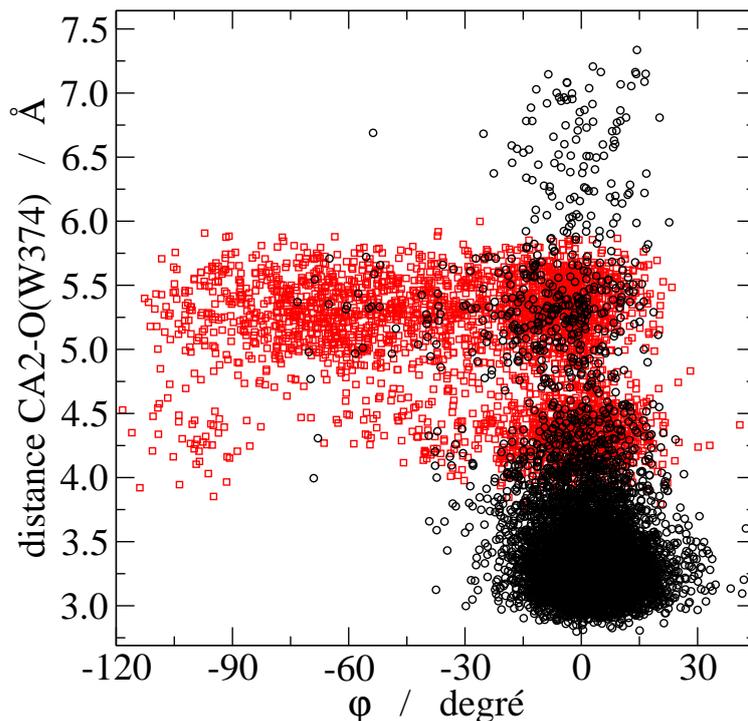


Figure III.19 – Corrélation entre φ et la distance entre l’atome de carbone CA2 et l’atome d’oxygène de wat374 dans la simulation de la GFP sans contrainte (en noir) et dans une simulation biaisée dans laquelle wat374 est éloignée du phénolate.

voit sur la figure III.19 que ces configurations correspondent en fait à des situations où la molécule wat374 s’est éloigné du phénolate.

Cette observation peut être mise à profit pour calculer le profil d’énergie libre partielle en fonction de φ . Il faudrait commencer par calculer le profil d’énergie libre partielle en fonction de la coordonnée $d_{\text{CA2-W374}}$ qui nous donnera la différence d’énergie libre entre les deux positions de la molécule d’eau wat374. On pourrait alors positionner l’une par rapport à l’autre les courbes $\mathcal{F}(\varphi)$ obtenues respectivement dans la simulation sans contrainte et dans la simulation dans laquelle on a contraint wat374 à s’éloigner du chromophore. Cependant, l’échantillonnage des configurations dans lesquelles $d_{\text{CA2-W374}}$ est supérieur à 4.5 Å dans la simulation de la GFP sans contrainte est trop faible pour pouvoir raccorder les courbes $\mathcal{F}(d_{\text{CA2-W374}})$ obtenues dans cette simulation et dans la simulation dans laquelle nous avons poussé wat374. Une étude complémentaire est donc nécessaire afin de calculer le profil d’énergie libre partielle en fonction de $d_{\text{CA2-W374}}$. Ce calcul devrait nous permettre d’échantillonner correctement toutes les positions de wat374 quelque soit la valeur de φ et ainsi éliminer les problèmes de raccord que nous avons rencontré.

Dans la littérature, l’algorithme WHAM pour *Weighted histogram Analysis Method* [53] est souvent mentionné pour calculer des profils d’énergie libre à partir de simulations biaisées et éventuellement à différentes températures dans des systèmes biologiques. Il est basé sur un calcul différent des constantes permettant de raccorder les courbes $\mathcal{F}_i(\varphi)$ entre elles. L’algorithme minimise l’incertitude sur le profil d’énergie libre obtenu à partir d’un ensemble de simulations et conduit à une

équation auto-cohérente qui doit être minimisée par itérations successives. Nous n'avons pas utilisé cet algorithme mais il paraît peu probable qu'il permette de résoudre les problèmes rencontrés ici. En effet, les principaux avantages de l'algorithme WHAM sont qu'il permet de minimiser et d'évaluer les erreurs statistiques dues à la taille des échantillons, et ne sont pas en liens avec les problèmes liés à l'échantillonnage de configurations au voisinage d'une barrière d'énergie potentielle.

III.4 Calcul du déclin de la fluorescence

Le profil d'énergie libre partielle obtenu au paragraphe III.3.3 nous donne la densité de probabilité que le chromophore anionique de la GFP dans son état électronique excité, à l'intérieur de la protéine, adopte une géométrie avec une valeur de φ donnée. Il nous permet de calculer la probabilité d'obtenir des configurations dans lesquelles, en accord avec le schéma de Martin *et al* [49] (figure III.2 page 109), le chromophore a une probabilité non nulle de retourner à l'état fondamental sans émettre de photon (figure III.20).

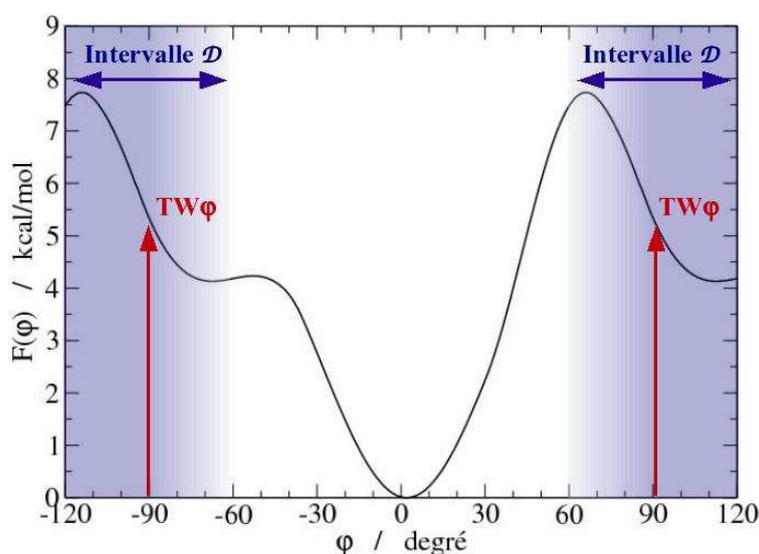


Figure III.20 – Représentation, de l'intervalle \mathcal{D} dans lequel une désactivation non radiative de l'état électronique excité par conversion interne est possible (en bleu) et de la position de la géométrie $TW\varphi$ (en rouge) à partir de laquelle le système peut atteindre une intersection conique et revenir à l'état fondamental de manière non radiative. La courbe noire représente le profil d'énergie libre partielle $\mathcal{F}(\varphi)$.

Nous nous sommes intéressés au calcul du rendement quantique de fluorescence et du déclin de la population de l'état excité fluorescent du chromophore de la GFP en incluant l'un des deux mécanismes de désactivations suivants :

- une désactivation de l'état excité fluorescent par conversion interne pour des géométries dans lesquelles l'angle de torsion φ prend des valeurs dans l'intervalle $\mathcal{D}=[60^\circ ; 120^\circ]$ (figure III.20). Ce mécanisme sera nommé "mécanisme de désactivation par conversion interne".

- une désactivation de l'état excité fluorescent par l'intermédiaire de l'intersection conique identifiée par Martin *et al* qui peut être atteinte à partir de la géométrie $TW\varphi$ dans laquelle φ vaut 90° (figure III.1 page 108). Ce mécanisme sera nommé "mécanisme de désactivation par une intersection conique".

Cette partie présente dans un premier temps les modèles cinétiques que nous avons utilisé. Ensuite, l'utilisation de la dynamique Brownienne pour alimenter les différents paramètres de ces modèles sera détaillée. Enfin, le rendement quantique de fluorescence, ϕ_f , et le déclin de la population de l'état excité seront calculés.

III.4.1 Modèle cinétique de l'état excité fluorescent

Pour calculer les constantes de vitesses des deux mécanismes de désactivation non radiatifs que nous avons considérés, nous avons construit deux modèles cinétiques différents. En effet, les deux mécanismes de désactivation ne peuvent pas être traités de la même manière. Pour le mécanisme de désactivation par une intersection conique, un seul point du domaine de variation de φ est intéressant. C'est celui pour lequel on se trouve dans la géométrie $TW\varphi$ (figure III.1 page 108). A l'inverse, pour le mécanisme de désactivation par conversion interne, une désactivation non radiative peut avoir lieu sur un ensemble de valeurs de φ comprises dans l'intervalle \mathcal{D} . Cette différence dans le type de domaine de valeurs de φ que l'on doit considérer impose une différence de traitement.

III.4.1.1 Cas du mécanisme de désactivation par conversion interne

Soit $N(t)$ la population de protéines dont le chromophore est dans le premier état électronique excité au temps t . En considérant que les deux seuls mécanismes de retour à l'état fondamental sont une désactivation radiative, la fluorescence, et une désactivation non radiative par conversion interne, l'évolution de la population $N(t)$ est donnée par

$$\frac{dN(t)}{dt} = -k_r N(t) - \mu Q(t) \quad (\text{III.4})$$

où k_r est la constante de vitesse de la fluorescence de l'état excité et μ est la probabilité qu'une désactivation non radiative par conversion interne ait lieu par unité de temps. La fonction $Q(t)$ représente la population de chromophores qui satisfont les deux conditions suivantes au temps t : la valeur de la torsion φ est comprise dans l'intervalle \mathcal{D} et le chromophore est toujours dans son état électronique excité (aucune désactivation radiative ou non radiative n'a eu lieu).

Pour évaluer la fonction $Q(t)$, nous avons supposé la relation suivante

$$\frac{Q(t)}{N(t)} = \frac{Q_{tot}(t)}{N(0)} \quad (\text{III.5})$$

où $Q_{tot}(t)$ est la population de chromophores tels que φ soit compris dans l'intervalle \mathcal{D} indépendamment des mécanismes de désactivation. Cette relation suppose que la dynamique de la coordonnée φ à l'état excité est indépendante du déclin de la population de l'état excité (évolution de $P(t)$). On

suppose donc que la probabilité que le chromophore retourne à l'état fondamental à un instant t , quelque soit le mécanisme, est indépendante de la dynamique de ce chromophore à ce même instant. On peut alors écrire l'équation III.4 de la façon suivante

$$\frac{dN(t)}{dt} = -k_r N(t) - \mu q(t) N(t) \quad (\text{III.6})$$

où la fonction $q(t)$ est la proportions de chromophore dont la coordonnée φ est comprise dans l'intervalle \mathcal{D} au temps t .

Nous avons extrait la valeur de μ des travaux expérimentaux réalisés par Gepshtein *et al* [101]. Ces auteurs ont mesuré expérimentalement le déclin de la fluorescence du chromophore anionique de la GFP en solution pour différentes températures et des solvants constitués d'un mélange d'eau et de glycerol. Ils utilisent ensuite un modèle cinétique qui leur permet de décrire un mécanisme de désactivation de la fluorescence par conversion interne. A l'aide de ce modèle et de leurs données ils ont calculé la constante de désactivation non radiative en fonction de la température. Dans notre modèle nous avons utilisé une borne supérieure de la valeur calculée par Gepshtein *et al*, $\mu = 48 \text{ ns}^{-1}$. De plus, nous avons supposé que μ est indépendant de φ dans l'intervalle \mathcal{D} est qu'il est nul partout ailleurs. Cette dernière approximation suppose que les surfaces d'énergie potentielle de l'état électronique excité et de l'état électronique fondamental du chromophore sont parallèles dans l'intervalle \mathcal{D} . Cette hypothèse se vérifie sur les calculs *ab initio* au niveau CASSCF//CASPT2 réalisés par Toniolo *et al* [48].

III.4.1.2 Cas du mécanisme de désactivation par une intersection conique

Dans ce mécanisme, le retour à l'état fondamental se fait *via* une intersection conique qui est accessible à partir de configurations du système correspondant à la géométrie $\text{TW}\varphi$ dans laquelle φ vaut $\pm 90^\circ$. La géométrie $\text{TW}\varphi$ peut en fait être considérée comme un état de transition. Dans la théorie de l'état de transition la probabilité pour qu'une désactivation non radiative ait lieu est proportionnelle à la probabilité que le système atteigne la géométrie $\text{TW}\varphi$ au temps t . Cette probabilité est donnée par la distribution des temps de premier passage à la géométrie $\text{TW}\varphi$.

Cependant, nous ne nous sommes pas placés dans les mêmes hypothèse que celles de la théorie de l'état de transition. Nous avons considéré, qu'une fois que le système atteint la géométrie $\text{TW}\varphi$ il retourne à l'état fondamental avec une probabilité λ différente de 1. Ce terme inclut la probabilité pour que lorsque le système atteint la géométrie $\text{TW}\varphi$, il accède à l'intersection conique et retourne à l'état fondamental. Du fait que la probabilité λ n'est pas égale à 1, nous avons considéré que le système peut passer plusieurs fois à la géométrie $\text{TW}\varphi$ (dans la théorie de l'état de transition on considère que toutes les trajectoires qui atteignent l'état de transition sont efficaces). L'évolution de la population $N(t)$ de chromophore à l'état électronique excité est alors donnée par

$$\frac{dN(t)}{dt} = -k_r N(t) - \sum_n \lambda^n p_n(t) N(0) \exp(-k_r t) \quad (\text{III.7})$$

où les fonctions $p_n(t)$ sont les distributions des temps de $n^{\text{ème}}$ passage à la géométrie $\text{TW}\varphi$. Ces

distributions sont une généralisation de la distribution des temps de premier passage et permettent de modéliser le fait que le système peut passer plusieurs fois à la géométrie $TW\varphi$. Le dernier facteur, $N(0)\exp(-k_r t)$ est la population de protéine dont le chromophore est encore à l'état excité si le seul mécanisme de désactivation possible était la fluorescence.

Afin de pouvoir calculer le déclin de fluorescence et le rendement quantique de fluorescence en prenant en compte l'un ou l'autre de ces deux mécanismes il faut résoudre les équations III.6 et III.7. Dans le cas du mécanisme de désactivation par conversion interne (équation III.7) il reste à déterminer la fonction $q(t)$. Dans le cas du mécanisme de désactivation par une intersection conique (équation III.7) il faut déterminer les distributions $p_n(t)$. Ces deux quantités ne peuvent ni être obtenues par une simulation de dynamique moléculaire, ni directement à partir du profil d'énergie libre partielle car nous avons vu que le mouvement que l'on cherche à modéliser est un événement rare par rapport aux échelles de temps de la dynamique du système (paragraphe III.3.2 page 118). En effet, les zones de l'espace des configurations dans lesquelles la probabilité pour que le chromophore retourne à l'état fondamental est non nulle ne correspondent pas à la géométrie d'équilibre du système (figure III.20 page 134). Nous avons évalué la hauteur des barrières que l'on doit franchir pour atteindre ces configurations à l'aide du profil d'énergie libre partielle $\mathcal{F}(\varphi)$ et on voit clairement qu'elles sont au moins 8 fois supérieures à l'énergie thermique (à 300 K, $k_B T$ vaut environ $0.6 \text{ kcal} \cdot \text{mol}^{-1}$). Pour obtenir un échantillonnage suffisant des mécanismes de désactivation que l'on considère, il est donc nécessaire d'utiliser une description du système moins détaillée, permettant d'obtenir des informations à une échelle de temps qui correspond à celle des mouvements que l'on souhaite étudier. En considérant que la coordonnée φ est la coordonnée de réaction, nous avons mis en oeuvre des simulations de dynamique Brownienne de cette coordonnée qui nous ont permis de calculer la fonction $q(t)$ et les distributions $p_n(t)$.

III.4.2 Echantillonnage d'évènement rare par dynamique Brownienne

La dynamique Brownienne est basée sur la séparation de l'ensemble des coordonnées du système en deux groupes. Le premier groupe contient les coordonnées du système qui décrivent la réaction chimique ou la déformation que l'on considère (les coordonnées de réaction). Le mouvement le long de ces coordonnées se déroule à une échelle de temps qui est celle du mécanisme étudié et qui est généralement plus longue que celle de la dynamique du système autour de sa position d'équilibre. Le second groupe contient les autres coordonnées pour lesquelles le temps caractéristique du mouvement est plus court que celui des coordonnées de réaction. Ce groupe contient les coordonnées peuvent être décrits par un effet moyen. La principale approximation que l'on fait lorsqu'on utilise la dynamique Brownienne consiste à ne considérer que la dynamique des coordonnées de réaction et à modéliser les autres coordonnées par une force de friction et une force aléatoire. Ces deux forces modélisent les échanges d'énergie entre les coordonnées de réaction et les autres coordonnées du système.

Dans notre cas, nous considérerons la torsion φ comme coordonnée de réaction. Comme on considère une torsion, les forces sont remplacées par des couples de forces et la masse est remplacée par un moment d'inertie, noté I . L'équation de Langevin qui remplace l'équation de Newton pour la

dynamique Brownienne s'écrit alors

$$I \ddot{\varphi} = \Gamma(\varphi) - f \dot{\varphi} + \tilde{\Gamma} \quad (\text{III.8})$$

où f est le coefficient de friction, $\Gamma(\varphi)$ est le couple de force qui dérive du potentiel dans lequel évolue la coordonnée φ et $\tilde{\Gamma}$ est un couple aléatoire. Un aperçu plus général de la dynamique Brownienne et de la résolution numérique de l'équation III.8 sont donnés en annexe C. Le coefficient de diffusion, D , est généralement plus utilisé que le coefficient de friction. Notamment dans les calculs des coefficients de transport en phase condensée. Dans le cas d'un mouvement de rotation, le coefficient de friction est relié au coefficient de diffusion par la relation [120, p 93]

$$D = \frac{k_B T}{f}$$

Dans notre étude du mouvement de rotation du phénol autour de son axe, le terme de friction est plus adapté. La force de friction représente effectivement une force de frottement fluide entre les atomes du phénolate et l'environnement du chromophore alors que les coefficients de diffusion sont plus adaptés dans le cas de l'étude de phénomène de transport. Nous utiliserons cependant le terme de coefficient de diffusion qui est plus généralement employé.

Le moment, d'inertie est obtenu par la relation

$$I = \sum_i m_i r_i^2$$

où la somme porte sur les atomes du cycle phénolate, m_i est la masse de l'atome i et r_i est la distance à l'axe passant par les atome CG2 et CZ. La torsion φ correspond à une rotation du cycle phénolate autour de cet axe qui est un axe de symétrie.

Le couple de force $\Gamma(\varphi)$ doit représenter le potentiel dans lequel évolue la coordonnée φ en faisant une moyenne sur toutes les configurations du système à une valeur de φ donnée. Il correspond en fait au profil d'énergie libre $\mathcal{F}(\varphi)$ que nous avons calculé au paragraphe précédent III.3.3. Dans notre cas, il représente l'effet de la protéine sur la coordonnée φ . Le couple de force est alors obtenu par la relation

$$\Gamma(\varphi) = - \frac{d\mathcal{F}(\varphi)}{d\varphi}$$

Pour calculer cette dérivée, nous avons utilisé une expression analytique du profil d'énergie libre partielle que nous avons ajusté sur la courbe de la figure III.16. Nous avons utilisé une expression telle que le profil d'énergie libre et donc le couple de force soient périodique, de période π . Cette périodicité est justifiée par la symétrie du cycle phénolate.

Cette modélisation très simple du système dans laquelle on ne considère plus qu'une seule coordonnée nous permet d'augmenter largement les échelles de temps de simulations tout en conservant l'effet de la protéine sur la dynamique de la coordonnée de réaction φ . Pour faire les simulations de dynamique Brownienne de la coordonnée φ nous avons écrit un programme utilisant les équations

présentées en annexe C. Cependant, il reste à définir le couple de force de friction et le couple de force aléatoire présentés dans l'équation III.8. Ces deux termes dépendent de deux paramètres : le coefficient de diffusion D et le pas de temps δt utilisé pour l'intégration de la dynamique. Les deux paragraphes suivants traitent de la détermination de ces deux grandeurs.

III.4.2.1 Calcul d'un coefficient de diffusion

Nous avons utilisé deux approches différentes pour calculer le coefficient de diffusion qui sont décrites ci-dessous. La première utilise une fonction d'auto-corrélation, la seconde plus originale utilise la vitesse moyenne limite du système.

a) Utilisation des fonctions d'auto-corrélation

La méthode la plus courante pour calculer un coefficient de diffusion consiste à utiliser la fonction d'auto-corrélation des fluctuations d'une coordonnée à l'équilibre [121, p 60]. Elles sont calculées à partir de simulations de dynamique moléculaire non contraintes, telles que celles décrites au paragraphe III.3.2. Une hypothèse est cependant nécessaire sur la forme du profil d'énergie libre partielle. On va supposer que celui-ci a une forme quadratique, de constante de force k . Le couple de force $\Gamma(\varphi)$ est alors donné par

$$\Gamma(\varphi) = -k(\varphi - \varphi_0) \quad (\text{III.9})$$

où φ_0 est la position du minimum du profil d'énergie libre qui est égale à la moyenne de φ sur la simulation sans contrainte. Cette approximation est valable si on considère les valeurs de φ petites en valeur absolue qui sont celles que l'on obtient effectivement par une simulation sans contrainte (figure III.11 page 122).

La fonction d'auto-corrélation à l'équilibre des fluctuations de la coordonnée φ est calculée à partir d'une trajectoire de dynamique moléculaire. Elle est donnée par la relation

$$C(t) = \frac{\langle (\varphi(t_0) - \bar{\varphi})(\varphi(t_0 + t) - \bar{\varphi}) \rangle_{t_0}}{\sigma_\varphi} \quad (\text{III.10})$$

$$C(t) = \frac{\langle \varphi(t_0)\varphi(t_0 + t) \rangle_{t_0} - \bar{\varphi}^2}{\sigma_\varphi} \quad (\text{III.11})$$

où $\bar{\varphi}$ est la moyenne de φ sur l'ensemble de la simulation et la moyenne sur t_0 porte sur une succession de conditions initiales régulièrement espacées le long de la trajectoire. Pour $t=0$, le numérateur de l'équation III.10 est égal aux fluctuations de φ sur la simulation et la fonction d'auto-corrélation est égale à 1. On dit qu'elle est normalisée. En pratique on utilise une trajectoire finie pour calculer la fonction d'auto-corrélation et on parcourt la trajectoire pour faire une moyenne sur t_0 . En conséquence, la statistique est grande pour des temps t courts mais généralement faible pour des temps t longs. Cependant, la fonction d'auto-corrélation devrait tendre rapidement vers zéro et seul les temps courts sont intéressants. Un exemple de fonction d'auto-corrélation est présentée sur la figure III.21.

Si on multiplie par $\varphi(t_0)$ l'équation III.8 et que l'on fait une moyenne sur t_0 , par exemple sur un grand nombre de trajectoires, on obtient

$$I\ddot{C}(t) = -kC(t) - \frac{k_B T}{D}\dot{C}(t)$$

Le terme correspondant au couple aléatoire $\tilde{\Gamma}$ s'annule. En effet, les variations de $\tilde{\Gamma}$ et de $\varphi(t_0)$ ne sont pas corrélées et la moyenne de leur produit est donc égal au produit de leur moyenne. Or la moyenne de $\tilde{\Gamma}$ est nulle par construction (équation C.3 annexe C). En intégrant, cette équation entre $t=0$ et $t=+\infty$, et en utilisant le fait que $C(0)=1$, $C(\infty)=0$, $\dot{C}(0)=0$ et $\dot{C}(\infty)=0$, on obtient que

$$\int_0^\infty C(t) dt = \frac{k_B T}{Dk}$$

Les valeurs de $C(0)$, $C(\infty)$ et $\dot{C}(\infty)$ utilisées sont une conséquence directe de la définition de la fonction d'auto-corrélation qui vaut 1 en $t=0$ et s'annule à l'infini (figure III.21). De plus, si on remplace t par $-t$ dans l'équation III.10, on voit que $C(t)=C(-t)$. La fonction d'auto-corrélation est donc une fonction paire et est symétrique par rapport à l'axe des ordonnées. De ce fait la dérivée en 0 de la fonction $C(t)$ est nulle. On peut donc calculer le coefficient de diffusion D , connaissant la constante de force k et en calculant l'intégrale de la fonction d'auto-corrélation des fluctuations de φ .

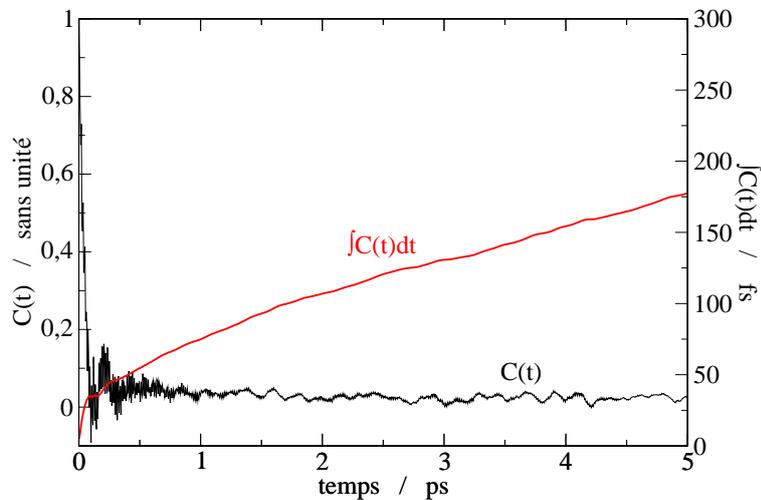


Figure III.21 – Fonction d'auto-corrélation $C(t)$ de la coordonnée φ sur une trajectoire dans l'ensemble (N,V,E) à l'équilibre (en noir). La courbe rouge est l'intégrale de la fonction $C(t)$ entre 0 et t .

La figure III.21 présente la fonction d'auto-corrélation obtenue sur une trajectoire de 1 ns dans l'ensemble (N,V,E)^a et l'intégrale de cette fonction. On remarque que la fonction d'auto-corrélation

^aL'utilisation de simulations dans l'ensemble (N,V,E) est recommandée pour calculer des fonctions d'auto-corrélation. En effet, le couplage du système avec un thermostat ou un barostat induit une corrélation artificielle dans les trajectoires car le contrôle de la température ou de la pression entraîne une modification des vitesses pi des positions à intervalle régulier.

se rapproche de 0 en une centaine de femtosecondes mais que contrairement à ce qui est attendu elle n'oscille pas exactement autour de 0. Ceci explique le fait que l'intégrale de la fonction d'auto-corrélation diverge et ne permet pas de calculer un coefficient de diffusion. Nous avons fait plusieurs essais pour améliorer le calcul de la fonction d'auto-corrélation qui n'ont pas été concluants et nous avons envisagé une autre approche pour calculer le coefficient de diffusion.

b) Utilisation de la vitesse limite

Pour calculer le coefficient de diffusion, nous avons étudié par simulation de dynamique moléculaire la relaxation du système à partir de configurations dans lesquelles le chromophore prend des valeurs de φ proches de 40° . Nous avons réalisé un ensemble de 260 trajectoires sans contrainte sur la coordonnée φ à partir de configurations que nous avons extraites des simulations biaisées présentées au paragraphe III.3.3. Les valeurs de φ dans ces configurations sont en moyenne de 43° avec un écart type de 2° . La figure III.22 présente l'évolution en fonction du temps de la moyenne de φ sur l'ensemble des trajectoires.

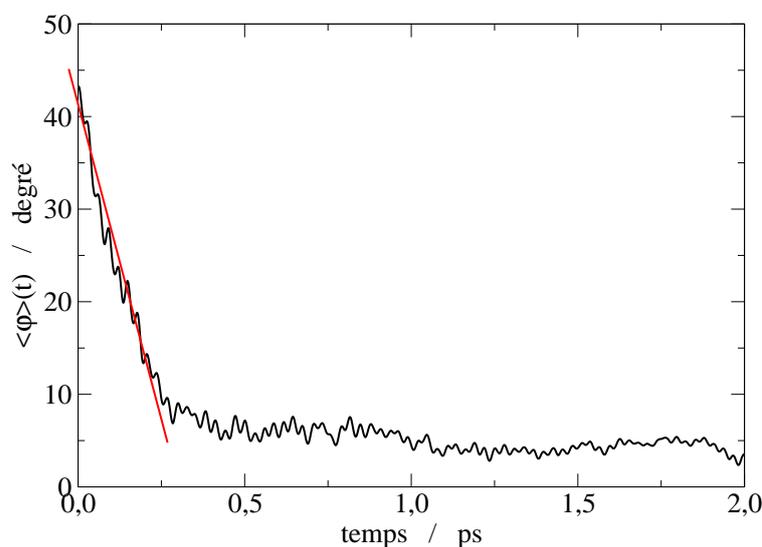


Figure III.22 – Valeur moyenne $\langle \varphi \rangle(t)$ sur un ensemble de trajectoires hors équilibre ayant comme point de départ des configurations dans lesquelles $\varphi \simeq 40^\circ$.

La trajectoire moyenne, $\langle \varphi \rangle(t)$ décroît rapidement et prend des valeurs inférieures à 10° après seulement 0.25 ps. Le système revient donc rapidement dans la zone de valeur de φ qui correspond à la région échantillonnée par la simulation sans contrainte présentée au paragraphe III.3.2. Cependant, la valeur de $\langle \varphi \rangle$ ne se stabilise pas autour de 0° . Le système est probablement bloqué dans un minimum local de la surface d'énergie potentielle et le retour dans le minimum global serait probablement observé en faisant des simulations beaucoup plus longues.

On remarque également que la courbe $\langle \varphi \rangle(t)$ est linéaire entre $t=0$ et $t=0.25$ ps (droite rouge sur la figure III.22). La vitesse de $\langle \varphi \rangle(t)$ a atteint une valeur limite que l'on peut utiliser pour calculer le coefficient de diffusion. En effet, cette vitesse limite traduit le fait que la dynamique du système a atteint un régime dans lequel elle est limitée par la diffusion. Si on fait une moyenne sur

un ensemble de trajectoires de l'équation de Langevin (équation III.8), on obtient la relation

$$I \frac{d^2}{dt^2} \langle \varphi \rangle = \langle \Gamma(\varphi) \rangle - \frac{k_B T}{D} \frac{d}{dt} \langle \varphi \rangle$$

où le terme correspondant au couple de force aléatoire $\tilde{\Gamma}$ s'annule étant donné que sa moyenne est nulle (équation C.4 annexe C). En supposant que la vitesse de $\langle \varphi \rangle(t)$ a atteint sa valeur limite, alors $\langle \varphi \rangle(t)$ est linéaire et la dérivée seconde s'annule. On obtient alors l'expression de la vitesse moyenne limite donnée par l'équation III.12.

$$\frac{d}{dt} \langle \varphi \rangle = -\frac{D}{k_B T} \langle \Gamma(\varphi) \rangle \quad (\text{III.12})$$

On utilise maintenant le fait que la courbure du profil d'énergie libre est faible dans l'intervalle $[0^\circ; 40^\circ]$ (figure III.16 page 129) ce qui nous permet de poser que $\langle \Gamma(\varphi) \rangle \simeq \Gamma_0$ où Γ_0 est indépendant de φ . Dans cette hypothèse, la vitesse moyenne de $\langle \varphi \rangle$ atteint une valeur limite qui est donnée par l'équation

$$\frac{d}{dt} \langle \varphi \rangle \simeq -\frac{D}{k_B T} \Gamma_0$$

On obtient alors la valeur suivante du coefficient de diffusion :

$$D = 2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$$

Cette valeur est en bon accord avec celles obtenues pour la rotation du cycle phénol d'une tyrosine dans une protéine bovine [122] ($2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$) et pour la dynamique d'un peptide d'alanine en solution aqueuse [123] ($1.5 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$) calculées à partir de fonction d'auto-corrélation à l'équilibre. La comparaison de la moyenne des trajectoires de relaxation du système, obtenues par dynamique moléculaire et par dynamique Brownienne avec la valeur calculée du coefficient de diffusion, montre que l'échelle de temps du retour à l'équilibre est bien reproduite (figure III.23).

L'utilisation de la relaxation du système pour déterminer le coefficient de diffusion est bien adaptée au fait que l'on souhaite étudier la dynamique de la coordonnée sur l'ensemble de son domaine de variation et pas seulement pour des valeurs proches de l'équilibre. En effet, cela nous permet de prendre en compte la diversité de l'environnement du chromophore suivant la valeur de φ . Cependant, plusieurs effets sont négligés. Tout d'abord, nous avons utilisé uniquement des trajectoires de la relaxation du chromophore pour lesquelles dans la configuration de départ, la valeur de φ est positive. Or nous avons vu au paragraphe III.3.3 que l'effet de l'environnement est fortement dissymétrique pour des valeurs de φ supérieures à 40° (en valeur absolue). D'autre part, nous avons choisi de prendre un coefficient de diffusion indépendant de φ . Cette approximation suppose que l'environnement du chromophore est homogène quelque soit la valeur de φ .

c) Effet du coefficient de diffusion sur la distribution des temps de premier passage

Nous avons calculé les distributions de temps de premier passage à la géométrie TW_φ pour plusieurs valeurs du coefficient de diffusion. Ces calculs nous permettront d'évaluer leurs sensibilité en fonction

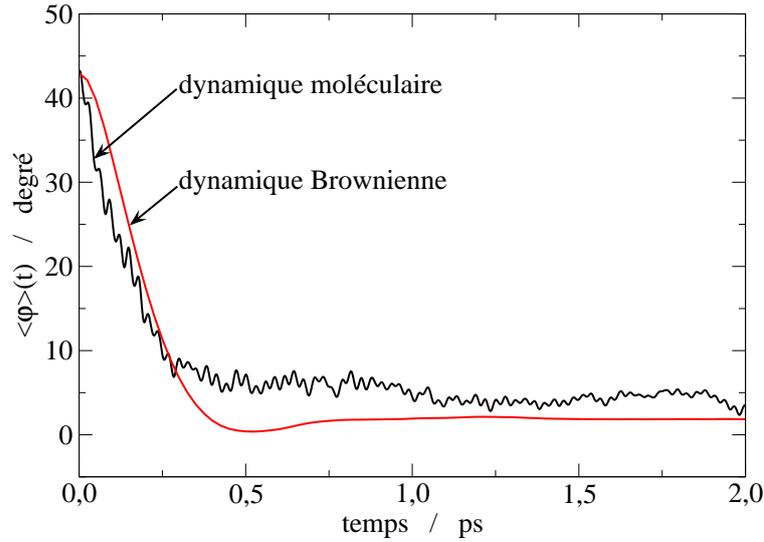


Figure III.23 – Comparaison de la moyenne des trajectoires obtenues par dynamique moléculaire et par dynamique Brownienne avec un coefficient de diffusion de valeur $D = 2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$

du coefficient de diffusion afin de vérifier s'il doit être obtenu avec plus de précision. Les distributions obtenues sont présentées sur la figure III.24. Elles ont été calculées en réalisant 10^5 trajectoires de 30 ns et en enregistrant pour chaque trajectoires le temps de premier passage à la géométrie $\text{TW}\varphi$.

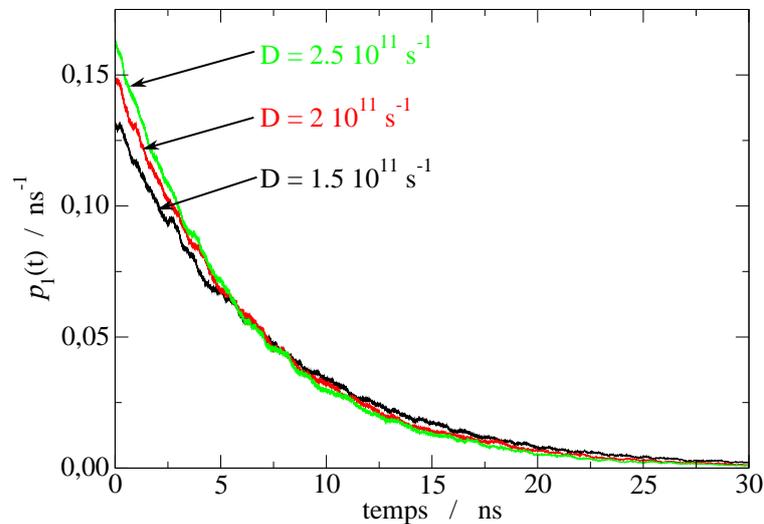


Figure III.24 – Distribution des temps de premier passage à la géométrie $\text{TW}\varphi$

Les distributions des temps de premier passages à la géométrie $\text{TW}\varphi$ sont peu sensibles à la valeur du coefficient de diffusion. En effet, une variation de 25% de la valeur du coefficient de diffusion modifie peu la distribution des temps de premier passages à la géométrie $\text{TW}\varphi$. Il semble donc que la dynamique Brownienne soit gouvernée par le couple de force $\Gamma(\varphi)$. La valeur de $D = 2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$ a donc été retenue pour les simulations de dynamique Brownienne.

III.4.2.2 Choix du pas de temps

Le pas de temps, δt , est crucial pour les simulations de dynamique Brownienne car la largeur de la distribution de probabilité qui est utilisée pour calculer les déplacements aléatoires des coordonnées et des vitesses dépend de δt (équation C.9 179). L'utilisation de valeurs de δt trop grandes peut conduire à un échantillonnage non physique. Cet effet est illustré sur la figure III.25.

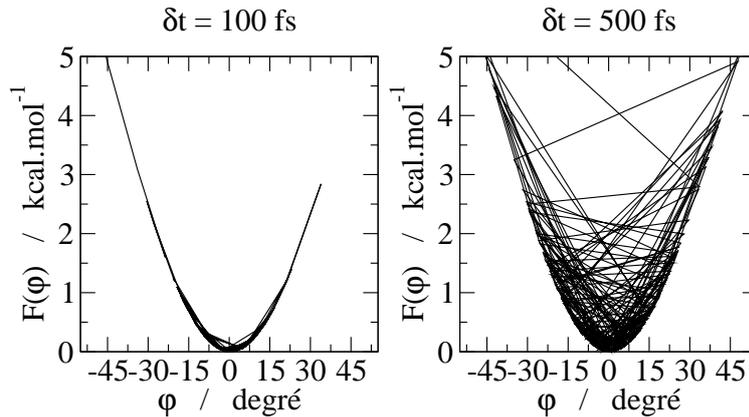


Figure III.25 – Déplacement de la coordonnée φ au cours d'une trajectoire Brownienne suivant que l'on utilise un pas de temps δt de 100 fs ou de 500 fs.

On remarque que pour un pas de temps de 100 fs, les déplacements successifs sur la coordonnée φ suivent bien la courbure du potentiel. A l'inverse, pour un pas de temps de 500 fs, un grand nombre de déplacements correspondent à des sauts d'un côté à l'autre du potentiel et sont non physiques. Le pas de temps considéré doit donc être d'autant plus petit que la courbure du potentiel est grande.

Le choix du pas de temps δt doit donc être un compromis entre une bonne description de la dynamique et un temps de simulation suffisamment court. Nous avons utilisé deux approches différentes pour estimer une valeur maximale du pas de temps.

Nous avons comparé la valeur de la largeur σ_{φ}^{rand} de la distribution à partir de laquelle sont calculés les déplacements aléatoires et l'écart type des valeurs de φ dans la simulation de dynamique moléculaire sans contrainte, σ_{φ}^{MD} qui donne un ordre de grandeur des valeurs de φ explorées (figure III.26). Pour que la trajectoire soit acceptable on considérera que σ_{φ}^{rand} doit être nettement inférieur à σ_{φ}^{MD} .

On remarque que σ_{φ}^{rand} est supérieur à σ_{φ}^{MD} pour des valeurs de δt supérieures à 200 fs, ce qui justifie l'allure des déplacements le long de la coordonnées φ sur la figure III.25. Pour que σ_{φ}^{rand} soit nettement inférieur à σ_{φ}^{MD} , les valeurs de δt à considérer doivent être inférieures à 50 fs.

Une autre approche consiste à prendre en compte le fait que, pour intégrer numériquement l'équation de Langevin (équation III.8), on suppose que le couple de force $\Gamma(\varphi)$ est constant pendant le pas de temps δt . On doit donc le choisir de sorte que la variation de $\Gamma(\varphi)$ soit faible. Par exemple on

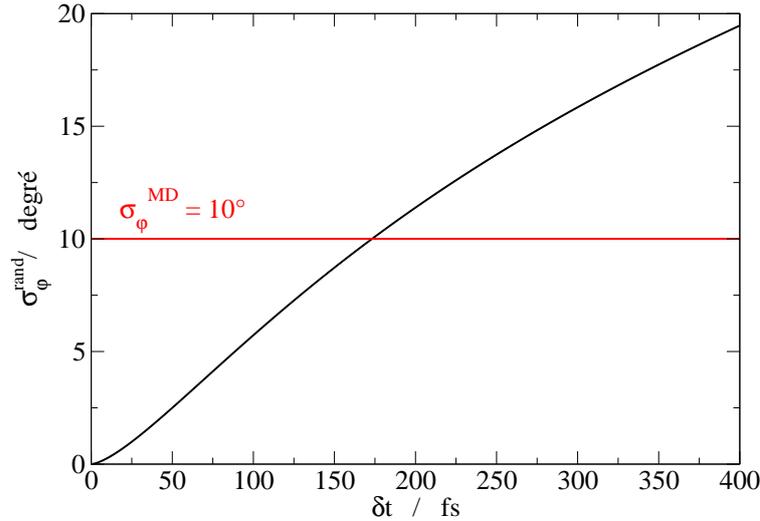


Figure III.26 – En noir, variation en fonction du pas de temps δt de la largeur de la distribution à partir de laquelle sont calculés les déplacements aléatoires. En rouge, écart type des valeurs de φ dans la simulation de dynamique moléculaire sans contraintes présentées au paragraphe III.3.2.

peut imposer que la variation de $\Gamma(\varphi)$ soit faible devant sa propre valeur $\Gamma(\varphi)$ ce qui s'exprime par

$$\dot{\varphi} \delta t \frac{d}{d\varphi} \Gamma(\varphi) < \Gamma(\varphi)$$

où $\dot{\varphi} \delta t$ est la variation de φ pendant le temps δt et le terme de gauche représente la variation de $\Gamma(\varphi)$ pendant δt . En considérant la vitesse maximale de φ , égale à la vitesse moyenne limite, donnée à l'équation III.12, on aboutit à une condition sur δt qui dépend de la courbure du profil d'énergie libre [124, 123].

$$\delta t \ll \delta t_{max} = \left(\frac{D}{k_B T} \left| \frac{d^2 \mathcal{F}(\varphi)}{d\varphi^2} \right| \right)^{-1} \quad (\text{III.13})$$

On retrouve ici le critère qualitatif selon lequel, le pas de temps doit être d'autant plus petit que la courbure du profil d'énergie libre est grande. Dans notre cas, la courbure est maximale pour des valeurs de φ dans l'intervalle $[-30^\circ ; 30^\circ]$ (figure III.16 page 129) où, comme dans le cas de la fonction d'auto-correlation, on peut supposer que le profil d'énergie libre a une forme harmonique (équation III.9). On peut alors en déduire une valeur pour la dérivée seconde du profil d'énergie libre, $k = 22.5 \text{ kcal.mol}^{-1}$. En utilisant cette valeur, on obtient $\delta t_{max} = 150 \text{ fs}$ avec $D = 2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$. Cette condition est en accord avec la première condition portant sur la largeur σ_φ^{rand} . Finalement, nous avons choisi un pas de temps $\delta t = 25 \text{ fs}$ qui satisfait les deux conditions et nous permet de réaliser très rapidement des trajectoires de plusieurs dizaines de nanosecondes.

III.4.3 Intégration des équations cinétiques

Avec les paramètres précédemment déterminés nous avons réalisé un ensemble de 10^6 trajectoires de dynamique Brownienne de l'état électronique excité du chromophore, d'une durée de 50 ns chacune.

Les conditions initiales de ces trajectoires ont été choisies de sorte qu'elles correspondent à un échantillon de la zone Franck-Condon. Nous avons donc pris des valeurs de $\varphi(t=0)$ tirées au hasard selon une distribution gaussienne centrée sur 0° et de largeur 4° qui représente la distribution des valeurs de φ dans l'état fondamental [24, 75]. Nous avons déterminé une valeur initiale de la vitesse de manière aléatoire à partir d'une densité de probabilité de Maxwell-Boltzmann à 300 K.

Les résultats présentés dans ce qui suit intègrent l'ensemble de ces trajectoires. Nous avons alors cherché à calculer la fonction $q(t)$ qui intervient dans le modèle de désactivation par conversion interne (équation III.6 page 136), et les distributions des temps de $n^{\text{ème}}$ passage à la géométrie TW φ , $p_n(t)$ qui interviennent dans le modèle de désactivation par une intersection conique (équation III.7 page 136). Puis nous avons résolu les équations cinétiques pour les deux modèles pour calculer le déclin de la fluorescence et le rendement quantique de fluorescence.

III.4.3.1 Cas du mécanisme de désactivation de la fluorescence par conversion interne

Dans le cas où l'on considère que le mécanisme de désactivation non radiatif de la fluorescence est le mécanisme par conversion interne, on doit résoudre l'équation III.6 page 136. Il faut alors déterminer la fonction $q(t)$.

a) Probabilité de se trouver dans la région $\varphi \in \mathcal{D} = [60^\circ; 120^\circ]$

La fonction $q(t)$ donne la proportion, au temps t , des protéines dont la coordonnée φ appartient au domaine \mathcal{D} . Pour calculer cette fonction, nous avons évalué, sur l'ensemble des trajectoires de dynamique Brownienne, le temps pendant lequel la trajectoire est restée dans l'intervalle \mathcal{D} entre l'instant $t=0$ et le temps t . Soit $T(t)$ ce temps de résidence dont la courbe est présentée sur la figure III.27.

La fonction $T(t)$ est nulle à temps très court et augmente ensuite linéairement en fonction du temps à partir de $t=2$ ps. A temps long, on a atteint un régime stationnaire et le temps de résidence moyen des trajectoires dans l'intervalle \mathcal{D} est donc proportionnel au temps. La proportion de chromophores dont la coordonnée φ se trouve dans l'intervalle \mathcal{D} , $q(t)$, est donc constante. Elle correspond à la pente de la courbe $T(t)$. On a la relation

$$q(t) = \frac{dT(t)}{dt} \simeq q_0 = 0.9 \cdot 10^{-3}$$

Les informations que l'on extrait de la figure III.27 sont, d'une part, que les simulations atteignent un équilibre dans lequel la proportion de chromophores dans l'intervalle \mathcal{D} est constante et, d'autre part, que cet équilibre est atteint au bout d'environ 1 ps.

b) Calcul de la constante de vitesse et du rendement quantique de fluorescence

Compte tenu du fait que la fonction $q(t)$ est une constante, l'équation III.6 est une simple équation différentielle du premier ordre qui a pour solution

$$P(t) = P(0) \exp[-(k_r + k_{nr})t] \tag{III.14}$$

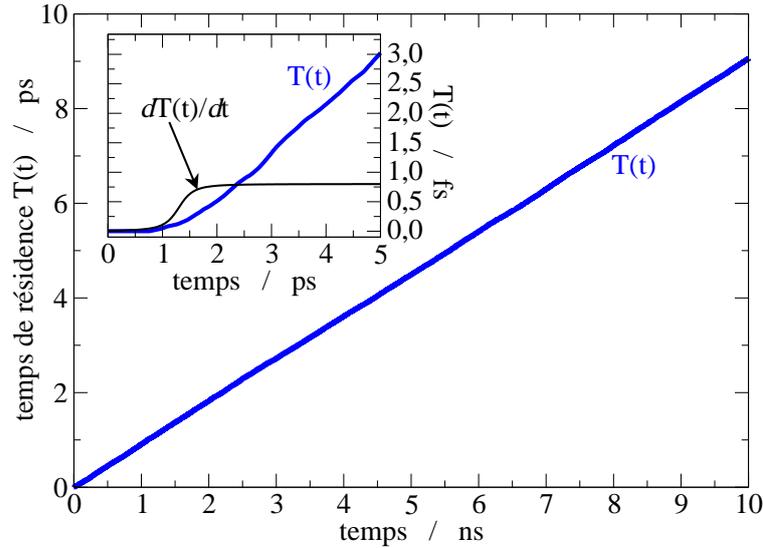


Figure III.27 – Temps de résidence, $T(t)$, de la coordonnée φ dans l'intervalle $[60^\circ ; 120^\circ]$ en fonction du temps (courbe bleu). Le cadre en haut à gauche est un agrandissement de la courbe $T(t)$ à temps court. La courbe noire représente la dérivée de $T(t)$ par rapport au temps.

où k_{nr} est la constante de vitesse de désactivation non radiative et vaut $k_{nr} = \mu q_0$. Les constantes de vitesse étant indépendantes du temps, l'expression du rendement quantique de fluorescence est simplement :

$$\phi_f = \frac{k_r}{k_r + k_{nr}} \quad (\text{III.15})$$

En utilisant la valeur de μ extraite de l'article de Gepshtein *et al* [101], on trouve

$$k_{nr} < 0.04 \text{ ns}^{-1} \quad (\text{III.16})$$

La valeur de k_{nr} obtenue est une borne supérieure compte tenu du fait que nous avons pris une borne supérieure de la valeur de μ . Connaissant le rendement quantique de la fluorescence de la GFP sauvage (0.8 [93]) et le temps de vie expérimental de la GFP (environ 3 ns [7]), on en déduit une valeur de la constante de vitesse radiative qui vaut $k_r = 0.27 \text{ ns}^{-1}$ (k_r est obtenue par l'équation I.4 page 13). La constante de vitesse associée au mécanisme de désactivation de la fluorescence par conversion interne est donc beaucoup plus faible que la constante de vitesse radiative. On peut donc en déduire que ce mécanisme contribue faiblement au déclin de la fluorescence.

En utilisant les valeurs des constantes de vitesse et l'équation III.15 on peut calculer une valeur du rendement quantique de fluorescence. On trouve $\phi_f = 0.87$. Le rendement quantique de fluorescence expérimental de la GFP que nous avons étudié est de 0.64 [117] (code 1EMA [18] de la *Protein Data Bank*). La contribution de ce mécanisme à la désactivation non radiative de la fluorescence est donc mineure.

III.4.3.2 Cas du mécanisme de désactivation de la fluorescence par une intersection conique

Dans le cas où l'on considère que le mécanisme de désactivation non radiatif de la fluorescence est le mécanisme par conversion interne, on doit résoudre l'équation III.7. Deux quantités sont inconnues : les distributions des temps de $n^{\text{ème}}$ passage $p_n(t)$ et la probabilité λ d'atteindre l'intersection conique à partir de la géométrie $TW\varphi$ et de retourner à l'état fondamental. Dans cette étude, la valeur de λ restera indéterminée et sera traitée comme un paramètre. L'évaluation de cette valeur est une question à par entière. Elle nécessiterait une connaissance plus précise de la surface d'énergie potentielle de l'état électronique excité du chromophore pour localiser correctement les intersections coniques. Il faudrait, d'autre part, évaluer la probabilité que le système retourne à l'état fondamental une fois qu'il a atteint l'intersection conique. Ceci peut se faire à l'aide de simulation de dynamique quantique [125, Chapitre 9], ou par un modèle plus simple de type Landau-Zener [126, p 164] qui nécessite cependant d'évaluer le couplage potentiel entre les états électroniques et sont généralement limités aux problèmes à 1 dimension.

a) Calcul des distributions des temps de $n^{\text{ème}}$ passages

Les distributions $p_n(t)$ sont obtenues directement à partir des simulation de dynamique Brownienne en construisant l'histogramme des temps de $n^{\text{ème}}$ passage à la géométrie $TW\varphi$ sur l'ensemble des trajectoires. Les distributions obtenues pour des valeurs de n allant de 1 à 8 sont présentées sur la figure III.28.

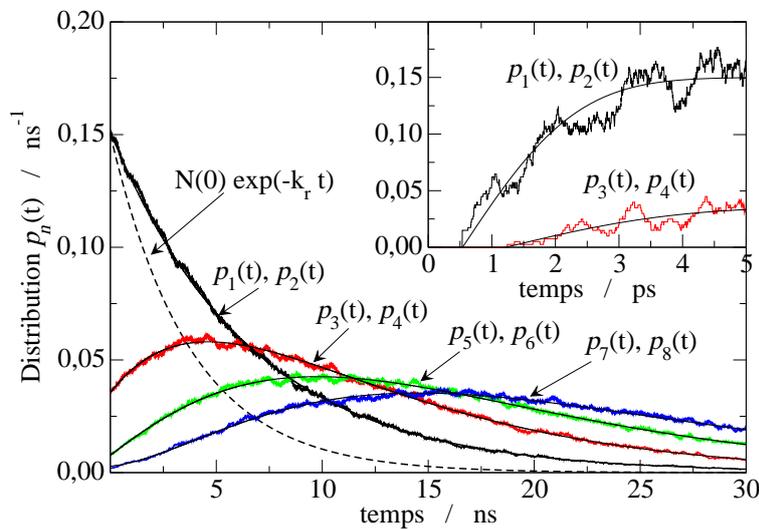


Figure III.28 – Distributions des temps de $n^{\text{ème}}$ passage obtenues par simulation de dynamique Brownienne. Le cadre en haut à droite est un agrandissement des distributions à temps courts. La courbe en pointillé, représente le déclin de la fluorescence dû à une désactivation purement radiative. Les courbes noires en traits pleins sont des ajustements par moindres carrés à l'aide d'une fonction exponentielle multipliée par un polynôme de degré inférieur ou égal à 2.

Les distributions $p_n(t)$ sont d'abord croissantes puis elles atteignent un maximum. Pour des temps supérieurs à quelques nanosecondes, elles diminuent exponentiellement. On remarque que les distri-

bution $p_{2n}(t)$ et $p_{2n+1}(t)$ ne sont pas différenciables. Ce comportement est dû au fait que pour des valeurs de $\varphi = \pm 90^\circ$ le profil d'énergie libre partielle ne présente pas un maximum mais au contraire une pente assez importante (figure III.16 page 129). Lorsqu'une trajectoire croise la géométrie TW φ dans le sens croissant du profil d'énergie libre partielle, elle recroise toujours la géométrie TW φ dans le sens décroissant du profil d'énergie libre dans un intervalle de temps court. Pour ces trajectoires, le temps de $2n^{\text{ème}}$ passage est donc très proche du temps de $2n+1^{\text{ème}}$ passage.

Dans les premières picosecondes, on remarque que les distributions $p_n(t)$ sont nulles puis augmentent rapidement pour atteindre une valeur maximale. Ce type de comportement est caractéristique des mouvements pour lesquels une barrière doit être franchie ou ceux pour lesquels la zone de l'espace des configurations dans laquelle le mécanisme peut avoir lieu ne se recouvrent pas avec la zone de l'espace des configurations associée à l'état initial. Ce deuxième cas est assez courant en photochimie. En effet, un grand nombre de mécanismes sont ultra-rapides du fait que la surface d'énergie potentielle ne présente pas de barrière. Cependant le processus ne démarrera pas instantanément après l'absorption d'un photon car le système doit atteindre la zone de l'espace des configurations dans laquelle le processus a lieu. C'est alors la dynamique du système qui limite la vitesse des processus.

Les distributions de $n^{\text{ème}}$ passage se croisent à temps long. A partir d'un certains temps, on peut considérer que la majorité des trajectoires ont atteint au moins n fois la géométrie TW φ . A partir de ce même temps la probabilité qu'une trajectoire atteigne la géométrie TW φ pour la $n+1^{\text{ème}}$ fois est donc plus importante que la probabilité qu'une trajectoire l'atteigne pour la $n^{\text{ème}}$ fois. Cela se traduit par le fait que pour chaque distribution $p_n(t)$ il existe un intervalle de temps dans lequel elle est au dessus des autres.

Sur la figure III.28, la courbe noire en pointillé représente le terme $N(0) \exp(-k_r t)$ qui intervient dans l'expression de la constante de désactivation non radiative (équation III.7 page 136). Pour des temps supérieur à 15 ns, ce terme prend des valeurs proches de 0. En conséquence, les distributions $p_n(t)$ pour n supérieur à 9, auront une faible contribution dans la somme sur les distributions de $n^{\text{ème}}$ passage de l'équation III.7. Ces distributions sont non nulles là où le terme $N(0) \exp(-k_r t)$ est nul et le produit des deux sera donc toujours nul.

b) Calcul du rendement quantique de la fluorescence

Pour calculer le rendement quantique de fluorescence de la GFP en incluant le mécanisme de désactivation non radiatif de la fluorescence par une intersection conique, on utilise l'équation I.3 page 13. Le rendement quantique de fluorescence ϕ_f est alors donné par l'expression :

$$\phi_f = \frac{1}{N(0)} k_r \int_0^\infty N(t) dt \quad (\text{III.17})$$

En remplaçant $N(t)$ par son expression dans l'équation III.7 on obtient

$$\begin{aligned}
 \phi_f &= \frac{1}{N(0)} \int_0^\infty \left[-\frac{dN(t)}{dt} - \sum_n \lambda^n p_n(t) N(0) \exp(-k_r t) \right] dt \\
 \phi_f &= 1 - \int_0^\infty \sum_n \lambda^n p_n(t) \exp(-k_r t) dt \\
 \phi_f &= 1 - \frac{1}{N(0)} \sum_n \lambda^n \int_0^\infty p_n(t) N(0) \exp(-k_r t) dt
 \end{aligned} \tag{III.18}$$

Les distributions $p_n(t)$ étant bruitées, il est préférable, pour calculer les intégrales de l'équation III.18, d'utiliser une expression analytique des distributions $p_n(t)$. Ces distributions atteignent rapidement leur valeur maximale, nous n'avons donc considéré que leur forme à temps long.

Les distributions des temps de n^{ème} passage de la figure III.28 peuvent être représentées par une fonction exponentielle multipliée par un polynôme de degré inférieur ou égal à 2. Nous avons ajusté les paramètres de ces fonctions pour les distributions $p_n(t)$ avec n inférieur ou égal à 8 par une méthode de moindre carré. Les expressions utilisées et les valeurs des paramètres pour chaque distribution sont rassemblées dans le tableau III.4. Ces courbes sont tracées en trait plein noir sur la figure III.28.

$p_n(t)$	Forme analytique et paramètres associés
$p_1(t), p_2(t)$	$a_1 \exp(-t/\tau_1)$ $\tau_1 = 6.60 \text{ ns}; a_1 = 1.5 \cdot 10^{-1} \text{ ns}^{-1}$
$p_3(t), p_4(t)$	$(a_3 t + b_3) \exp(-t/\tau_3)$ $\tau_3 = 6.59 \text{ ns}; a_3 = 1.8 \cdot 10^{-2} \text{ ns}^{-2}; b_3 = 3.5 \cdot 10^{-2} \text{ ns}^{-1}$
$p_5(t), p_6(t)$	$(a_5 t^2 + b_5 t + c_5) \exp(-t/\tau_5)$ $\tau_5 = 6.59 \text{ ns}; a_5 = 1.0 \cdot 10^{-3} \text{ ns}^{-3}; b_5 = 8.4 \cdot 10^{-3} \text{ ns}^{-2}; c_5 = 7.7 \cdot 10^{-3} \text{ ns}^{-1}$
$p_7(t), p_8(t)$	$(a_7 t^2 + b_7 t + c_7) \exp(-t/\tau_7)$ $\tau_7 = 7.69 \text{ ns}; a_7 = 1.0 \cdot 10^{-4} \text{ ns}^{-3}; b_7 = 1.4 \cdot 10^{-3} \text{ ns}^{-2}; c_7 = 2.8 \cdot 10^{-3} \text{ ns}^{-1}$

Tableau III.4 – Expressions analytiques employées pour modéliser les distributions des temps de n^{ème} passage et les valeurs des paramètres de ces expressions obtenues par une méthode de moindres carrés.

A partir de l'équation III.18 et de la valeur des intégrales calculées analytiquement avec les expressions du tableau III.4, on peut calculer des valeurs de ϕ_f en fonction de λ . Nous avons tout d'abord cherché à déterminer la convergence de ϕ_f en fonction de n , le nombre de distribution de n^{ème} passage que l'on considère.

La figure III.29 présente la convergence en fonction de n du rendement quantique de la fluorescence ϕ_f pour une valeur de λ égale à $1/2$. On remarque que les distributions $p_n(t)$ pour n supérieur à 6 ne jouent pas un rôle majeur dans la valeur finale du rendement quantique de fluorescence. Cette

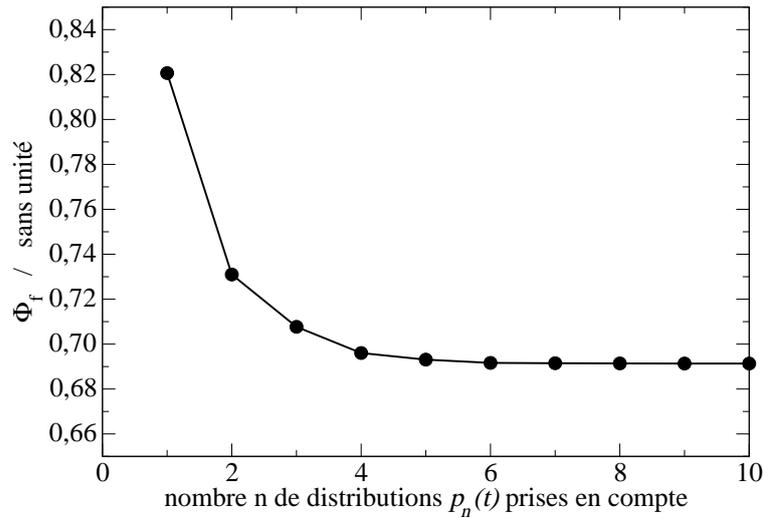


Figure III.29 – Convergence de la valeur du rendement quantique de fluorescence ϕ_f en fonction du nombre de distribution de $n^{\text{ème}}$ passage que l'on considère pour une valeur de λ de $1/2$.

observation est en accord avec la forme des distributions $p_n(t)$ par rapport au terme $N(0) \exp(-k_r t)$ (figure III.28). Pour des valeurs de n élevées, le produit de ce terme par la fonction $p_n(t)$ est nul et la contribution de la distribution $p_n(t)$ est donc négligeable. On peut donc en déduire qu'en moyenne le chromophore dans son état électronique excité atteindra la géométrie $TW\varphi$ au plus 6 fois avant de revenir à l'état fondamental si λ est de l'ordre de $1/2$.

La figure III.30 présente les valeurs du rendement quantique de la fluorescence, ϕ_f , calculées avec l'équation III.18 et en fonction de la valeur de λ . On remarque que pour $\lambda=0$, le rendement quantique de fluorescence vaut 1 et que plus λ augmente, plus ϕ_f diminue. Ce comportement est en accord avec le modèle utilisé. Si $\lambda=0$, le seul processus de désactivation de l'état excité est la fluorescence. Le nombre de photon émis est donc égal au nombre de photons absorbés et le rendement quantique de fluorescence vaut 1. Par contre, plus λ augmente, plus le processus non radiatif est efficace ce qui fait chuter le rendement quantique de fluorescence.

La valeur moyenne du rendement quantique sur les valeurs possibles de λ est du même ordre de grandeur que les valeurs expérimentales. Cette valeur moyenne peut être vue comme la valeur de ϕ_f quelque soit la valeur de λ , en supposant que toutes les valeurs de λ sont équiprobables. Le modèle utilisé est donc capable de donner un ordre de grandeur raisonnable du rendement quantique de fluorescence bien que λ soit inconnu. Ce résultat est un grand avantage compte tenu des difficultés liées au calcul de λ pour des systèmes de la taille du chromophore de la GFP.

c) 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 (équation I.2 page 12). Pour calculer le déclin de la fluorescence, il faut résoudre l'équation différentielle III.7. C'est une équation du premier ordre avec

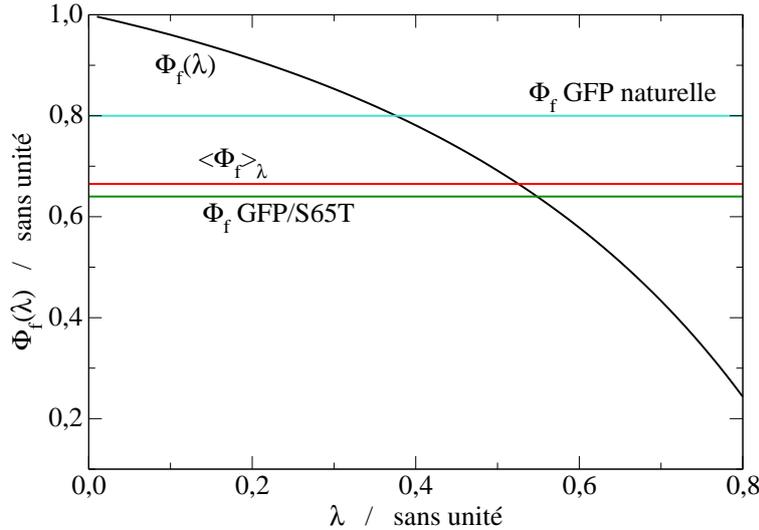


Figure III.30 – Valeurs du rendement quantique ϕ_f obtenues en fonction de la probabilité λ . La courbe rouge représente la moyenne des valeurs de ϕ_f sur les valeurs de λ . La courbe verte représente la valeur de ϕ_f dans le mutant de la GFP que nous avons étudié (code 1EMA [18] de la *Protein Data Bank*). La courbe bleu représente la valeur de ϕ_f dans la GFP sauvage.

un second membre dépendant du temps. La solution de cette équation est

$$N(t) = N(0) \exp(-k_r t) \left[1 - \sum_n \lambda^n \int_0^t p_n(t') dt' \right] \quad (\text{III.19})$$

La fonction $N(t)$ peut être facilement évaluée quelque soit t en utilisant les expressions analytiques du tableau III.4 et nous donne le déclin de la population de l'état excité en fonction du temps. Nous avons calculé le déclin pour plusieurs valeurs de λ en incluant les distributions $p_n(t)$ jusqu'à n égale 6. Les courbes obtenues sont présentées sur la figure III.31 a, en échelle logarithmique.

Les déclins obtenus sont linéaires sur une échelle logarithmique et présentent donc un comportement mono-exponentiel. La pente de ces déclins de la fluorescence est inversement proportionnelle au temps de vie de l'état excité qui inclut la fluorescence et la désactivation non radiative par conversion interne. On trouve un temps de vie de 3.3 ns pour une valeur de λ égale à 0.25 et 2.6 ns pour une valeur de λ égale à 0.5. Ces deux valeurs encadrent la valeur expérimentale du temps de vie pour la GFP sauvage qui est de 3 ns et sont cohérentes avec les valeurs du rendement quantique de la fluorescence. En effet, les valeurs de ϕ_f obtenues pour ces deux valeurs de λ encadrent la valeur de ϕ_f de la GFP sauvage. Ce type de modèle nous donne donc des résultats en bon accord avec les observations expérimentales sur la GFP. Le fait que le déclin de fluorescence soit monoexponentiel peut paraître surprenant compte tenu de l'expression de la constante de vitesse de désactivation non radiative que nous avons employées. La figure III.31 b, montre en fonction du temps l'évolution du terme entre crochet dans l'équation III.19. Ce terme présente des variations proches d'une fonction exponentielle sauf pour pour des valeurs de λ proches de 0.4. Le produit de deux exponentielles étant toujours une fonction exponentielle, on obtient donc des déclins de la population de l'état excité qui sont approximativement mono-exponentiels.

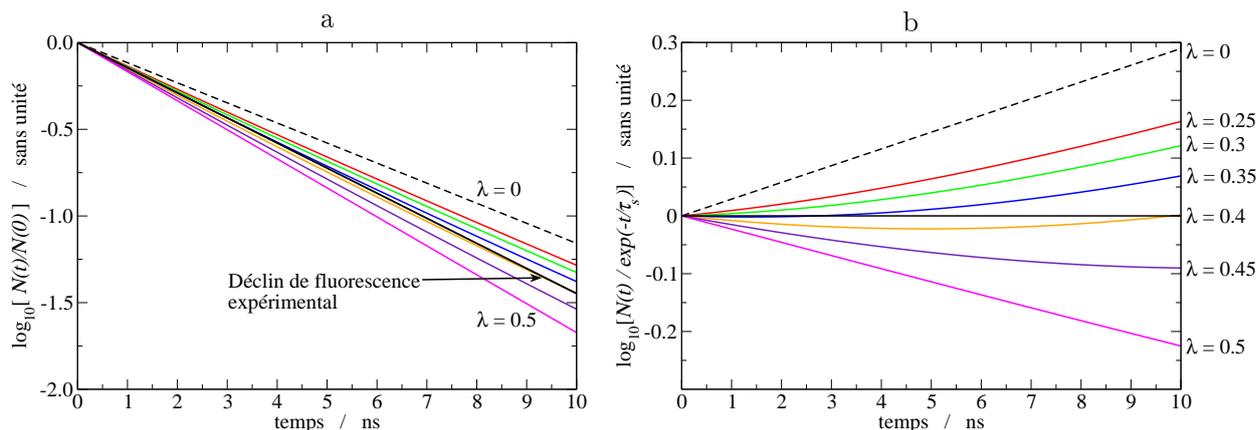


Figure III.31 – a : Déclin de la population de l'état électronique excité obtenu à partir du logarithme décimal de l'équation III.19 pour plusieurs valeurs de λ ($\lambda=0$ et λ variant de 0.25 à 0.5 par pas de 0.05). b : Rapport du déclin de la population de l'état excité obtenu à partir de l'équation III.19 au déclin de la population de l'état excité expérimental qui est monoexponentiel et dont le temps de vie est $\tau_s=3$ ns, pour plusieurs valeurs de λ . Sur les deux figures, les couleurs correspondent aux mêmes valeurs de λ , la courbe noire en pointillé est relative au déclin de fluorescence en absence de processus de désactivation et la courbe noire en trait plein est relative au déclin de fluorescence expérimental.

Sans information supplémentaire qui permettrait de préciser la valeur de λ , on ne peut pas conclure précisément sur l'efficacité du mécanisme de désactivation de la fluorescence par une intersection conique au voisinage de la géométrie $TW\varphi$. Cependant, la valeur moyenne du rendement quantique de fluorescence sur l'ensemble des valeurs de λ est du même ordre de grandeur que le rendement quantique de fluorescence expérimental. De plus, pour des valeurs de λ de l'ordre de 0.4, le temps de vie de l'état excité obtenu à partir de notre modèle est proche de celui mesuré expérimentalement pour la GFP sauvage. Il semble donc que si la probabilité λ est de cette ordre de grandeur, ce mécanisme aurait une efficacité non négligeable et il pourrait contribuer au déclin non radiatif de la population de chromophores à l'état électronique excité.

III.5 Conclusions et perspectives

Dans ce chapitre, nous avons présenté une approche, combinant des simulations de dynamique moléculaire biaisées et des simulations de dynamique brownienne, et, permettant d'étudier la dynamique du chromophore dans son premier état électronique excité à l'intérieur de la protéine. Un changement d'échelle est nécessaire pour étudier les processus photophysiques associées à cette dynamique et à une échelle de temps de quelques nanosecondes. L'association de ces deux techniques de simulation nous a permis d'étudier la dynamique d'une coordonnée du système, identifiée comme étant la coordonnée de réaction, sur une grande échelle de temps par rapport aux simulations tout atome. Dans ce travail nous avons réalisé un million de trajectoires de la coordonnée φ de 50 ns chacune, ce qui correspondrait à une simulation d'une milliseconde. Ce changement d'échelle est possible grâce à l'utilisation de la dynamique brownienne mais nécessite de décrire le système de manière plus

phénoménologique.

Nous avons utilisé cette approche pour étudier le mécanisme de la désactivation de la fluorescence de la GFP par une déformation du chromophore en fonction de la coordonnée de torsion φ . Nous nous sommes appuyés sur les résultats de Martin *et al* (figure III.2 page 109) qui suggèrent qu'une désactivation de la fluorescence par conversion interne est probable pour des valeurs de φ dans l'intervalle $\mathcal{D}=[60^\circ ; 120^\circ]$ et que le chromophore peut retourner à l'état fondamental grâce à un intersection conique accessible à partir de la géométrie perpendiculaire $TW\varphi$. Bien que le potentiel intra-moléculaire du chromophore dans le vide en fonction φ montre que la géométrie la plus stable est la géométrie $TW\varphi$, le profil d'énergie libre partielle en fonction de φ que nous avons calculé par la méthode du *umbrella sampling* montre que dans la protéine, cette géométrie se situe plusieurs $\text{kcal}\cdot\text{mol}^{-1}$ au dessus de la géométrie plane fluorescente. Nous avons identifié les acides aminés qui contribuent le plus à ce que le chromophore reste dans une géométrie plane. Il est apparu qu'il s'agit principalement de deux thréonine, Thr61 et Thr203, ainsi qu'une chaîne de molécules d'eau structurales situées au voisinage du chromophore. Le calcul du profil d'énergie libre partielle à l'intérieur de la protéine est cependant compliqué car il existe un grand nombre de minima locaux et de barrières qui rendent difficile l'échantillonnage de la surface d'énergie potentielle.

Les simulations de dynamique brownienne ont nécessité de déterminer le coefficient de diffusion rotationnelle du cycle phénolate à l'intérieur de la protéine. Nous l'avons évalué en calculant la vitesse limite de la coordonnée φ sur des trajectoires de la relaxation du système à partir de géométries où φ prend des valeurs proches de 40° . Nous avons extrait des trajectoires de dynamique brownienne les données dont nous avons besoin pour calculer la constantes de vitesse du mécanismes de désactivation. Dans le cas de la désactivation par une conversion interne dans l'intervalle \mathcal{D} il s'agit du temps de résidence dans cet intervalle. Dans le cas de la désactivation de la fluorescence par une intersection conique il s'agit des temps de $n^{\text{ème}}$ passage à la géométrie $TW\varphi$. Nous avons alors calculé le rendement quantique de fluorescence de la GFP en présence de l'un ou l'autre de ces deux processus. Il en résulte que la désactivation par conversion interne semble ne contribuer que faiblement au déclin de la fluorescence. La désactivation de la fluorescence par une intersection conique accessible à partir de la géométrie $TW\varphi$ pourrait être plus efficace mais il faudrait préciser la valeur de λ pour pouvoir conclure de manière plus franche.

Outre la localisation plus précise des intersections coniques sur la surface d'énergie potentielle de l'état électronique excité du chromophore et la probabilité de retourner à l'état fondamental dans ces géométries (que nous avons modélisé par le paramètre λ) plusieurs améliorations ou alternatives peuvent être envisagées. Tout d'abord, le champ de force que nous avons utilisé pourrait être amélioré d'une part par un traitement simultané et couplé des coordonnées φ et τ et, d'autre part, en utilisant des charges effectives qui s'adaptent à la géométrie. Un modèle prenant en compte plusieurs formes de Lewis [127] dont le poids varie en fonction de la géométrie et de l'environnement pourrait être envisagé. D'autre part, on peut se demander quel est le bon potentiel à utiliser dans le cas de la dynamique brownienne. En effet, on a vu que le mouvement de la coordonnée φ est fortement corrélé à la position de molécules d'eau ou d'acides aminés proches du chromophore qui font obstacle à la torsion φ . Une méthode permettant de traiter ces obstacles de façon explicite dans la dynamique brownienne, en conservant une seule coordonnée, pourrait être envisagée.

Conclusion générale

Nous avons présenté l'étude de propriétés photophysiques de protéines fluorescentes au moyen de différents outils théoriques (chimie quantique, simulation de dynamique moléculaire, simulation de dynamique brownienne, modèle cinétique et analyse statistique), adaptés au mieux, aux problèmes traités. Ce travail s'inscrit dans un axe transversal du laboratoire de chimie physique, qui réunit des expérimentateurs et des théoriciens dans le but d'étudier la relation entre la structure, la dynamique et les propriétés photophysiques des protéines fluorescentes. D'un point de vue théorique, la mise en oeuvre et l'élaboration de méthodes permettant d'étudier ce type de système est une question à part entière et a motivé une partie de cette thèse.

Dans un premier temps, nous avons étudié par simulation moléculaire, l'effet du changement de pH sur la structure de la Cerulean et son spectre d'absorption. Nous avons fait des simulations à deux pH différents et pour deux conformations différentes du chromophore. Ces calculs nous ont permis d'établir que le décalage du spectre d'absorption vers le bleu, observé expérimentalement par Malo *et al* lorsque le pH diminue de pH 8 à pH 5, est la conséquence d'une isomérisation du chromophore. Nous avons aussi montré que cette isomérisation est liée au changement de l'orientation de la chaîne latérale de l'acide aminé en position 148.

Ces résultats ouvrent deux types de perspectives. Tout d'abord, dans l'objectif d'améliorer les propriétés photophysiques des mutants de la CFP, il serait intéressant d'approfondir expérimentalement et théoriquement le rôle de l'acide aminé en position 148 et son effet sur l'isomérisation du chromophore. Par ailleurs, il serait intéressant pour ce type d'étude de mettre au point des méthodes permettant de prendre en compte la distribution des micro-états de protonation et pas seulement le micro-état de plus basse énergie libre.

La deuxième partie de cette thèse concerne l'étude d'un mécanisme de désactivation de la fluorescence. Nous avons proposé une approche, combinant des simulations de dynamique moléculaire biaisée et des simulations de dynamique brownienne, afin de déterminer la cinétique d'un mécanisme de désactivation non radiative de l'état électronique excité, lié à une torsion du chromophore. Cette approche permet de faire un changement d'échelle de temps. Nous avons pu ainsi obtenir des distributions de temps de $n^{\text{ème}}$ passage aux géométries critiques et en déduire quantitativement le déclin de fluorescence (rendement quantique et temps de vie). Les résultats obtenus sont directement comparables aux résultats expérimentaux.

La poursuite de ce deuxième sujet nécessite de développer des méthodes permettant de décrire simultanément l'état fondamental et l'état excité du chromophore quelque soit sa géométrie, par

exemple à l'aide d'un champ de force matriciel. Il est également nécessaire de développer des méthodes dynamiques permettant de prendre en compte, à temps long, la présence d'obstacle le long de la coordonnée étudiée (voir le rôle des thréonine vis à vis de la torsion du chromophore). Les outils développés dans cette thèse et leurs futurs développements permettront de progresser dans la compréhension de la relation entre l'isomérisation du chromophore, le pH et le déclin de la fluorescence qui sont étroitement liés dans les protéines fluorescentes.

Annexe A

Structure des protéines

Les protéines de tous les êtres vivants sont des chaînes qui ont en commun avec les polymères d'être constituées d'un petit nombre d'éléments qui se répètent, les acides aminés. Contrairement aux polymères, les protéines ne sont pas périodiques. C'est la diversité d'enchaînement de ces briques de bases que sont les acides aminés qui conduit à l'ensemble des protéines du vivant et à leur diversité. L'agencement dans l'espace des acides aminés détermine la structure tridimensionnelle de la protéine. Cet agencement est guidé par les propriétés physico-chimiques des acides aminés qui composent la protéine, par leur ordre dans la séquence, ainsi que par l'environnement. Il est le résultat de la minimisation de l'énergie libre du système contrôlée par l'optimisation des interactions hydrophiles et hydrophobes entre les acides aminés entre eux et avec le milieu environnant. Au sein de la structure protéique, on peut distinguer plusieurs niveaux de structuration.

A.1 Les acides aminés et la structure primaire

La chaîne primaire est le premier niveau de structuration. Elle correspond à la séquence des acides aminés. Les acides aminés que l'on rencontre dans les protéines que nous avons étudiées sont au nombre de 20 (tableau A.1). Excepté la proline, ils présentent tous une structure similaire. Ils sont constitués d'une fonction acide carboxylique et d'une fonction amine reliées entre elles par un carbone aliphatique, appelé carbone *alpha*, C_α , qui est un carbone asymétrique (à l'exception de la glycine). Les acides aminés sont donc des molécules chirales. Ce C_α est substitué par un hydrogène et une chaîne que l'on appelle chaîne latérale (figure A.1). C'est cette chaîne latérale qui donne son identité à chacun des acides aminés.

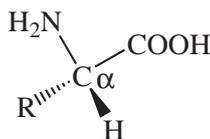


Figure A.1 – Représentation de Cramer d'un acide aminé. Le groupement R représente la chaîne latérale.

Dans le cadre du travail réalisé dans cette thèse, il est utile de distinguer deux types d'acides aminés :

- ceux dont la chaîne latérale porte une fonction acido-basique (acide carboxylique alcool, thiol, amine ou amide) peuvent échanger des protons avec le solvant et donc être sous une forme protonée ou non protonée suivant le pH. Il s'agit des acides aminés "à chaîne latérale polaire" du tableau A.1.
- ceux qui ne portent pas de telles fonctions et qui n'échangent pas de proton. Il s'agit des acides aminés "à chaîne latérale non polaire" du tableau A.1.

On définit l'état de protonation standard d'un acide aminé portant une fonction acido-basique comme son état de protonation lorsqu'il est isolé (ne fait pas partie d'une protéine) en phase aqueuse à pH 7. Dans leur état standard, l'acide aspartique et l'acide glutamique sont sous leur forme basique, non protoné et anionique. La tyrosine et la cystéine sont sous leur forme acide, protoné et neutre. L'arginine et la lysine sont sous leur forme acide, protonée et cationique. L'histidine est un cas particulier, à pH 7 elle sous sa forme basique, neutre. Un seul des atomes d'azote du cycle imidazole porte un proton et il y a donc deux possibilités pour la position du proton (δ ou ϵ).

C'est la présence de la fonction acide carboxylique et de la fonction amine portées par le $C\alpha$ qui donne aux acides aminés la capacité de former une chaîne. L'acide carboxylique d'un acide aminé peut s'additionner sur la fonction amine d'un autre acide aminé pour former une fonction amide qui relie les deux acides aminés (figure A.2). Cette liaison s'appelle une liaison peptidique. L'enchaînement ordonné des acides aminés de l'atome N-terminal (défini comme l'atome d'azote du groupement amine du premier acide aminé) à l'atome C-terminal (défini comme l'atome de carbone du groupement COOH du dernier acide aminé) forme la séquence primaire de la protéine. L'ensemble des atomes constituant les liaisons peptidiques entre les acides aminés est appelé squelette protéique.

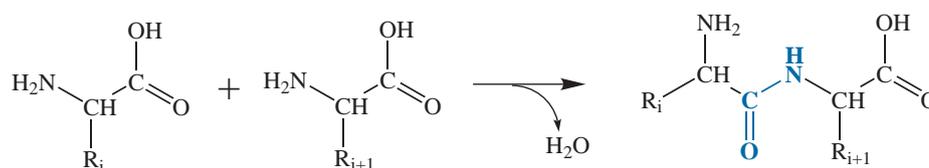


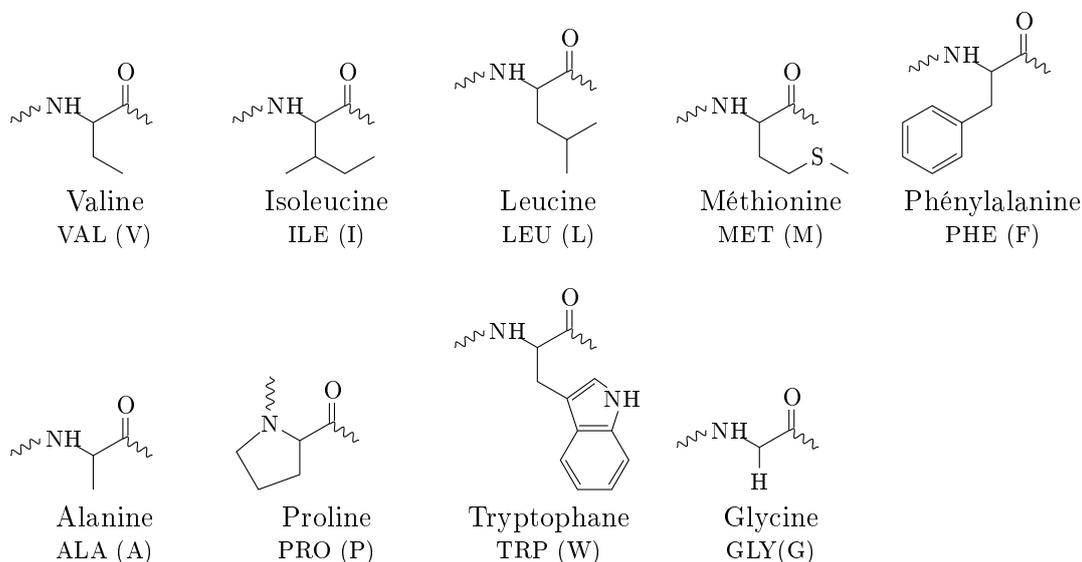
Figure A.2 – Formation de la liaison peptidique (en bleu) entre deux acides aminés.

A.2 Structures secondaires

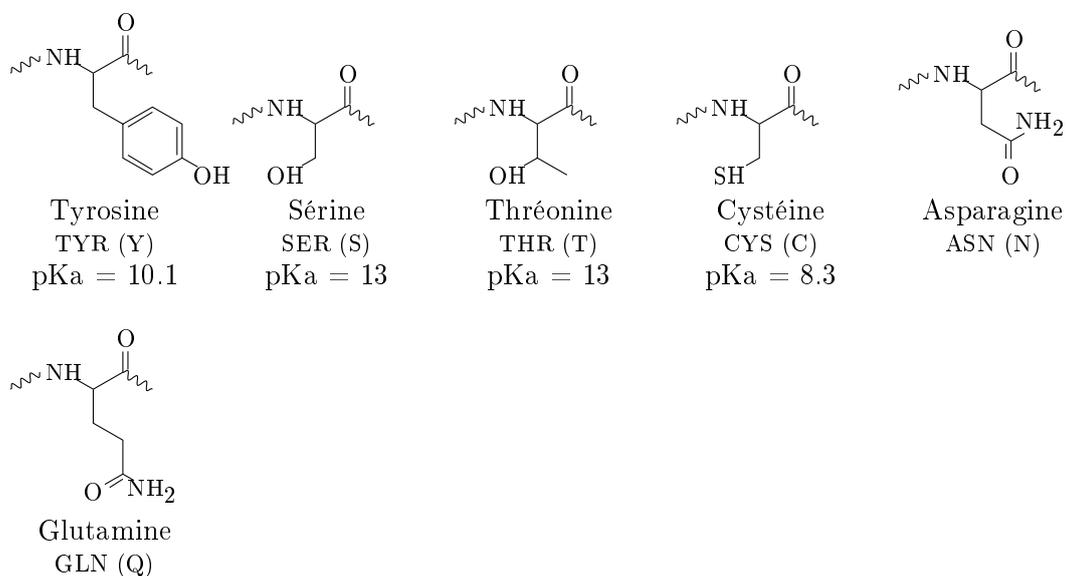
Le deuxième niveau de structuration de la structure des protéines correspond aux structures secondaires. Il s'agit de structures tridimensionnelles locales formées par l'agencement d'acides aminés et contrôlées par la formation de réseaux de liaison hydrogène. Une courte séquence d'acides aminés est appelée un brin peptidique. Les conformations adoptées par un brin sont contrôlées par les angles dièdres ϕ , ψ et ω représentés sur la figure A.3. Les valeurs prises par ces angles sont soumises à certaines contraintes. Du fait de la conjugaison entre le groupement carbonyle et le doublet non liant du NH du squelette protéique, l'angle ω prend uniquement des valeurs proches de 0° ou 180° .

Tableau A.1 – Tableau regroupant les vingt acides aminés principaux ainsi que leur nom, les codes à une lettre et à trois lettres utilisé pour simplifier l'écriture et, s'il y a lieu, le pKa de la chaîne latérale.

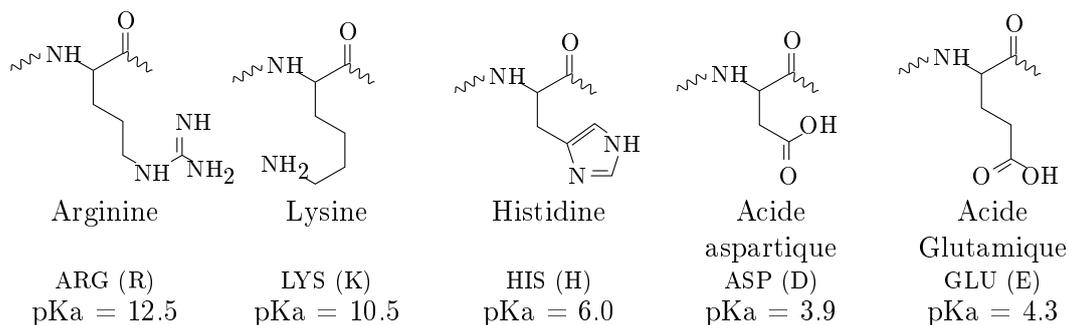
Acides aminés à chaîne latérale non polaire



Acides aminés à chaîne latérale polaire non chargée dans l'état standard



Acides aminés à chaîne latérale polaire chargée dans l'état standard



La gamme des valeurs de ϕ et ψ est plus étendue, mais les valeurs de ces angles sont corrélées. Le diagramme de Ramachandran (figure A.3) représente la densité des couples (ϕ, ψ) construite sur un grand nombre de protéine. On voit que seules certains couples de valeurs (ϕ, ψ) sont effectivement rencontrés. Les deux régions les plus denses (aux alentours de $\phi=120^\circ$, $\psi=-100^\circ$ et aux alentours de $\phi=-60^\circ$, $\psi=-100^\circ$) correspondent à des structures bien déterminées.

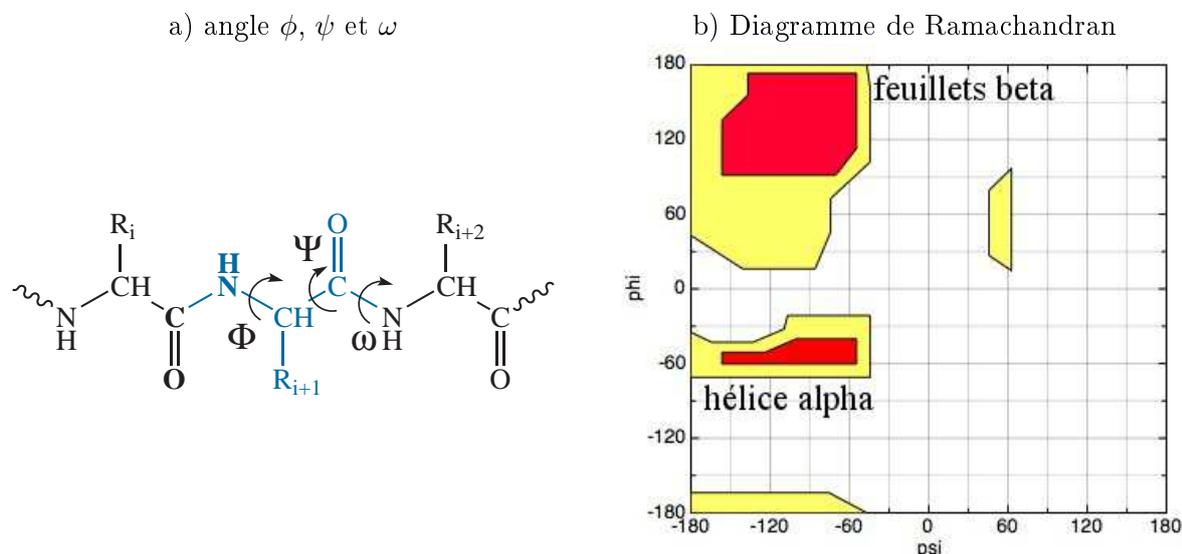


Figure A.3 – Géométrie d'un brin peptidique. a) Définitions des angles dièdres ϕ , ψ et ω du squelette protéique. b) Diagramme de Ramachandran présentant les couples (ϕ, ψ) accessibles (en jaune). Les régions rouges correspondent aux conformations les plus stables.

La région centrée autour des couples ($90^\circ \leq \phi \leq 170^\circ$, $-150^\circ \leq \psi \leq -60^\circ$) correspond à la structure dite en "feuillet β ". Un feuillet β se crée par formation de liaisons hydrogène entre les atomes du squelette protéique de deux brins peptidiques. Il peut se former de façon parallèle ou antiparallèle (figure A.4) suivant le sens des deux brins qui, par convention, va du N-terminal au C-terminal. Dans les deux cas, un feuillet peut être formé par une succession de brins reliés deux à deux. Dans les protéines que nous avons étudiées ce type de structure secondaire délimite l'intérieur et l'extérieur de la protéine. Le long de brins formant un feuillet β , on observe une alternance d'acides aminés polaires et non polaires dont la chaîne latérale s'oriente respectivement vers l'extérieur et vers l'intérieur de la protéine.

La région du diagramme de Ramachandran centrée autour des valeurs ($40^\circ \leq \phi \leq 60^\circ$, $-150^\circ \leq \psi \leq -50^\circ$) correspond à la structure secondaire dite en "hélice α " (figure A.4). Cette structure se crée par la formation de liaisons hydrogène entre l'atome d'oxygène du groupement carbonyle de l'acide aminé i et le groupement NH de l'acide aminé $i + 4$. Il en résulte qu'un tour d'hélice contient 3.6 acides aminés et que son pas est proche de 5.4 Å. L'alignement le long d'une hélice α des groupements carbonyle du squelette protéique fait apparaître un moment dipolaire important colinéaire à l'axe de l'hélice. D'autres types de structure en hélice existent. Elles diffèrent de l'hélice α par le réseau de liaisons hydrogène formé. Le tour d'hélice et le pas sont alors différents. Le reste des acides aminés forment soit des feuillets β ou des hélices α avec des défauts, soit des structures moins organisées (pelote statistique ou *random coil*).

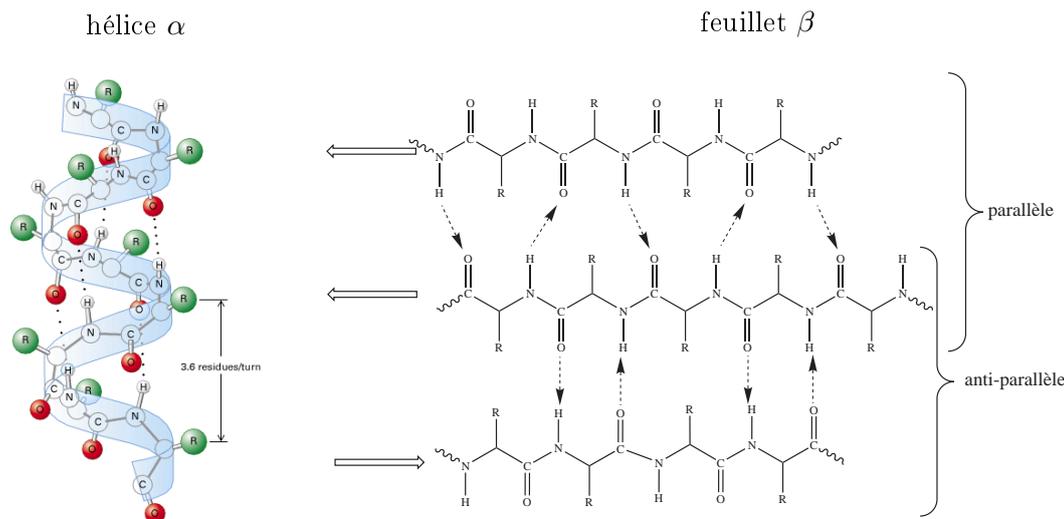


Figure A.4 – Représentation des structures secondaires en hélice α (à gauche) et en feuillet β (à droite). Les deux brins supérieurs forment un feuillet parallèle, les deux brins inférieurs forment un feuillet antiparallèle. L'image de gauche est issue du site www.chusa.jussieu.fr.

A.3 Structure tertiaire et structure quaternaire

L'agencement spatial relatif des différentes structures secondaires de la protéine constitue la structure tertiaire de la protéine. Cette structure se forme par l'intermédiaire des interactions entre les chaînes latérales des acides aminés. Il peut s'agir de la formation de liaisons hydrogène, de ponts salins (entre une chaîne latérale cationique et une chaîne latérale anionique) ou de ponts disulfures entre deux acides aminés contenant un atome de soufre. Le repliement de la structure tertiaire peut également être induit par les interactions hydrophobes qui correspondent à l'enfouissement des chaînes latérales hydrophobes afin de minimiser la perte d'entropie liée à leur désolvatation, ou par un alignement de longues hélices α qui est stabilisé par l'interaction entre les dipôles électrostatiques de ces structures.

Il arrive que la structure finale de la protéine soit obtenue par une association de plusieurs chaînes peptidiques, appelées alors monomères ou sous-unités, qui s'agencent pour former un complexe macromoléculaire. La structure obtenue par l'agencement de sous-unités est appelée structure quaternaire. On peut citer par exemple la protéine chaperonne GroEL qui assiste les protéines lors de leur repliement et qui est constituée de deux cylindres, formés chacun par un ensemble de 14 sous-unités.

D'autres protéines comme les cytochromes nécessitent un cofacteur, c'est à dire une molécule ou une macromolécule présente dans le milieu cellulaire, qui s'insère dans la structure tertiaire et confère à la protéine des propriétés spécifiques. Pour les cytochromes il s'agit d'une porphyrine complexée par un métal qui confère à la protéine des propriétés d'oxydo-réduction.

Annexe B

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. Le cas des électrons est considéré en annexe D.

B.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.

B.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 $\mathcal{U}(q^{3N})$ 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 B.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 B.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.

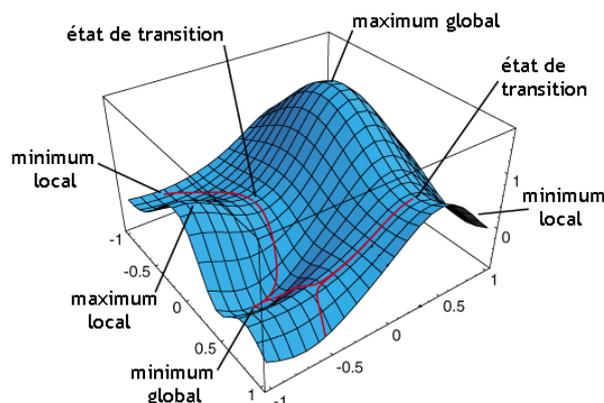


Figure B.1 – Schéma d’une surface d’énergie potentielle simple en deux dimensions. Sur la figure sont indiqués les points spécifiques de la surface d’énergie potentielle. Ils correspondent aux points pour lesquels les dérivées de la surface d’énergie potentielle en fonction des deux coordonnées sont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

B.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.

B.1.3 Modèle “tout atome” et coordonnées internes

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 B.2).

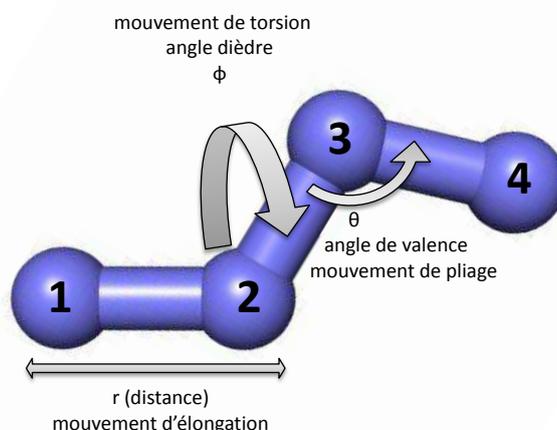


Figure B.2 – Définition des coordonnées internes d'une chaîne d'atomes. L'élongation est décrite par la distance r entre les atomes 1 et 2. Le pliage est décrit par l'angle de valence θ entre les atomes 2, 3 et 4. La torsion est mesurée par un angle dièdre ϕ , défini comme l'angle entre les plans passant 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.

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* [3, 4]. L'expression analytique de l'énergie potentielle, \mathcal{U}_{ff99} , associée à ce champ de force est la suivante :

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

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.

- **Energies de liaison et de pliage**

Les deux premiers termes représentent l'énergie potentielle en fonction de l'élongation d'une liaison et du pliage de deux liaisons (figure B.2) décrites respectivement par les coordonnées r et θ . Ce sont des termes harmoniques de constantes de force k_r et k_θ et dont la position d'équilibre est donnée par r_0 et θ_0 . Ces paramètres ont été optimisés de sorte que les simulations de petits peptides reproduisent au mieux des données de spectroscopie infrarouge [3].

- **Energie de torsion**

Le troisième terme est l'énergie potentielle en fonction d'un angle dièdre qui représente les mouvements de torsion autour d'une liaison. Ce terme est périodique et modélisé par une série de Fourier. La phase, ϕ_0 , la périodicité n et k_n qui est le nombre d'angles dièdres qui varie au cours de la torsion, dépendent de l'hybridation des deux atomes engagés dans la liaison. Ces paramètres ont été déterminés sur un ensemble de molécules modèles, les plus simples possibles, permettant de modéliser tous les types de torsion [3].

- **Interactions de van der Waals**

Le quatrième terme représente les interactions entre atomes non liés de type van der Waals. Des paramètres Lennard-Jones, σ_i et ε_i , ont tout d'abord été optimisés pour chaque type d'atome afin que les simulations reproduisent au mieux des données thermodynamiques expérimentales (comme les enthalpies de vaporisation). Ensuite pour une paire d'atomes i et j , les paramètres Lennard-Jones des atomes i et j individuels sont combinés en utilisant les règles de Lorentz-Berthelot [3].

$$\begin{aligned}\varepsilon_{ij} &= \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \\ \sigma_{ij} &= \frac{\sigma_{ii} + \sigma_{jj}}{2}\end{aligned}$$

- **Interactions électrostatiques**

Les interactions électrostatiques entre les atomes i et j sont modélisées par un potentiel coulombien. Les charges effectives q_i sont les seuls paramètres de ce champ de force qui ne sont pas définis par le type l'atome mais séparément pour chaque atome de la molécule. Elles sont ajustées par la méthode RESP, de sorte qu'elles reproduisent au mieux le potentiel électrostatique créé par la molécule calculé au niveau HF/6-31G* sur une grille de point. Les auteurs du champ de force de AMBER ont largement contribué au développement de la méthode RESP [43, 44] pour ajuster des charges effectives.

Modèles de molécule d'eau

L'eau constitue le solvant dans lequel évoluent la plupart des systèmes biologiques. Les interactions entre l'eau et la protéine sont très importantes car elles stabilisent la structure de la protéine. Le modèle utilisé pour les molécules d'eau est donc tout aussi important que le champ de force utilisé pour la protéine.

Un grand nombre de modèle d'eau existe, celui que nous avons utilisé est le modèle TIP3P [128] qui est celui avec lequel le champ de force *ff99* a été optimisé [3]. Il est composé de trois centres atomiques qui portent chacun une charge effective. La charge effective de l'oxygène est de -0.834 électrons, celle des hydrogènes est de 0.417 électrons et les atomes d'hydrogène présentent un rayon et une énergie de van der Waals nuls. C'est un modèle semi-rigide, dans lequel la seule coordonnée de déformation est la distance entre les deux atomes d'hydrogène. Les paramètres de ce modèle ont été optimisés de sorte qu'il reproduise les propriétés de l'eau liquide à température ambiante et pression atmosphérique.

B.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.

B.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

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

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, $\mathcal{U}(q^{3N})$, et donc de l'expression analytique du champ de force selon :

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

Les trajectoires sont obtenues par intégration numérique des équations du mouvement par différence finie. Etant donné un pas de temps δt 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 δt petit) et un temps de calcul limité (utilisation de grandes valeurs de δt). Dans la suite de programme AMBER, avec lequel les simulations de dynamique moléculaire ont été faites, l'algorithme utilisé est celui dit "Leap-Frog" [121, p 80]. 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 δt est un paramètre important d'une simulation de dynamique moléculaire. On considère le plus souvent que δt 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 [129] 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 B.2.3.

B.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 important 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 B.3).

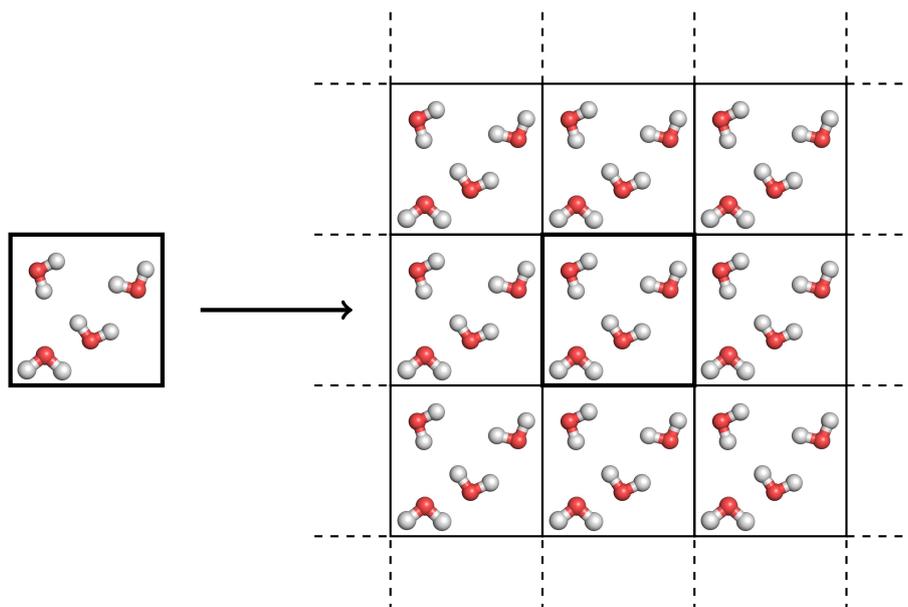


Figure B.3 – Conditions périodiques appliquées à un système en deux dimensions. La cellule correspondant à la boîte de simulation est encadrée, les autres cellules sont des images de la boîte de simulation primitive obtenues par translation. La flèche représente le déplacement d'une molécule qui sort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

B.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 B.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.

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

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 électrostatiques car ce sont des interactions à long portée (elles décroissent en $1/r_{ij}$) et la somme de l'équation B.2 a une convergence qui dépend de la façon dont on fait tendre le système vers l'infini.

La méthode d'Ewald [130, 121] 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 B.3).

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

où \mathbf{R} représente une position dans l'espace, \mathbf{R}_i représente la position de la charge q_i et les fonctions $g_{q_i}(\mathbf{R})$ représentent des densités de charges gaussiennes centrées en \mathbf{R}_i et dont la charge totale vaut q_i . Dans le premier terme de l'équation B.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 B.3 pour calculer l'énergie d'interaction électrostatique est un avantage car il remplace la somme de l'équation B.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* [131, 132]. Celui-ci est identique à la méthode d'Ewald pour le calcul du premier terme de l'équation B.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.

B.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 oeuvre 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 appelés 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 :

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

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 [133, 134, 135, 136]. 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 [133] 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 τ_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_0 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 B.4. A chaque pas de temps on ajuste les vitesses v_i des

atomes en les multipliant par un facteur correctif λ 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 τ_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 [136]. Un aperçu plus détaillé de l'équation de Langevin est donnée en annexe C page 177.

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 [137, p 159] :

$$P = \frac{2}{3V} \left(\sum_{i=1}^N \frac{1}{2} m_i v_i^2 - \sum_{i=1}^N \vec{q}_i \cdot \vec{F}_i \right)$$

où \vec{q}_i est le vecteur position de l'atome i et \vec{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 τ_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 μ qui est donné par la relation :

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

où τ_P est l'unique paramètre de ce modèle et β est la compressibilité isotherme de l'eau.

B.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 B.4).

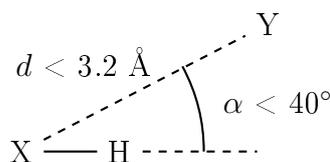


Figure B.4 – Le critère utilisé pour savoir si une liaison hydrogène existe comporte la distance X–Y et l'angle \widehat{HXY} . Les valeurs les plus souvent utilisées sont telles que d soit inférieur à 3.2 \AA et que α soit inférieur à 40° .

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

B.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.

B.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 maximum de la densité et elle est de l'ordre 1 \AA à 2 \AA .

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 I.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 (Na^+ ou 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 $5 - 10 \text{ kcal} \cdot \text{mol}^{-1}$). 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.

B.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⁻¹). 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 C

Simulation de dynamique Brownienne

La dynamique Brownienne est une approche qui permet d'étudier des échelles de temps plus longue que la dynamique moléculaire tout atome. Dans une simulation de dynamique moléculaire tout atome, on considère que le pas de temps utilisé pour intégrer la trajectoire doit être inférieur d'un ordre de grandeur au temps caractéristique du mouvement le plus rapide. Il en résulte que les mouvements les plus lents ne sont que très faiblement échantillonnés car les temps de simulations seraient trop long pour acquérir une statistique satisfaisante. Une façon de faire est de modéliser les coordonnées rapides par des forces aléatoires et un terme de friction. Les équations du mouvement de Newton sont alors remplacées par les équations du mouvement de Langevin.

Cette annexe présente les équations de Langevin, les paramètres qui interviennent et la façon dont elles sont résolues numériquement.

C.1 Les équations du mouvement

Nous allons considérer que, sur l'ensemble des coordonnées du système, on peut isoler une coordonnée q selon laquelle le mouvement est lent par rapport au mouvement sur les autres coordonnées. L'équation de Langevin, qui remplace celle de Newton, est la suivante :

$$m\ddot{q} = F(t) - \xi\dot{q} + \tilde{R} \tag{C.1}$$

où m est l'inertie associée à la coordonnée q , $F(t)$ est la force exercée par l'environnement le long de la coordonnée q , ξ est le coefficient de friction et \tilde{R} représente une force aléatoire de moyenne nulle et non corrélée à la coordonnée q (équation C.4). Du fait de la présence d'un terme aléatoire cette équation est dite stochastique.

L'équation de Langevin représente la dissipation d'énergie due aux échanges entre la coordonnée lente et les coordonnées rapides. La force aléatoire et la force de friction, qui est similaire à une force de frottement fluide, représentent les chocs avec les autres particules du système. La force

aléatoire satisfait les trois relations suivantes :

$$\langle \tilde{R}(t) \tilde{R}(0) \rangle = 2 m k_B T \xi \delta(t) \quad (\text{C.2})$$

$$\langle \tilde{R}(t) q(t) \rangle = 0 \quad (\text{C.3})$$

$$\langle \tilde{R}(t) \rangle = 0 \quad (\text{C.4})$$

où la moyenne porte sur une trajectoire. L'équation C.3 impose que la force aléatoire qui s'applique sur la coordonnée q est indépendante de cette coordonnée. L'équation C.2 relie le coefficient de friction avec la force aléatoire. Cette relation est nécessaire pour que la trajectoire puisse atteindre un équilibre, s'il existe.

A la place du coefficient de friction, on utilise aussi le coefficient de diffusion, D . Dans le cas d'un mouvement de rotation ils sont reliés par la relation d'Einstein :

$$\xi_{rot} = \frac{k_B T}{D_{rot}}$$

où D_{rot} est en s^{-1} .

C.2 Résolution numérique de l'équation de Langevin

L'équation C.1 est résolue par intégration numérique avec un pas d'intégration δt . Nous avons utilisé l'algorithme d'Ermak [121, p 263] qui est un algorithme du même type que celui de Verlet. Il impose de faire l'hypothèse que la force $F(t)$ varie peu sur l'intervalle de temps δt .

$$q(t + \delta t) = q(t) + c_1 \delta t \dot{q}(t) + c_2 \delta t^2 \ddot{q}(t) + \delta q^{rand} \quad (\text{C.5})$$

$$\dot{q}(t + \delta t) = c_0 \dot{q}(t) + (c_1 - c_2) \delta t \ddot{q}(t) + c_2 \delta t \ddot{q}(t + \delta t) + \delta \dot{q}^{rand} \quad (\text{C.6})$$

où δq^{rand} et $\delta \dot{q}^{rand}$ sont des termes aléatoires et les coefficients c_0 , c_1 et c_2 ont pour expressions :

$$\begin{aligned} c_0 &= \exp(-\xi \delta t) \\ c_1 &= \frac{1 - c_0}{\xi \delta t} \\ c_2 &= \frac{1 - c_1}{\xi \delta t} \end{aligned} \quad (\text{C.7})$$

Les termes aléatoires δq^{rand} et $\delta \dot{q}^{rand}$ sont corrélés et sont obtenus à partir de la densité de probabilité suivante :

$$\rho(\delta q^{rand}, \delta \dot{q}^{rand}) = \frac{1}{2\pi\sigma_q\sigma_v(1-c_{rand}^2)^{1/2}} \exp\left\{-\frac{1}{2(1-c_{rand}^2)}\left(\left(\frac{\delta q^{rand}}{\sigma_q}\right)^2 + \left(\frac{\delta \dot{q}^{rand}}{\sigma_v}\right)^2 - 2c_{rand}\left(\frac{\delta q^{rand}}{\sigma_q}\right)\left(\frac{\delta \dot{q}^{rand}}{\sigma_v}\right)\right)\right\} \quad (C.8)$$

où σ_q est l'écart type du déplacement aléatoire sur la coordonnée q , σ_v est l'écart type de la variation aléatoire de la vitesse de la coordonnée q , et c_{rand} est la corrélation entre ces deux variables aléatoires.

Ces trois paramètres sont donnés par les équations :

$$\begin{aligned} \sigma_q^2 &= \delta t^2 \frac{k_B T}{m} (\xi \delta t)^{-1} \left(2 - \frac{3 - 4e^{-\xi \delta t} + e^{-2\xi \delta t}}{\xi \delta t}\right) \\ \sigma_v^2 &= \frac{k_B T}{m} (1 - e^{-2\xi \delta t}) \end{aligned} \quad (C.9)$$

$$c_{rand}\sigma_q\sigma_v = \delta t \frac{k_B T}{m} (\xi \delta t)^{-1} (1 - e^{-\xi \delta t})^2 \quad (C.10)$$

Du fait de la présence d'un terme aléatoire dans l'équation C.1, une trajectoire unique obtenue avec des conditions initiales données n'a pas de sens particulier. On doit calculer la moyenne sur un grand nombre de trajectoires de la propriété considérée.

C.3 Tirage des nombres aléatoires corrélés

Les termes aléatoires des équations utilisées pour résoudre numériquement l'équation de Langevin, δq^{rand} et $\delta \dot{q}^{rand}$ sont corrélés (équation C.8), c'est à dire que leurs variations sont liées. Ce paragraphe présente une méthode efficace pour échantillonner des variables aléatoires corrélées. Dans ce qui suit on considère deux variables aléatoires X et Y de variance σ_X et σ_Y et de corrélation c . La corrélation est reliée aux valeurs moyennes, \bar{X} et \bar{Y} des variables X et Y par la relation :

$$\begin{aligned} c &= \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i X_i Y_i - \bar{X} \bar{Y}} \end{aligned}$$

On va supposer la densité de probabilité suivante pour les coordonnées X et Y qui est de la même forme que celle de l'équation C.8 :

$$\rho(X, Y) \propto \exp\left\{-\frac{1}{2(1-c^2)}\left((X/\sigma_X)^2 + (Y/\sigma_Y)^2 - 2c(X/\sigma_X)(Y/\sigma_Y)\right)\right\}$$

Pour simplifier les notations, nous utiliserons des variables réduites : $x = X/\sigma_X$ et $y = Y/\sigma_Y$. La densité de probabilité s'écrit alors :

$$\rho(x, y) \propto \exp\left(-\frac{x^2 + y^2 - 2cxy}{2(1 - c^2)}\right)$$

Une première méthode pour obtenir des valeurs de x et y consiste à utiliser la méthode du rejet [64, p 281] : on tire aléatoirement un couple (x, y) et on accepte le couple obtenu avec la probabilité $\rho(x, y)$.

Une deuxième méthode consiste à faire un changement de variables pour décorrélérer les variables x et y . Pour cela on écrit la densité de probabilité sous forme matricielle, avec \mathbf{M} une matrice de dimension 2 et \mathbf{V} un vecteur dans l'espace défini par les variables x et y .

$$\rho(x, y) \propto \exp\left\{-\frac{1}{2(1 - c^2)} (\mathbf{V}^t \mathbf{M} \mathbf{V})\right\} \quad \mathbf{M} = \begin{pmatrix} 1 & -c \\ -c & 1 \end{pmatrix}$$

On exprime ensuite le produit $\mathbf{V}^t \mathbf{M} \mathbf{V}$ dans la base qui diagonalise \mathbf{M} .

$$\mathbf{D} = \begin{pmatrix} 1 + c & 0 \\ 0 & 1 - c \end{pmatrix} \quad \mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

où $\mathbf{D} = \mathbf{U}^t \mathbf{M} \mathbf{U}$ et U est la transformation unitaire qui diagonalise \mathbf{M} . Soit z_1 et z_2 les nouvelles variables qui sont les composantes de $\mathbf{U} \mathbf{V}$. La densité de probabilité s'écrit alors :

$$\rho(z_1, z_2) \propto \exp\left\{-\left(\frac{(z_1)^2}{2(1 + c)} + \frac{(z_2)^2}{2(1 - c)}\right)\right\}$$

On calcule ensuite x et y de la façon suivante :

1. on tire aléatoirement deux nombres z_1 et z_2 selon la densité $\rho(z_1, z_2)$
2. on calcule X et Y avec les équations :

$$\begin{aligned} X &= \frac{1}{\sqrt{2}} (z_1 + z_2) \sigma_X \\ Y &= \frac{1}{\sqrt{2}} (z_1 - z_2) \sigma_Y \end{aligned}$$

Cette façon de faire est plus efficace que la simple méthode du rejet. En effet, si les deux nombres sont fortement corrélés, par la méthode du rejet, il faut essayer un grand nombre de couples avant d'en accepter un. De plus, si on utilise une fonction permettant de tirer aléatoirement une chaîne de nombres aléatoires gaussiens aucun tirage n'est rejeté ce qui accentue encore l'efficacité.

Annexe D

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.

D.1 Fonction d'onde et densité électronique

La fonction d'onde électronique du système, Ψ_i , 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 Ψ_i est antisymétrique. Chaque déterminant représente une configuration électronique du système :

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

où χ_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 Ψ_i est approchée par un seul déterminant, et les OM sont obtenues par diagonalisation de l'opérateur de Fock [138, p 297], 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 :

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

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 [138], 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.

D.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 [139] 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 [140] ont proposé d'écrire l'énergie comme une fonctionnelle de la densité, de la façon suivante :

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

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 D.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 D.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 oeuvre des méthode de dynamique quantique *ab 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 [83] et BLYP [84,85]. Le dernier type de fonctionnelles, appelées meta-GGA, inclue le laplacien de la densité électronique. La fonctionnelle B3LYP [88] 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.

D.3 Etats excités et transitions électroniques

D.3.1 Opérateur d'échelle

Soit Ψ_1 la fonction d'onde de l'état électronique fondamental. Soit Ψ_j la fonction d'onde d'un état électronique excité. L'opérateur d'échelle, \hat{T} , est l'opérateur permettant de construire la fonction d'onde Ψ_j à partir de Ψ_1 .

$$\Psi_j = \hat{T} \Psi_1$$

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 χ_k par l'OM virtuelle χ_l . L'opérateur d'échelle devient alors :

$$\hat{T} = \sum_k \sum_l a_{kl} \hat{t}_{kl} \tag{D.2}$$

où la somme sur k porte sur les OM occupées et la somme sur l porte sur les OM vacantes. Les coefficients a_{kl} 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_{nn+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.

D.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 ν 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 ν 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 :

$$f = \frac{|M_{ij}|^2}{|M_{12}^{harm}|^2} \quad (\text{D.3})$$

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 :

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

D.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 a_{kl} 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.

Bibliographie

- [1] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr.T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Rev C.02, 2004. Gaussian, Inc., Wallingford, CT.
- [2] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. AMBER 9, 2006. University of California, San Fransisco.
- [3] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Fergusson, D.C. Spellmeyer, 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.
- [4] 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.
- [5] Bernard Valeur. *Molecular Fluorescence – Principles and Applications*. Wiley-VCH, Weinheim (Germany), Weinheim (Germany), 2006.
- [6] M. Mostafavi and T. Gustavsson. *Réaction ultrarapides en solutions - approches expérimentales et théoriques*. CNRS édition, 2006.
- [7] R.Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem.*, 67 :509–544, 1998.
- [8] M. Zimmer. Green Fluorescent Protein (GFP) : Application, Structure, and Related Photo-physical Behavior. *Chem. Rev.*, 102 :759–781, 2002.

- [9] M. Chalfie and S.R. Kain. *Green fluorescent protein : properties, applications and protocols*, volume 47. New York : Wiley, methods of biochemical analysis edition, 2005.
- [10] K.F. Sullivan and S.A. Kay. *Green Fluorescent Proteins. Methods in Cell Biology*, volume 58. Academic Press, 1999.
- [11] O. Shimomura, F.H. Johnson, and Y. Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan aequorea. *J. Cell. Comp. Physiol.*, 59 :223–239, 1962.
- [12] F.H. Johnson, O. Shimomura, Y. Saiga, L.C. Gershman, G.T. Reynolds, and J.R. Waters. Quantum efficiency of cypridina luminescence, with a note on that of aequorea. *J. Cell. Comp. Physiol.*, 60(1) :85–103, 1962.
- [13] H. Morise, O. Shimomura, F.H. Johnson, and J. Winant. Intermolecular energy transfer in the bioluminescent system. *Biochemistry*, 13 :759–781, 1974.
- [14] J.G. Morin and J.W. Hastings. Energy transfer in a bioluminescent system. *J. Cell. Physiol.*, 77(3) :313–318, 1970.
- [15] D.C. Prasher, V.K. Eckenrode, W.W. Ward, F.G. Prendergast, and M.J. Cormier. Primary structure of the aequorea victoria green-fluorescent protein. *Gene*, 111(2) :229–233, 1991.
- [16] M. Chalfie, Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263 :802–805, 1994.
- [17] M.A. Perozzo, K.B. Ward, R.B. Thompson, and W.W. Ward. X-ray diffraction and time-resolved fluorescence analyses of aequorea green fluorescent protein crystals. *J. Biol. Chem.*, 263(16) :7713–7716, 1988.
- [18] M. Ormö, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, and S.J. Remington. Crystal structure of the aequorea victoria green fluorescent protein. *Science*, 273(5280) :1392–1395, 1996.
- [19] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Research*, 28 :235–242, 2000.
- [20] F. Yang, L.G. Moss, and G.N. Phillips Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol.*, 14(10) :1246–1251, 1996.
- [21] O. Shimomura. Structure of the chromophore of aequorea green fluorescent protein. *FEBS Lett.*, 104(2) :220–222, 1979.
- [22] A.B. Cubitt, R. Heim, S.R. Adams, A.E. Boyd, L.A. Gross, and R. Tsien. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.*, 20 :448–455, 1995.
- [23] B.G. Reid and G.C. Flynn. Chromophore formation in green fluorescent protein. *Biochemistry*, 36(22) :6786–6791, 1997.
- [24] V. Helms, T.P. Straatsma, and J.A. McCammon. Internal dynamics of green fluorescent protein. *J. Phys. Chem. B*, 103 :3263–3269, 1999.
- [25] N.C. Shaner, G.H. Patterson, and M.W. Davidson. Advances in fluorescent protein technology. *J. Cell. Sci.*, 120 :4247–4260, 2007.

-
- [26] A.A. Pakhomov and V.I. Martynov. GFP family : Structural insights into spectral tuning. *Chemistry & Biology*, 15(8) :755–764, 2008.
- [27] R. Heim, D.C. Prasher, and R.Y. Tsien. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 91(26) :12501–12504, 1994. Premiere publi CFP.
- [28] G.H. Patterson, R.N. Day, and D. Piston. Fluorescent protein spectra. *J. Cell Sci.*, 114 :837–838, 2001.
- [29] R.M. Wachter, M.A. Elsliger, K. Kallio, G.T. Hanson, and S.J. Remington. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure with folding & design*, 6(10) :1267–1277, 1998.
- [30] A. Miyawaki, J. Llopis, R. Heim, J.M. McCaffery, J.A. Adams, M. Ikura, and R.Y. Tsien. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature*, 388 :882–887, 1997.
- [31] V.A. Romoser, P.M. Hinkle, and A. Persechini. Detection in living cells of Ca^{2+} -dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. a new class of fluorescent indicators. *J. Biol. Chem.*, 272 :13270–13274, 1997.
- [32] P.R. Selvin. The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.*, 7 :730–734, 2000.
- [33] S.S. Vogel, C. Thaler, and S.V. Koushik. Fanciful FRET . *Sci. STKE*, (331) :re2, 2006.
- [34] B. Domingo, R. Sabariego, F. Picazo, and J. Llopis. Imaging FRET standards by steady-state fluorescence and lifetime methods. *Microsc. Res. Tech.*, 70(12) :2007, 1010-1021.
- [35] A. Villoing, M. Ridhoir, B. Cinquin, M. Erard, L. Alvarez, G. Vallverdu, P. Pernot, R. Grailhe, F. Mérola, and H. Pasquier. Complex Fluorescence of the Cyan Fluorescent Protein : Comparisons with the H148D Variant and Consequences for Quantitative Cell Imaging. *Biochemistry*, 47 :12483–12492, 2008.
- [36] M.A. Rizzo, G.H. Springer, B. Granada, and D.W. Piston. An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.*, 22(4) :445–449, 2004.
- [37] G.D. Malo, L.J. Pouwels, M.T. Wang, A. Weichsel, W.R. Montfort, M.A. Rizzo, D.W. Piston, and R.M. Wachter. X-ray structure of cerulean GFP : A tryptophan-based chromophore useful for fluorescence lifetime imaging. *Biochemistry*, 46(35) :9865–9873, 2007.
- [38] J. Lippincott-Schwartz, N. Altan-Bonnet, and G.H. Patterson. Photobleaching and photoactivation : following protein dynamics in living cells. *Nat. Cell. Biol.*, 5 :S7–S14, 2003.
- [39] K.A. Lukyanov, D.M. Chudakov, S. Lukyanov, and V.V. Verkhusha. Photoactivatable fluorescent proteins. *Nat. Rev. Mol. Cell. Biol.*, 6 :885–891, 2005.
- [40] S.J. Remington. Fluorescent proteins : maturation, photochemistry and photophysics. *Curr. Opin. Struct. Biol.*, 16(6) :714–721, 2006.
- [41] W.J. Jorgensen and J. Tirado-Rives. The opl [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6) :1657–1666, 1988.

- [42] U.C. Singh and P.A. Kollman. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.*, 5(2) :129–145, 1984.
- [43] C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.*, 97(40) :10269–10280, 1993.
- [44] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollman. Application of resp charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.*, 115(21) :9620–9631, 1993.
- [45] J. Ridard and B. Lévy. Effective atomic charges in alanine dipeptide. *J. Comput. Chem.*, 20(5) :473–482, 1999.
- [46] B. Lévy and M. Enescu. Theoretical study of methylene blue : a new method to determine partial atomic charges ; investigation of the interaction with guanine. *J. Mol. Struct. (Theochem.)*, 432(3) :235–245, 1998.
- [47] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9) :1157–1174, 2004.
- [48] 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.
- [49] 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.
- [50] G.M. Torrie and J.P. Valleau. Monte carlo free energy estimates using non-boltzmann sampling : Application to the sub-critical lennard-jones fluid. *Chem. Phys. Lett.*, 28(4) :578–581, 1974.
- [51] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation : Umbrella sampling. *J. Comput. Phys.*, 23(2) :187–199, 1977.
- [52] J.M. Bui, R.H. Henchman, and J.A. McCammon. The dynamics of ligand barrier crossing inside the acetylcholinesterase gorge. *Biophys. J.*, 85(4) :2267–2272, 2003.
- [53] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, and J.M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecule. 1. the method. *J. Comput. Chem.*, 13(8) :1011–1021, 1992.
- [54] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, and P.A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16(11) :1339–1350, 1995.
- [55] J. Kästner and W. Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method : “umbrella integration”. *J. Chem. Phys.*, 123(14) :144104, 2005.
- [56] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268 :1144–1149, 1995.

- [57] D. Bashford and M. Karplus. pka's of ionizable groups in proteins : atomic detail from a continuum electrostatic model. *Biochemistry*, 29(44) :10219–10225, 1990.
- [58] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, and A. Onufriev. H++ : a server for estimating pK(a)s and adding missing hydrogens to macromolecules, <http://biophysics.cs.vt.edu/H++>. *Nucleic Acids Research*, 33 :W368–W371, 2005.
- [59] R. Anandakrishnan and A. Onufriev. Analysis of basic clustering algorithms for numerical estimation of statistical averages in biomolecules. *J. Comput. Biol.*, 15(2) :165–184, 2008.
- [60] H. Li, A.D. Robertson, and J.H. Jensen. Very fast empirical prediction and rationalization of protein pka values. *Proteins*, 61(4) :704–721, 2005.
- [61] D.C. Bas, D.M. Rogers, and J.H. Jensen. Very fast prediction and rationalization of pka values for protein-ligand complexes. *Proteins*, 73(3) :765–783, 2008.
- [62] B. Efron. Bootstrap methods : Another look at the jackknife. *Ann. Statist.*, 7(1) :1–26, 1979.
- [63] A.R. Henderson. The bootstrap : A technique for data-driven statistics. using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta*, 359(1-2) :1–26, 2005.
- [64] W.H. Press, W.T. Vetterling, S.A. Teukolsky, and B.P. Flannery. *Numerical Recipes in Fortran - The art of scientific Computing*. Cambridge University Press, 1992.
- [65] R. Heim and R.Y. Tsien. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.*, 6 :178–182, 1996.
- [66] M.H.J. Seifert, D. Ksiazek, M.K. Azim, P. Smialowski, N. Budisa, and T.A. Holak. Slow exchange in the chromophore of a green fluorescent protein. *J. Am. Chem. Soc.*, 124 :7932–7942, 2002.
- [67] J.H. Bae, M. Rubini, G. Jung, G. Wiegand, M.H.J. Seifert, J.S. Kim M.K. Azim, J.S. Kim, A. Zumbusch, T.A. Holak, R. Huber, L. Moroder, and N. Budisa. Expansion of the Genetic Code Enables Design of a Novel Gold Class of Green Fluorescent Proteins. *J. Mol. Biol.*, 328 :1071–1081, 2003.
- [68] D. Sinnecker, P. Voigt, N. Hellwig, and M. Schaefer. Reversible photobleaching of enhanced green fluorescent proteins. *Biochemistry*, 44(18) :7085–7094, 2005.
- [69] I. Demachy, J. Ridard, H. Laguitton-Pasquier, E. Durnerin, G. Vallverdu, P. Archirel, and B. Lévy. Cyan Fluorescent Protein : Molecular Dynamics, Simulations, and Electronic Absorption Spectrum. *J. Phys. Chem. B*, 109 :24121–24133, 2005.
- [70] Myriam Rydhoir. Relation structure - propriétés de fluorescence de variants spectraux de la enhanced cyan fluorescent protein, 2007. Mémoire de fin d'étude du Conservatoire National des Arts et Métier.
- [71] V. Adam, M. Lelimosin, S. Boehme, G. Desfonds, K. Nienhaus, M.J. Field, J. Wiedenmann, S. McSweeney, G.U. Nienhaus, and D. Bourgeois. Structural characterization of irisfp, an optical highlighter undergoing multiple photo-induced transformations. *Proc. Natl. Acad. Sci. USA*, 105(47) :18343–18348, 2008.
- [72] R. Nifosi, P. Amat, and V. Tozzini. Variation of spectral, structural, and vibrational properties within the intrinsically fluorescent proteins family : A density functional study. *J. Comput. Chem.*, 28(4) :2366–2377, 2007.

- [73] J.Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara, and H. Nakatsuji. Excited states of GFP chromophore and active site studied by the SAC-CI method : Effect of protein-environment and mutations. *J. Comput. Chem.*, 28(15) :2443–2452, 2007.
- [74] J. Mongan, D.A. Case, and J.A. McCammon. Constant ph molecular dynamics in generalized born implicit solvent. *J. Comput. Chem.*, 25(16) :2038–2948, 2004.
- [75] N. Reuter, H. Lin, and W. Thiel. Green Fluorescent Proteins : Empirical Force Field for the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics Simulations of the Wild Type and S65T Mutant. *J. Phys. Chem. B*, 106 :6310–6321, 2002.
- [76] M.K. Gilson. Modeling protonation equilibria in biomolecules. In W.F. van Gunsteren, P.K. Weiner, and A.J. Wilkinson, editors, *Computer simulation of biomolecular systems, Theoretical and experimental applications*, volume Volume 3, pages 199–222. Kluwer Academic Publishers, 1997.
- [77] X. Shu, K. Kallio, X. Shi, P. Abbyad, P. Kanchanawong, W. Childs, S.G. Boxer, and S.J. Remington. Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 1. Mutagenesis and structural studies. *Biochemistry*, 46(43) :12005–12013, 2007.
- [78] G. Vallverdu, I. Demachy, J. Ridard, and B. Lévy. Using biased molecular dynamics and brownian dynamics in the study of fluorescent proteins. *J. Mol. Struct. (Theochem.)*, 898 :73–81, 2009.
- [79] 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.
- [80] L.V. Schafer, G. Groenhof, A.R. Klingen, G.M. Ullmann, M. Boggio-Pasqua, M.A. Robb, and H. Grubmueller. Photoswitching of the fluorescent protein asFP595 : Mechanism, proton pathways, and absorption spectra. *Angew. Chem. Int. Ed.*, 46(4) :530–536, 2007.
- [81] J.C. Slater. *Quantum Theory of Molecular and Solids. Vol. 4 : The Self-Consistent Field for Molecular and Solids*. McGraw-Hill, New York, 1974.
- [82] S.H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations : a critical analysis. *Can. J. Phys.*, 58(8) :1200–1211, 1980.
- [83] J.P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18) :3865–3868, 1996.
- [84] A.D. Becke. Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A*, 38(6) :3098–3100, 1988.
- [85] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37 :785–789, 1988.
- [86] J.P. Perdew. Density-functional approximation for the correlation-energy of the inhomogeneous electron-gas. *Phys. Rev. B*, 33(12) :8822–8824, 1986.

-
- [87] J.M. Tao, J.P. Perdew, V.N. Staroverov, and G.E. Scuseria. Climbing the density functional ladder : Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91(146401), 2003.
- [88] A.D. Becke. Density-Functional Thermochemistry .3. The Role Of Exact Exchange. *J. Chem. Phys.*, 98(7) :5648–5652, 1993.
- [89] V.N. Staroverov, G.E. Scuseria, J.M. Tao, and J.P. Perdew. Comparative assessment of a new nonempirical density functional : Molecules and hydrogen-bonded complexes. *J. Chem. Phys.*, 119(23) :12129–12137, 2003.
- [90] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters : The pbe0 model. *J. Chem. Phys.*, 110(13) :6158–6170, 1999.
- [91] Q.K. Timerghazin, H.J. Carlson, C. Liang, R.E. Campbell, and A. Brown. Computational prediction of absorbance maxima for a structurally diverse series of engineered green fluorescent protein chromophores. *J. Phys. Chem. B*, 112(8) :2533–2541, 2008.
- [92] P. Beroza, D.R. Fredkin, M.Y. Okamura, and G. Feher. Protonation of interacting residues in a protein by a monte carlo method : Application to lysozyme and the photosynthetic reaction center of rhodobacter sphaeroides. *Proc. Natl. Acad. Sci. USA*, 88(13) :5804–5808, 1991.
- [93] G. Striker, V. Subramaniam, C.A.M. Seidel, and A. Volkmer. Photochromicity and fluorescence lifetimes of green fluorescent protein. *J. Phys. Chem. B*, 103(40) :8612–8617, 1999.
- [94] C. Scharnagl, R.A. Raupp-Kossmann, and F. Fischer. Molecular basis for ph sensitivity and proton transfert in green fluorescent protein : Protonation and conformational substates from electrostatic calculations. *Biophys. J.*, 77 :1839–1857, 1999.
- [95] O. Vendrell, R. Gelabert, M. Moreno, and J. Lluch. Potential energy landscape of the photoinduced multiple proton-transfert process in the green fluorescent protein : Classical molecular dynamics and multiconfigurational electronic structure calculations. *J. Am. Chem. Soc.*, 128 :3564–3574, 2006.
- [96] O. Vendrell, R. Gelabert, M. Moreno, and J.M. Lluch. A potential energy function for heterogeneous proton-wires. ground end photoactive states of the proton-wire in the green fluorescent protein. *J. Chem. Theory Comput.*, 4(7) :1138–1150, 2008.
- [97] K. Brejc, T.K. Sixma, P.A. Kitts, S.R. Kain, R.Y. Tsien, and M. Ormö. Structural basis for dual excitation and photoisomerization of the aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 94(6) :2306–2311, 1997.
- [98] A.D. Kummer, C. Kompa, H. Niwa, T. Hirano, S. Kojima, and M.E. Michel-Beyerle. Viscosity-dependent fluorescence decay of the GFP chromophore in solution due to fast internal conversion. *J. Phys. Chem. B*, 106(30) :7554–7559, 2002.
- [99] D. Mandal, T. Tahara, and S.R. Meech. Excited-state dynamics in the green fluorescent protein chromophore. *J. Phys. Chem. B*, 108(3) :1102–1108, 2004.
- [100] K.M. Solntsev, O. Poizat, J. Dong, J. Rehaut, Y.B. Lou, C. Burda, and L.M. Tolbert. Meta and para effects in the ultrafast excited-state dynamics of the green fluorescent protein chromophores. *J. Phys. Chem. B*, 112(9) :2700–2711, 2008.

- [101] R. Gepshtein, D. Huppert, and N. Agmon. Deactivation mechanism of the green fluorescent chromophore. *J. Phys. Chem. B*, 110(9) :4434–4442, 2006.
- [102] H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi, and F.I. Tsuji. Chemical nature of the light emitter of the aequorea green fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 93(24) :13617–13622, 1996.
- [103] W. Weber, V. Helms, J.A. McCammon, and P.W. Langhoff. Shedding light on the dark and weakly fluorescent states of green fluorescent proteins. *Proc. Natl. Acad. Sci. USA*, 96(11) :6177–6182, 1999.
- [104] A.A. Voityuk, M.E. Michel-Beyerle, and N. Rosch. Structure and rotation barriers for ground and excited states of the isolated chromophore of the green fluorescent protein. *Chem. Phys. Lett.*, 296(3-4) :269–276, 1998.
- [105] C. Scharnagl and R.A. Raupp-Kossmann. Solution pK(a) values of the green fluorescent protein chromophore from hybrid quantum-classical calculations. *J. Phys. Chem. B*, 108(1) :477–489, 2004.
- [106] V. Helms, C. Winstead, and P.W. Langhoff. Low-lying electronic excitations of the green fluorescent protein chromophore. *J. Mol. Struct. (Theochem.)*, 506(Sp. Iss. SI) :179–189, 2000.
- [107] 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.
- [108] 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.
- [109] S.L. Maddalo and M. Zimmer. The role of the protein matrix in green fluorescent protein fluorescence. *Photochem. Photobiol.*, 82(2) :367–372, 2006.
- [110] C.M. Megley, L.A. Dickson, S.L. Maddalo, G.J. Chandler, and M. Zimmer. Photophysics and dihedral freedom of the chromophore in yellow, blue, and green fluorescent protein. *J. Phys. Chem. B*, 113 :302–308, 2009.
- [111] M.C. Chen, C.R. Lambert, J.D. Urgitis, and M. Zimmer. Photoisomerization of green fluorescent protein and the dimensions of the chromophore cavity. *Chem. Phys.*, 270(1) :157–164, 2001.
- [112] N.Y.A. Baffour-Awuah and M. Zimmer. Hula-twisting in green fluorescent protein. *Chem. Phys.*, 303(1-3) :7–11, 2004.
- [113] R.S.H. Liu and A.E. Asato. The primary process of vision and the structure of bathorhodopsin - a mechanism for photoisomerization of polyenes. *Proc. Natl. Acad. Sci. USA*, 82(2) :259–263, 1985.
- [114] R.S.H. Liu and D.T. Browne. A bioorganic view of the chemistry of vision - HT-n and BP-m,n mechanisms for reactions of confined, anchored polyenes. *Acc. Chem. Res.*, 19 :42–48, 1986.

-
- [115] R.S.H. Liu and G.S. Hammond. Hula-twist a volume-conserving mechanism of cis-trans photoisomerization. *Chem. Eur. J.*, 7(21) :4536–4544, 2001.
- [116] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory : fifty years after kramers. *Reviews of Modern Physics*, 62(2) :251–341, 1990.
- [117] G.H. Patterson, S.M. Knobel, W.D. Sharif, S.R. Kain, and D.W. Piston. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.*, 73(5) :2782–2790, 1997.
- [118] R. Nifosi and V. Tozzini. Molecular dynamics simulations of enhanced green fluorescent proteins : Effects of f64l, s65t and t203y mutations on the ground-state proton equilibria. *Proteins : Struct., Funct., Genet.*, 51(3) :378–389, 2003.
- [119] S.S. Patnaik, S. Trohalaki, and R. Pachter. Molecular modeling of green fluorescent protein : Structural effects of chromophore deprotonation. *Biopolymers*, 75(6) :441–452, 2004.
- [120] J.A. McCammon and S.C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, 1997.
- [121] M.P. Allen and D.J. Tiddesley. *Computer simulation of liquids*. Oxford University Press, 1987.
- [122] J.A. McCammon and M. Karplus. Dynamics of tyrosine ring rotations in a globular protein. *Biopolymers*, 19 :1375–1405, 1980.
- [123] G. Hummer and I.G. Kevrekidis. Coarse molecular dynamics of a peptide fragment : Free energy, kinetics, and long-time dynamics computations. *J. Chem. Phys.*, 118(23) :10762–10773, 2003.
- [124] D.S. Chekmarev, T. Ishida, and R.M. Levy. Long-time conformational transitions of alanine dipeptide in aqueous solution : Continuous and discrete-state kinetic models. *J. Phys. Chem. B*, 108(50) :19487–19495, 2004.
- [125] W. Domcke, D.R. Yarkony, and H. Köppel. *Advanced Series in Physical Chemistry, Vol 15, Conical Intersections - Electronic structure, dynamics & spectroscopy*. World Scientific Publishing, 2004.
- [126] M.S. Child. *Molecular Collision Theory*. Academic Press London and New York, 1984.
- [127] S. Olsen and R.H. McKenzie. A diabatic three-state representation of photoisomerization in the green fluorescent protein chromophore. *J. Chem. Phys.*, 130(18), 2009.
- [128] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2) :926–935, 1983.
- [129] 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.
- [130] P.P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3) :253–287, 1921.

- [131] T.A. Darden, D. York, and L. Pedersen. Particle mesh Ewald : An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98 :10089–10092, 1993.
- [132] U. Essmann, L. Perera, M.L. Berkowitz, T.A. Darden, H. Lee, and L.G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19) :8577–8593, 1995.
- [133] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gusteren, A. Di Nola, and J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8) :3684–3690, 1984.
- [134] T.A. Andrea, W.C. Swope, and H.C. Andersen. The role of long ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.*, 79 :4576–4584, 1983.
- [135] H.C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72 :2384–2393, 1980.
- [136] R.J. Loncharich, B.R. Brooks, and R.W. Pastor. Langevin dynamics of peptides : The frictional dependence of isomerization rates of n-acetylalanyl-n-methylamider. *Biopolymers*, 32(5) :523–535, 2004.
- [137] D. Frenkel and B. Smit. *Understanding molecular simulation - From algorithm to applications*. Academic Press, second edition edition, 2002.
- [138] J.L. Rivail. *Elements de chimie quantique à l'usage des chimistes*. Edition du CNRS, 1989.
- [139] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B) :B864–B871, 1964.
- [140] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A) :A1133–A1138, 1965.

Liste des figures

I.1	Diagramme de Perrin-Jablonski représentant les phénomènes photophysiques possibles dans un état excité. RV signifie relaxation vibrationnelle, CI signifie conversion interne et CS signifie croisement inter-système.	10
I.2	a) et b) Structure tridimensionnelle de la GFP vue de face et de dessus. Les feuillets rouges définissent le <i>cleft</i> qui est une faiblesse de la structure de la GFP. c) Vue du chromophore de la GFP à l'intérieur de la protéine. Il est lié à l'hélice α centrale par deux liaisons covalentes.. . . .	15
I.3	Mécanisme proposé par Cubitt <i>et al</i> [22] pour la synthèse intra-moléculaire du chromophore de la GFP	16
I.4	Orbitale moléculaire HOMO et LUMO du chromophore de la GFP obtenue par un calcul B3LYP/6-31G*.	16
I.5	Echantillon des protéines fluorescentes actuellement disponibles.	17
I.6	Représentation et numérotation des atomes du chromophore de la Cerulean en conformation <i>cis</i> et <i>trans</i> et du chromophore anionique de la GFP.	20
I.7	Schéma représentant une chaîne d'acides aminés reliés par des liaisons peptidiques. Le cadre délimite l'entité associée à un acide aminé standard dans AMBER.	21
I.8	Courbes d'énergie potentielle des torsions autour des liaisons C12-C11 (angle dièdre τ) et C11-C7 (angle dièdre φ) pour les deux conformations <i>cis</i> et <i>trans</i> du chromophore. Les calculs <i>ab initio</i> ont été fait au niveau B3LYP/6-31G* sur des géométries partiellement optimisées. <i>ff99</i> : valeurs calculées avec le champ de force <i>ff99</i>	32
I.9	Comparaison entre la courbe d'énergie potentielle de l'état excité du chromophore calculée par Toniolo <i>et al</i> et celle calculée avec le champ de force <i>ff99</i> après ajustement des paramètres associés à la torsion φ	35
I.10	Courbes $\mathcal{F}_i(q)$ obtenues à partir d'une simulation sans contraintes et de simulations biaisées.	40
I.11	Deux courbes $\mathcal{F}_i(q)$ issues de deux fenêtres consécutives et les deux histogrammes biaisés des valeurs de q correspondants.	41

I.12	Schéma permettant de visualiser la méthode du <i>bootstrap</i> . L'échantillon initial E_i est obtenu à partir de l'espace complet des éléments par une méthode d'échantillonnage (dynamique moléculaire, tirage aléatoire ...). Les échantillons de <i>bootstrap</i> sont obtenus en tirant aléatoirement N éléments parmi les N éléments de l'échantillon initial avec remise : un élément peut donc être tiré plusieurs fois.	46
I.13	a) Incertitude $u(\langle x \rangle)$ sur la valeur moyenne de x (en noir) et $u(\sigma_x)$ sur l'écart type de x (en rouge), en fonction du nombre N de valeurs de x dans l'échantillon initial. b) En rouge, distribution $\mathcal{D}^b(\langle x \rangle)$ des valeurs moyennes calculées sur les N_b échantillons de <i>bootstrap</i> dans le cas où l'échantillon initial contient $N=1000$ valeurs de x (en haut) et $N=50$ valeurs de x (en bas). Le trait vert représente la vraie valeur moyenne de la distribution utilisée et le trait noir représente la valeur moyenne de l'échantillon initial.	48
II.1	a : 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). b : Chromophore dans la conformation <i>trans</i> . c : Chromophore dans la conformation <i>cis</i>	51
II.2	a) 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 <i>cis</i> . b) Spectre d'absorption de la Cerulean en solution en fonction du temps après un saut de pH de pH 8 à pH 5. A $t = 0$ (rouge ○) pH 8, 20 min (violet □), 40 min (vert ◇), 60 min (orange △), 90 min (bleu ▽), 150 min (noir ×) pH 5.	53
II.3	Longueur d'onde de la première transition électronique et force d'oscillateur associée (entre parenthèses) pour la conformation <i>cis</i> du chromophore neutre et du chromophore cationique et la conformation <i>trans</i> du chromophore neutre.	54
II.4	Courbe de titration de Asp148 dans la Cerulean. L'ordonnée représente la probabilité que Asp148 soit dans sa forme acide (protoné et neutre). Les deux molécules en haut à gauche et en bas à droite représentent respectivement la forme acide et la forme basique d'un acide aspartique.	57
II.5	Les quatre positions initiales envisagées pour le proton acide de Asp148.	62
II.6	Noms des atomes employés pour Asp148. Les angles θ , α et β correspondent respectivement aux torsions autour des liaisons CA-CB, CB-CG et CG-ODH. La torsion θ est mesurée par l'angle dièdre C-CA-CB-CG, la torsion α est mesurée par l'angle dièdre CA-CG-CB-ODH et la torsion β est mesurée par l'angle dièdre CB-CG-ODH-HD.	63

- II.7 Déplacement carré moyen en fonction du temps des C_α des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β en noir, de l'ensemble des C_α en rouge et de l'ensemble des atomes de la protéine en orange dans la simulation 1. 64
- II.8 En noir, orientation de la chaîne latérale de Asp148 dans les simulations 1, 2 et 3. Il n'y a pas de liaison intra-acide aminé entre l'acide carboxylique et le squelette protéique de Asp148, l'hydrogène HD est orienté à l'extérieur du groupe carboxylique et les deux atomes d'oxygène du groupe carboxylique sont proches du groupe NH de l'indole du chromophore. La géométrie bleu transparente est issue de la structure cristalline, le proton acide étant positionné selon l'hypothèse d'une liaison hydrogène intra-acide aminé. 66
- II.9 Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à partir de la simulation 1 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 3.2 Å pour la distance et 40° pour l'angle. 68
- II.10 a) Représentation du *cleft* entre les brins 6 (en jaune) et 7 (en vert) dans la Cerulean. La chaîne latérale de Asp148 (brin 6) est également représentée. b) Schéma représentant les acides aminés constituant les brins 6 et 7 et les liaisons entre les fonctions amides de chaque paire d'acides aminés. Les traits pleins représentent les liaisons hydrogène présentes dans la structure RX. Les tirets représentent les liaisons hydrogène manquantes. 69
- II.11 Trajectoires de la distance entre l'atome ODH de Asp148 et l'atome d'oxygène du carbonyle de Lys166 (en noir), de la moyenne des deux longueurs des liaisons hydrogènes entre Tyr151 et Asn164 (en rouge) et de la densité de la boîte de simulation (en violet). Le temps $t=0$ correspond au relâchement des contraintes et les barres verticales séparent les différents type de simulation. 73
- II.12 Vue des interactions entre des atomes d'une protéine centrale (en bleu) et deux protéines voisines dans la région (en rouge) des brins 6 et 7 de la protéine centrale. La protéine jaune est obtenue par l'opération de symétrie $(-x - a, -y, z - c)$ et la protéine rose est obtenue par l'opération de symétrie $(x - a, -y - b, -z)$. Les acides aminés représentés sont ceux dont au moins un atome est interaction avec un atome de la région rouge de la protéine centrale (tableau II.5).. . . . 76

II.13	Vue des interactions inter-protéines dans le cristal de la GFP(structure RX code 1EMA). La protéine centrale est à gauche, la protéine de droite est obtenue par l'opération de symétrie (x , $-y + b$, $-z + c$). La région représentée en rouge représente les acides aminés de la protéine centrale inclus dans la recherche des contacts avec les protéines voisines. Les acides aminés représentés sont ceux dont au moins un atome est interaction avec un atome de la région rouge de la protéine centrale (tableau II.5), soit directement, soit par l'intermédiaire d'une molécule d'eau.	78
II.14	Trajectoires de la distance entre les atomes d'oxygène du groupe COO^- de Asp148 et l'atome N9 du chromophore de la Cerulean sur l'ensemble de la simulation à pH 8 avec le chromophore en conformation <i>cis</i> .La simulation de la protéine sans contrainte commence à $t=0.1$ ns.	79
II.15	Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à partir de la simulation à pH 8 avec le chromophore en conformation <i>cis</i> . 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 \AA pour la distance et 40° pour l'angle.. . . .	81
II.16	Profil d'énergie libre en fonction de la torsion θ autour de la liaison CA-CB de Asp148 à pH 8 (angle dièdre C-CA-CB-CG.	82
II.17	Déplacement carré moyen, en fonction du temps, des C_α des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β (en noir), de l'ensemble des C_α (en rouge) et de l'ensemble des atomes de la protéine (en orange) dans la simulation pH8- <i>trans</i>	83
II.18	Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à partir de la simulation à pH 8 avec le chromophore en conformation <i>trans</i> . 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 \AA pour la distance et 40° pour l'angle.. . . .	84
II.19	Superposition des structures de la Cerulean au cours de la simulation pH8- <i>trans</i> et de la structure RX de la ECFP dans sa conformation A, ou dans sa conformation B. Le brin 6 qui porte His148 dans la ECFP et Asp148 dans la Cerulean est mis en évidence. Il est représenté en vert dans la Cerulean et en rose dans la ECFP. Les acides aminés Asp148 et His148 sont représentés respectivement en vert et en rose.	85

II.20	Spectres d'absorption de la Cerulean, à pH 5 avec le chromophore en conformation <i>cis</i> , pH5- <i>cis</i> (noir), à pH 8 avec le chromophore en conformation <i>cis</i> , pH8- <i>cis</i> (pointillés rouges) et à pH 8 avec le chromophore en conformation <i>trans</i> , pH8- <i>trans</i> (trait plein rouge).	88
II.21	a) Variation en fonction du nombre N_e d'échantillons choisis au hasard par la méthode du <i>bootstrap</i> , de la valeur moyenne (en noir, échelle à gauche) et de l'écart type (en rouge, échelle à droite) des valeurs de λ_{max} calculés sur les N_e échantillons. b) Spectre d'absorption moyen (en noir) et fluctuations du spectre d'absorption (en marron).	89
II.22	Variation de l'incertitude sur λ_{max} en fonction du nombre N_{sub} d'éléments dans l'échantillon primitif.	90
II.23	Spectres d'absorption calculés avec les fonctionnelles B3LYP, BLYP, BP86, PBE, PBE0, SVWN, TPSS et TPSSh à partir de 200 configurations extraites de la simulation à pH 5 avec le chromophore en conformation <i>cis</i>	92
II.24	Distribution $p(\Delta\lambda_{max} DFT)$ du décalage du maximum du spectre d'absorption en fonction de la fonctionnelle utilisée a) lorsque le pH varie de pH 5 à pH 8 le chromophore restant en conformation <i>cis</i> b) lorsque le pH varie de pH 5 à pH 8 et le chromophore passe de sa conformation <i>cis</i> à sa conformation <i>trans</i> . La distribution noire avec des carrés est la moyenne des distributions ; elle est obtenue par l'équation II.4.	93
II.25	Orbitale moléculaire du chromophore de la Cerulean en conformation <i>cis</i> obtenues au niveau B3LYP/6-31G*. La valeur de l'orbitale sur l'isosurface tracée est 0.02.	96
II.26	Spectres d'absorption a) pour la simulation pH5- <i>cis</i> b) pour la simulation pH8- <i>cis</i> et c) pour la simulation pH8- <i>trans</i> où la contribution à l'absorbance de chaque état excité est explicitée : en noir le premier état excité, en rouge le deuxième état excité, en vert le troisième état excité et en bleu le quatrième état excité.	97
II.27	Densités de probabilité de la différence d'énergie entre l'orbitale HOMO et l'orbitale LUMO du chromophore construites à partir des calculs TDDFT utilisés pour faire les spectres d'absorption, à pH5 lorsque le chromophore est en conformation <i>cis</i> (en noir), à pH8 lorsque le chromophore est en conformation <i>cis</i> (en rouge) et à pH 8 lorsque le chromophore est en conformation <i>trans</i>	99
II.28	Différence d'énergie libre, $\Delta\mathcal{F}$, à pH 7, entre l'état échantillonné par Monte Carlo et l'état de référence (dans cet état, tous les acides aminés acido-basiques sont neutres), en fonction du nombre de pas Monte Carlo (en noir). Valeur moyenne de $\Delta\mathcal{F}$ en fonction du nombre de pas Monte Carlo (en rouge).	101
II.29	Histogrammes représentant la probabilité des micro-états de protonation en fonction de leur énergie libre associée à différents pH. La valeur $\Delta\mathcal{F}=0$ kcal·mol ⁻¹ a été attribuée au micro-état le plus probable.	102

II.30	En noir, échelle de gauche, nombre de micro-états de protonation nécessaires pour obtenir un échantillon de micro-états de protonation tel que la probabilité que le micro-état de protonation de la protéine appartienne à cet échantillon soit de 0.8. En rouge, valeur du pKa (en abscisse) des acides aminés acido-basiques et, en ordonnée (échelle de droite) leur nombre dans la Cerulean.	103
III.1	Aperçu des géométries $TW\varphi$ (à gauche) et $TW\tau$ (à droite).	108
III.2	Schéma proposé par Martin <i>et al</i> pour représenter la surface d'énergie potentielle de l'état excité fluorescent du chromophore de la GFP. Les flèches rouges et vertes en pointillé représentent des transitions électronique non radiatives. Les flèches noires en pointillé représentent les chemins sur la surface d'énergie potentielle de l'état électronique excité envisagés par Martin <i>et al</i>	109
III.3	Aperçu de l'encombrement stérique du chromophore de la GFP, figure a, et des liaisons hydrogène limitant le mouvement de l'oxygène du phénolate, figure b. . .	110
III.4	a) Histogramme des valeurs de φ obtenues au cours de la simulation de une nanoseconde. b) Valeurs moyennes et écart type des coordonnées φ , τ , pY_{CA2} et pY_{CAG2} qui concerne le pont reliant les deux cycles aromatiques.	114
III.5	Trajectoire moyenne, $\langle \varphi \rangle$ en noir, et trajectoire de l'écart type, σ_φ en rouge, de la valeur absolue de la torsion φ en fonction du temps sur les 1500 trajectoires hors équilibre.	115
III.6	Distribution des temps de premier passage à la géométrie $TW\varphi$ sur l'ensemble des trajectoires.	116
III.7	Courbes $\mathcal{F}_i(\varphi)$ d'énergie libre partielle de la coordonnée φ obtenues pour chaque simulation biaisée. Pour chaque simulation biaisée, seules les valeurs de φ pour lesquelles l'histogramme des valeurs de φ est supérieur à 5% de son maximum ont été reportées sur la figure.	117
III.8	Profil d'énergie libre partielle en fonction de la coordonnée φ en phase aqueuse. La courbe noire en pointillé et obtenue par symétrie par rapport à l'axe passant par $\varphi=-90^\circ$	118
III.9	Déplacement carré moyen en fonction du temps des C_α des acides aminés qui forment une structure secondaire de type hélice α ou feuillet β en noir, de l'ensemble des C_α en rouge et de l'ensemble des atomes de la protéine en orange. 120	
III.10	Réseau de liaisons hydrogène dans l'environnement du chromophore obtenu à partir de la simulation. 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.	121

III.11	a) Histogrammes des valeurs de φ au cours de la simulation. b) Trajectoire de φ sur les 6 ns de la simulation de la GFP avec le chromophore dans son état excité fluorescent.	122
III.12	Histogrammes des valeurs de φ obtenues au cours des 15 simulations biaisées et de la simulation sans contraintes (en noir).	124
III.13	Variation de la valeur moyenne des coordonnées φ (en noir), τ (en rouge), pyr_{CA2} (en vert) et pyr_{CG2} (en bleu) en fonction de la valeur du paramètre φ_0 du potentiel de biais. Les points pour $\varphi_0=0^\circ$ correspondent aux valeurs obtenues pour la simulation sans contraintes. Les barres d'erreur représente les écarts type. Le trait pointillé représente la droite $\langle \varphi \rangle = \varphi_0$	125
III.14	Variation de la moyenne du RMSd et de son écart type (barre d'erreur) en fonction de la moyenne de φ dans chaque simulation biaisée pour (i) les $\text{C}\alpha$ des acides aminés impliqués dans les feuillets β et l'hélice α centrale (en noir), (ii) l'ensemble des carbone α de la protéine (en rouge) et (ii) l'ensemble des atomes des acides aminés de la cage dans laquelle se trouve le chromophore dans la protéine (en bleu) (acides aminés 61, 62, 63, 68, 69, 94, 96, 145, 148, 150, 165, 167, 183, 203, 205, 222, 224). Les points pour $\varphi_0=0^\circ$ correspondent aux valeurs obtenues pour la simulation sans contrainte.	127
III.15	Aperçu du chromophore et des acides aminés et des molécules d'eau qui exercent un couple important sur le phénolate, à gauche dans le cas de $\varphi > 0^\circ$ et à droite dans le cas de $\varphi < 0^\circ$	128
III.16	a) Courbes d'énergie libre partielle en fonction de φ obtenues pour chaque simulation biaisée listée dans le tableau III.2. b) Profil d'énergie libre complet obtenu après a raccordé les différents morceaux de la figure a). c) Comparaison du profil d'énergie libre partielle obtenu par la méthode standard (en noir) et par la méthode de McCammon (en rouge). Les courbes noires en pointillé correspondent à la courbe rouge translatée de $\pm 180^\circ$	129
III.17	a : Histogrammes biaisés des valeurs de φ obtenus pour quatre simulations biaisées (courbes en trait plein). Les courbes en pointillé représentent les potentiels de biais utilisés. La couleur de l'histogramme et du potentiel de biais associé correspondent. b : Courbes $\mathcal{F}_i(\varphi)$ obtenues pour les quatre simulations biaisées dont les histogrammes biaisés sont présentés sur la figure a.	130
III.18	Déplacement de la molécule d'eau wat374 dû à la rotation du cycle phénol. Les points rouge et vert correspondent respectivement aux positions de l'atome d'oxygène de wat374 au cours des fenêtres correspondant à $\varphi_0 = -20^\circ$ et $\varphi_0 = -80^\circ$	132
III.19	Corrélation entre φ et la distance entre l'atome de carbone CA2 et l'atome d'oxygène de wat374 dans la simulation de la GFP sans contrainte (en noir) et dans une simulation biaisée dans laquelle wat374 est éloignée du phénolate.	133

III.20	Représentation, de l'intervalle \mathcal{D} dans lequel une désactivation non radiative de l'état électronique excité par conversion interne est possible (en bleu) et de la position de la géométrie TW φ (en rouge) à partir de laquelle le système peut atteindre une intersection conique et revenir à l'état fondamental de manière non radiative. La courbe noire représente le profil d'énergie libre partielle $\mathcal{F}(\varphi)$	134
III.21	Fonction d'auto-corrélation $C(t)$ de la coordonnée φ sur une trajectoire dans l'ensemble (N,V,E) à l'équilibre (en noir). La courbe rouge est l'intégrale de la fonction $C(t)$ entre 0 et t	140
III.22	Valeur moyenne $\langle \varphi \rangle(t)$ sur une ensemble de trajectoires hors équilibre ayant comme point de départ des configurations dans lesquelles $\varphi \simeq 40^\circ$	141
III.23	Comparaison de la moyenne des trajectoires obtenues par dynamique moléculaire et par dynamique Brownienne avec un coefficient de diffusion de valeur $D = 2 \cdot 10^{11} \text{ rad}^2 \text{ s}^{-1}$	143
III.24	Distribution des temps de premier passage à la géométrie TW φ	143
III.25	Déplacement de la coordonnée φ au cours d'une trajectoire Brownienne suivant que l'on utilise un pas de temps δt de 100 fs ou de 500 fs.	144
III.26	En noir, variation en fonction du pas de temps δt de la largeur de la distribution à partir de laquelle sont calculés les déplacements aléatoires. En rouge, écart type des valeurs de φ dans la simulation de dynamique moléculaire sans contraintes présentées au paragraphe III.3.2.	145
III.27	Temps de résidence, $T(t)$, de la coordonnée φ dans l'intervalle $[60^\circ ; 120^\circ]$ en fonction du temps (courbe bleu). Le cadre en haut à gauche est un agrandissement de la courbe $T(t)$ à temps court. La courbe noire représente la dérivée de $T(t)$ par rapport au temps.	147
III.28	Distributions des temps de n ^{ème} passage obtenues par simulation de dynamique Brownienne. Le cadre en haut à droite est un agrandissement des distributions à temps courts. La courbe en pointillé, représente le déclin de la fluorescence dû à une désactivation purement radiative. Les courbes noires en traits pleins sont des ajustements par moindre carré à l'aide d'une fonction exponentielle multipliée par un polynôme de degré inférieur ou égal à 2.	148
III.29	Convergence de la valeur du rendement quantique de fluorescence ϕ_f en fonction du nombre de distribution de n ^{ème} passage que l'on considère pour une valeur de λ de 1/2.	151
III.30	Valeurs du rendement quantique ϕ_f obtenues en fonction de la probabilité λ . La courbe rouge représente la moyenne des valeurs de ϕ_f sur les valeurs de λ . La courbe verte représente la valeur de ϕ_f dans le mutant de la GFP que nous avons étudié (code 1EMA [18] de la <i>Protein Data Bank</i> . La courbe bleu représente la valeur de ϕ_f dans la GFP sauvage.	152

III.31	a : Déclin de la population de l'état électronique excité obtenu à partir du logarithme décimal de l'équation III.19 pour plusieurs valeurs de λ ($\lambda=0$ et λ variant de 0.25 à 0.5 par pas de 0.05). b : Rapport du déclin de la population de l'état excité obtenu à partir de l'équation III.19 au déclin de la population de l'état excité expérimental qui est monoexponentiel et dont le temps de vie est $\tau_s=3$ ns, pour plusieurs valeurs de λ . Sur les deux figures, les couleurs correspondent aux mêmes valeurs de λ , la courbe noire en pointillé est relative au déclin de fluorescence en absence de processus de désactivation et la courbe noire en trait plein est relative au déclin de fluorescence expérimental.	153
A.1	Représentation de Cramer d'un acide aminé. Le groupement R représente la chaîne latérale.. . . .	157
A.2	Formation de la liaison peptidique (en bleu) entre deux acides aminés.	158
A.3	Géométrie d'un brin peptidique. a) Définitions des angles dièdres ϕ , ψ et ω du squelette protéique. b) Diagramme de Ramachandran présentant les couples (ϕ, ψ) accessibles (en jaune). Les régions rouges correspondent aux conformations les plus stables.	160
A.4	Représentation des structures secondaires en hélice α (à gauche) et en feuillet β (à droite). Les deux brins supérieurs forment un feuillet parallèle, les deux brins inférieurs forment un feuillet antiparallèle. L'image de gauche est issue du site www.chusa.jussieu.fr	161
B.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.	164
B.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 θ entre les atomes 2, 3 et 4. La torsion est mesurée par un angle dièdre ϕ , 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.	165
B.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.. . .	169
B.4	Le critère utilisé pour savoir si une liaison hydrogène existe comporte la distance X-Y et l'angle \widehat{HXY} . Les valeurs les plus souvent utilisées sont telles que d soit inférieur à 3.2 Å et que a soit inférieur à 40°.	173

Liste des tableaux

I.1	Charges effectives, en unité atomique, obtenues pour le chromophore de la GFP dans son état fondamental, S_0 , et dans son premier état excité fluorescent, S_1 . La numérotation des atomes est présentée sur la figure I.6 page 20.	25
I.2	Charges effectives, en ua (fraction de la charge d'un électron), portées par les atomes de la fonction amide d'un acide aminé neutre dans le champ de force <i>ff99</i>	26
I.3	Charges effectives, en unité atomique, obtenues pour le chromophore de la Cerulean en conformation <i>trans</i> et en conformation <i>cis</i> . La valeur du dipôle de la ligne <i>charges</i> correspond au dipôle créé par les charges effectives. La valeur de la ligne RHF correspond au dipôle calculé au niveau RHF/6-31G*. Colonnes <i>ff99</i> : valeurs des charges correspondantes dans le champ de force <i>ff99</i>	28
I.4	Paramètres de l'énergie potentielle de torsion autour des liaisons C12-C11 (angle dièdre φ) et C11-C7 (angle dièdre τ) pour les conformations <i>cis</i> et <i>trans</i> du chromophore. L'angle φ est défini comme l'angle dièdre entre les atomes C12-C11-C7-C8 et l'angle τ est défini comme l'angle dièdre entre les atomes N13-C12-C11-C7.	33
I.5	Valeurs de la constante de force k et de la position du minimum ϕ_0 de la contrainte harmonique utilisée, de la forme $k(\phi - \phi_0)^2$, pour les coordonnées τ , pyr_{CA2} et pyr_{CG2} . L'angle de torsion τ est caractérisé par les angles dièdres N2-CA2-CB2-CG2 et C2-CA2-CB2-CG2. Les pyramidalisations pyr_{CA2} et pyr_{CG2} sont données par les angles dièdres impropres N2-C2-CA2-CB2 et CD1-CD2-CG2-CB2.	34
II.1	Liste des 20 micro-états de protonation de plus basse énergie à pH 5. Chaque ligne représente un micro-état dans lequel 0 signifie que l'acide aminé est dans sa forme basique (non protoné) et 1 signifie qu'il est dans sa forme acide (protonée). Chaque colonne correspond à un acide aminé. Ils sont donnés par numéro croissant : Glu6, Glu17, Asp19, Glu32, Glu34, His77, Glu90, Glu95, Asp102, Glu115, Asp117, Glu124, Asp129, Glu132, Asp133, His139, Glu142, Asp148, Glu172, Asp173, Asp180, Asp190, Asp197, His217. La dernière colonne est l'énergie libre associée au micro-état en $\text{kcal} \cdot \text{mol}^{-1}$	58

II.2	Liste des 20 micro-états de protonation de plus basse énergie à pH 8. Chaque ligne représente un micro-état dans lequel 0 signifie que l'acide aminé est non protoné et 1 signifie qu'il est protoné. Chaque colonne correspond à un acide aminé. Ils sont donnés par numéro croissant : Asp19, His25, Tyr39, Lys101, Glu142, Asp148, Lys158, His199. La dernière colonne est l'énergie libre associée au micro-état en kcal · mol ⁻¹	60
II.3	Valeur moyenne et déviation standard (entre parenthèses) des torsions (en degrés) et des distances (en Å) qui permettent de caractériser les conformations de la chaîne latérale de Asp148 obtenues le long des quatre simulations effectuées à pH 5. Les noms des atomes et les torsions sont définis sur la figure II.6. Les données de la ligne <i>référence</i> sont extraites de la structure RX de la Cerulean. Sur cette ligne l'atome d'oxygène ODH est défini comme étant celui qui est supposé porter un proton dans la structure RX.	65
II.4	Distances, en Å, entre l'atome d'oxygène du carbonyle du squelette protéique et l'atome d'azote de la fonction amide des paires d'acides aminés constituant les brins 6 et 7, (figure II.10 b page 69). Chaque paire d'acides aminés peut former deux liaisons hydrogène (figure A.4 page 161). La première valeur est la distance entre l'atome d'azote du premier acide aminé et l'atome d'oxygène du second acide aminé, la seconde valeur est la distance entre l'atome d'oxygène du premier acide aminé et l'azote du second acide aminé. Les lignes 1 à 6 concernent les distances mesurées dans la structure RX de la référence indiquée. Les valeurs de la ligne 7 sont les moyennes obtenues pour chaque distance dans la simulation 1.	70
II.5	Contacts entre protéines dans le cristal de la Cerulean (structure RX code 2Q57) entre des atomes de la protéine centrale (colonne 1) et des atomes d'une protéine voisine dans le cristal (colonne 2) obtenue par l'opération de symétrie donnée en colonne 4. La distance entre les atomes (colonne 3) est donnée en Å. Les noms des atomes sont ceux indiqués dans la structure RX.	76
II.6	Plus courtes distances dans le cristal de la GFP, en Å, entre des atomes de la protéine centrale (colonne 1) et des atomes d'une protéine voisine dans le cristal (colonne 4) par l'intermédiaire de molécules d'eau cristallographiques (colonne 2). La protéine voisine est obtenue par l'opération de symétrie (x , $-y + b$, $-z + c$). Les noms des atomes sont ceux indiqués dans la structure RX de la GFP, entrée IEMA de la <i>Protein Data Bank</i>	77
II.7	Valeur moyenne et déviation standard (entre parenthèse) des torsions (en degrés) qui permettent de caractériser les conformations de la chaîne latérale de Asp148 obtenues pour les simulation pH5- <i>cis</i> (simulation 1), pH8- <i>cis</i> (sur les trois dernières nanosecondes) et pH8- <i>trans</i> . Les noms des atomes et les torsions sont définis sur la figure II.6 page 63. Les données de la ligne <i>référence</i> sont extraites de la structure RX de la Cerulean.	80

II.8	Valeur moyenne de $\Delta\lambda_{max}$ et incertitude $u(\Delta\lambda_{max})$ compte tenu de l'effet de la taille de l'échantillon utilisé et de la fonctionnelle de la densité utilisée.. . . .	95
II.9	Valeur moyenne et fluctuations (entre parenthèses) de la force d'oscillateur associée au deux premiers états excités en fonction des différentes situations envisagées concernant le pH et la conformation du chromophore.	96
III.1	Valeur moyenne et écart type des coordonnées φ , τ , pyr_{CA2} et pyr_{CG2} qui concerne le pont reliant les deux cycles aromatiques du chromophore.	120
III.2	Constante de force, k en $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$, et position du minimum, φ_0 du potentiel de biais employé dans les simulations biaisées utilisées pour construire le profil d'énergie libre en fonction de φ dans la GFP. La durée correspond au temps exploité dans chaque simulation.	123
III.3	Valeurs moyennes des couples Γ ($\text{kcal}\cdot\text{mol}^{-1}\cdot\text{degre}^{-1}$) exercé sur le cycle phénolate par les acides aminés proches du chromophore dans une sélection de simulations biaisées caractérisées par la valeur φ_0 du minimum du potentiel de biais. $\langle \varphi \rangle$ représente la valeur moyenne de φ dans ces simulations. Les lignes intitulées "total" correspondent au couple total exercé par l'ensemble des acides aminés de la protéine et des molécules d'eau cristallographiques. La ligne "chromophore" correspond à la contribution du potentiel intramoléculaire du chromophore au couple total.. . . .	126
III.4	Expressions analytiques employées pour modéliser les distributions des temps de $n^{\text{ème}}$ passage et les valeurs des paramètres de ces expressions obtenues par une méthode de moindres carrés.	150
A.1	Tableau regroupant les vingt acides aminés principaux ainsi que leur nom, les codes à une lettre et à trois lettres utilisé pour simplifier l'écriture et, s'il y a lieu, le pKa de la chaîne latérale.. . . .	159

Résumé :

Ce manuscrit présente une étude théorique de protéines de la famille de la *Green Fluorescent Protein*, GFP. Ces protéines permettent grâce à leur propriétés de fluorescence 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.

La première partie de ce manuscrit présente l'étude par simulation moléculaire de l'effet du changement de pH sur la structure de la Cerulean et sur son spectre d'absorption. Ces calculs nous ont permis d'établir que le décalage du spectre d'absorption, observé expérimentalement en fonction du pH, est dû à une isomérisation du chromophore liée au changement de l'orientation de la chaîne latérale d'un acide aminé proche.

La deuxième partie de ce manuscrit aborde l'étude d'un mécanisme de désactivation de la fluorescence dans la GFP. Nous avons proposé une approche, combinant des simulations de dynamique moléculaire biaisée et de dynamique brownienne, afin de déterminer la cinétique d'un mécanisme de désactivation de la fluorescence lié à une torsion du chromophore. Nous avons pu obtenir des distributions de temps de n^{ème} passage aux géométries critiques et en déduire des informations quantitatives sur le déclin de fluorescence.

Les outils développés et leurs futurs développements permettront de progresser dans la compréhension de la relation entre l'isomérisation du chromophore, le pH et le déclin de la fluorescence qui sont étroitement liés dans les protéines fluorescentes.

Abstract :

This Phd presents a theoretical study of proteins of the Green Fluorescent Protein family. These proteins allow to obtain new insight into a lot of in vivo biological processes. However, their photophysical properties are not always well controlled. Theoretical studies are complementary to experimental studies. It allows us to obtain a better understanding of microscopic processes and to contribute to the interpretation of photophysical properties of these proteins.

The first part of this Phd deals with a molecular dynamics study of the effect of the pH on the structure and the absorption spectra of the Cerulean. This study demonstrates that the shift of the absorption spectra experimentally observed is due to a chromophore isomerisation linked to a motion of the side chain of an amino acid close to the chromophore.

The second part of this Phd focuses on the study of the fluorescent decay of the Green Fluorescent Protein. We have developed an approach involving biased molecular dynamics simulation and Brownian dynamics simulations in order to determine the kinetics of a quenching mechanism. This approach allowed us to get the nth passage time distributions at critical geometries and to compute quantitative data of the fluorescent decay.

The tools we have developed in this Phd and their future improvements will allow to go on in the understanding of the relation between chromophore isomerisation, pH and fluorescent decay, which are strongly related in fluorescent proteins.