Université Paris-Sud
École Doctorale : Chimie de Paris-Sud
Laboratoire de Chimie Physique

Étude théorique de l'extinction de fluorescence des protéines fluorescentes : champ de forces, mécanisme moléculaire et modèle cinétique

thèse de doctorat
soutenue le 18/07/2012
par

Gabriella Jonasson

Composition du jury

Oliver Nüsse	*Président du jury*	Professeur, Université Paris-Sud
Nathalie Reuter	*Rapporteur*	Group Leader, Universitetet i Bergen, Norvège
Serge Antonczak	*Rapporteur*	Professeur, Université Nice-Sophia Antipolis
David Perahia	*Examinateur*	Directeur de Recherche, ENS de Cachan
Rodolphe Vuilleumier	*Examinateur*	Professeur, ENS de Paris
Isabelle Demachy	*Directeur de thèse*	Professeur, Université Paris-Sud
Fabienne Mérola	*Membre invité*	Directeur de Recherche, Université Paris-Sud

Résumé

Les protéines fluorescentes, comme la GFP (green fluorescent protein), sont des protéines naturellement fluorescentes qui sont utilisées pour leur rôle de marqueur, permettant de localiser des protéines dans les cellules et d'en suivre les déplacements. De nombreuses études expérimentales et théoriques ont été menées ces dix dernières années sur les protéines fluorescentes. De là, se forge une compréhension essentiellement qualitative du rôle de la protéine vis-à-vis de l'obtention ou non d'une émission radiative : il apparaît que la protéine permet la fluorescence en bloquant les processus qui la désactivent ; ces processus de désactivation sont très rapides et efficaces (à l'échelle de la picoseconde) dans le cas du chromophore seul, et ils sont bien identifiés comme étant des torsions autour des liaisons intercycles (τ et φ). Dans la protéine, la sensibilité des temps de vie de fluorescence à des mutations proches ou non du chromophore, à des modifications de pH ou de température laisse supposer un contrôle de la dynamique du chromophore par différents paramètres, sans qu'ils soient pour autant identifiés et mis en relation.

Une étude de la dynamique de la protéine permettrait de faire la lumière sur les mécanismes responsables de ces phénomènes photophysiques pour lesquels une analyse structurale ne suffit pas. Cependant l'étude de la dynamique est limitée par la taille du système (>30 000 atomes), par l'échelle de temps des phénomènes photophysiques considérés (dizaine de nanosecondes) et par le fait que les deux torsions τ et φ sont fortement couplées dans l'état excité du chromophore. Ces trois facteurs excluent les méthodes de dynamique existantes aujourd'hui ; dynamique quantique (AIMD), dynamique mixte classique-quantique (QM/MD) et dynamique moléculaire classique (MD).

Nous avons surmonté le problème par la modélisation de la surface d'énergie potentielle de torsion du chromophore à l'état excité basée sur des calculs quantiques de haute précision, par une interpolation des valeurs obtenues par une expression analytique appropriée en fonction des angles de torsion τ et φ et avec une précision suffisante pour reproduire des barrières de l'ordre de la kcal/mol, et enfin, par l'implémentation de cette expression analytique dans le programme parallèle AMBER. Une deuxième difficulté théorique concerne la simulation et l'analyse statistique d'événements peu fréquents à l'échelle de la nanoseconde, et dont on ne connaît pas le chemin de réaction, ici les déformations de la protéine et du chromophore conduisant aux géométries favorables à la conversion interne.

Grâce à ces développements et aux simulations qu'ils ont permises, nous avons réalisé la première modélisation de la désactivation non-radiative par conversion interne à l'échelle de la nanoseconde dans trois protéines fluorescentes différentes. L'analyse des dynamiques moléculaires classiques nous donne une évaluation quantitative des temps de vie de l'extinction de fluorescence, en accord avec les données expérimentales. Par ailleurs elle nous a permis

d'identifier les mouvements moléculaires concertés de la protéine et du chromophore conduisant à cette extinction.

De ces résultats, émerge une représentation plus complète du mécanisme qui libère la torsion du chromophore ou qui la déclenche : il peut venir d'un mouvement spécifique de la protéine, qui se produit à l'échelle de la nanoseconde, ou bien de plusieurs mouvements spécifiques, plus fréquents (rupture de liaisons hydrogène, rotation de chaînes latérales, dynamique d'agrégats d'eau), mais qui coïncident seulement à l'échelle de la nanoseconde. Ces mouvements spécifiques n'ont pas un coût énergétique important mais la nécessité de leur coïncidence crée un délai de l'ordre de quelques nanosecondes alors que dans le vide la torsion se produit en quelques picosecondes. Dans le cas des protéines étudiées, on a identifié en grande partie les mécanismes et les acides aminés qui sont impliqués.

Summary

Fluorescent proteins, like GFP (green fluorescent protein), are efficient sensors for a variety of physical-chemical properties and they are extensively used as markers in living cells imaging. These proteins have been widely studied both experimentally and theoretically the last decade. The comprehension of the protein's role in the regulation of the radiative emission is today essentially qualitative : it appears that the protein enables the fluorescence by blocking the processes that deactivates it ; the deactivating processes are very quick and efficient (on the picosecond time scale) when the chromophore is isolated, and they are identified as being the torsions around the central bonds of the chromophore (τ and φ). The fluorescence lifetimes of a protein is very sensitive to mutations in the vicinity of the chromophore, to modifications in pH or in temperature. This seems to indicate a control of the dynamics of the chromophore by different parameters, that are not necessarily identified.

A study of the dynamics of the protein would allow a deeper understanding of the mechanisms that are responsible for the fluorescence quenching. From a theoretical point of view, one is faced with three difficulties in this type of study : the size of the system (>30 000 atoms including a water box), the required time scale (tens of nanoseconds) and the fact that the torsions τ and φ are strongly coupled in the excited state of the chromophore. We must thus rule out the already existing dynamics methods : quantum dynamics (AIMD), mixed classical-quantum dynamics (QM/MD) and classical molecular dynamics (MD).

We have overcome this problem by modeling the torsional potential energy surface of the chromophore in the first excited state trough high precision quantum calculations, by interpolating the energy values with an analytical fitting expression depending on the torsions τ and φ and with a precision high enough to reproduce barriers of the order of 1 kcal/mol, and lastly, by implementing this fitting expression in a parallelized version of the MD program AMBER. Another theoretical difficulty concerns the simulation and the statistical analysis of rare events on the nanosecond time scale without knowing the reaction path in advance, i.e. the deformations of the protein and of the chromophore leading to geometries where the internal conversion is favored.

As a result of these developments and of the simulations they have enabled, we have been able to model, for the first time, the non-radiative deactivation by internal conversion at the nanosecond time scale in three different fluorescent proteins. The analysis of the classical molecular dynamics gives us a quantitative evaluation of the lifetime of the fluorescence extinction, in agreement with experimental results. In addition, it has allowed us to identify the concerted molecular movements between the protein and the chromophore leading to this extinction.

A more complete representation of the mechanism that liberates or provokes the chromophore torsion emerges from these results : it could be a specific movement of the protein, that occurs on

the nanosecond timescale, or several specific movements that occur more frequently (breakage of a hydrogen bond, rotation of side chains, dynamics of a water cluster), but that coincide only on the nanosecond time scale. These specific movements do not have a high energy cost but the need for them to coincide creates a delay of several nanoseconds compared to the chromophore torsion in vacuo which occurs after a few picoseconds. In the proteins we have studied (GFP, YFP and Padron), we have identified the principle components of the mechanisms and the amino acids that are implicated in this chromophore-protein interplay.

Remerciements

En préambule de ce mémoire, je souhaite adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de ces formidables trois années de doctorat.

En premier lieu, j'adresse de sincères remerciements à Mehran Mostafavi, directeur du laboratoire de Chimie Physique, ainsi qu'à David Lauvergnat, animateur du groupe ThéoSim, pour m'avoir accueilli chaleureusement au sein du laboratoire dans le cadre d'un contrat doctoral.

Deuxièmement, je remercie les rapporteurs de cette thèse, Nathalie Reuter et Serge Antonczak, pour l'intérêt qu'ils ont porté à mon travail. Merci également aux autres membres du jury qui ont accepté d'examiner ce travail : Oliver Nüsse, David Perahia, Rodolphe Vuilleumier et Fabienne Mérola.

Mes remerciements s'adressent également aux personnes qui m'ont proposé le sujet de thèse et qui m'ont encadrée tout au long de ces années de thèse : Isabelle Demachy, Jacqueline Ridard et Bernard Lévy. Je leur suis profondément reconnaissante pour le temps conséquent qu'ils m'ont accordé, leurs conseils et encouragements ainsi que leurs qualités sur le plan scientifique, pédagogique et humain. J'ai énormément appris à leurs côtés et j'ai pris un grand plaisir à travailler avec eux.

Le laboratoire de Chimie Physique est un cadre privilégié pour effectuer une thèse. Parmi ceux qui ont contribué à mes réflexions, je remercie tout spécialement Pascal Pernot, Aurélien de la Lande, Fabien Cailliez et David Lauvergnat.

Un grand merci à Jean-Marie Teuler pour l'implication dans la partie informatique de ce travail, j'ai beaucoup apprécié son enthousiasme, sa sympathie et son humour.

Je remercie également Pierre Archirel, avec qui j'ai pu partagé le bureau, pour sa gentillesse, ses réponses à mes nombreuses questions.

Je désire en outre remercier toutes les personnes du troisième étage (incluant ceux précédemment cités) pour leur accueil chaleureux et leur soutien amical. Merci aux adeptes de la cafétéria du LCP pour avoir rendu la pause déjeuner si plaisante avec des sujets de discussion divers et variés, des fous rires et plein de chocolat. Merci aussi à Fabien et Aurélie pour votre compagnie fort sympathique aux conférences où nous sommes allés ensemble.

Encore un merci à Hélène Pasquier, Fabienne Mérola, Marie Erard et Asma Fredj du groupe Biophysique, pour toutes les discussions captivantes dans le cadre de l'axe transversal P3F.

J'ai eu la chance tout au long de ces trois années d'être aidée par des personnes extrêmement efficaces et toujours souriantes de l'administration et de la logistique qui m'ont rendu la vie

plus facile. Merci beaucoup donc à Séverine Bourguignon, Marie-Françoise Lecanu, Delphine Lelièvre, Anne Morel et Joël François.

Un grand merci à mes amis, en particulier Aude et Céline avec qui j'ai pu partager les moments difficiles et les moments heureux pendant ma thèse. Bientôt à vous les phi !

Ma reconnaissance va aussi à ceux qui ont plus particulièrement assuré le soutien affectif de ce travail doctoral : ma famille ainsi que la famille Donon. Je souhaite remercier spécialement Damien pour son soutien, sa patience et sa confiance en moi tout au long de la thèse.

Slutligen vill jag tacka min älskade familj, Kristoffer, Mikaela, Lena och Håkan. Utan deras stöd och uppmuntran skulle jag troligen inte sitta och skriva det här idag.

Table des matières

I. Introduction générale **13**

II. Protéines fluorescentes **17**

II.1. Historique . . . 17
II.1.1. Découverte et caractérisation de la protéine fluorescente verte . . . 17
II.1.2. La famille des protéines fluorescentes . . . 19
II.2. Photophysique . . . 21
II.2.1. Phénomènes photophysiques . . . 21
II.2.2. Caractérisation de la fluorescence . . . 22
II.2.3. Photophysique des protéines fluorescentes . . . 23
II.2.4. L'influence de l'environnement . . . 25
II.2.5. Mesures expérimentales . . . 25
II.3. Protéines étudiées . . . 25
II.3.1. La GFP . . . 25
II.3.2. La YFP . . . 26
II.3.3. Padron . . . 26

III. Potentiel du chromophore - NADiA **29**

III.1. Contexte . . . 29
III.2. Calculs *ab initio* . . . 30
III.2.1. La méthode . . . 30
III.2.2. Les géométries . . . 33
III.3. Interpolation de la surface d'énergie potentielle . . . 34
III.3.1. Définitions des angles dièdres τ et φ . . . 34
III.3.2. Procédure d'interpolation . . . 36
III.3.3. La surface d'énergie potentielle . . . 39
III.4. Implémentation de NADiA dans Amber . . . 41
III.4.1. Algorithme du potentiel NADiA . . . 41
III.4.2. Traitement des interactions entre atomes non liés . . . 43
III.4.3. Lecture et partage de données . . . 44
III.5. Simulations dans le vide . . . 44
III.6. Analyse d'incertitude . . . 47
III.7. Potentiel de l'imidazolinone . . . 48
III.7.1. Procédure . . . 48
III.7.2. Simulation DFT/MD . . . 49
III.7.3. Simulation MD avec un champ de forces standard . . . 50

III.7.4. Calibration du potentiel 51
III.8. Autres paramètres du champ de force 52
III.8.1. Acide aminé non-standard 52
III.8.2. Angles dièdres 53
III.8.3. Angles de valence 54
III.8.4. Charges effectives 55
III.9. Conclusion du chapitre 60

IV. Mécanismes moléculaires de torsion du chromophore dans les protéines fluorescentes 61
IV.1. Analyse de la GFP 61
IV.1.1. L'environnement du chromophore 61
IV.1.2. Dynamique moléculaire à l'état fondamental 63
IV.1.3. Dynamique moléculaire à l'état excité 65
IV.1.3.1. Analyse des torsions en φ 67
IV.1.3.2. Analyse des torsions en τ 77
IV.1.4. Étude de mouvements collectifs de la protéine 80
IV.1.5. Conclusion 83
IV.2. Analyse de la YFP 84
IV.2.1. L'environnement du chromophore 84
IV.2.2. Dynamique moléculaire à l'état fondamental 85
IV.2.2.1. Le π- stacking entre le chromophore et Tyr203 88
IV.2.3. Dynamique moléculaire à l'état excité 89
IV.2.3.1. Analyse du mécanisme moléculaire des torsions τ 91
IV.2.3.2. Analyse des torsions en φ 99
IV.2.4. Conclusion 100
IV.3. Analyse de Padron 101
IV.3.1. L'environnement du chromophore 101
IV.3.2. Dynamique moléculaire à l'état fondamental 103
IV.3.3. Dynamique moléculaire à l'état excité 105
IV.3.4. Possibilité d'isomérisation 108
IV.3.4.1. Potentiel interne du chromophore 108
IV.3.4.2. Les contraintes de la protéine 109
IV.3.5. Conclusion 110

V. La durée de vie non-radiative 111
V.1. Caractéristiques de la transition $S_1 \rightarrow S_0$ dans les géométries perpendiculaires . . 111
V.1.1. La force d'oscillateur $S_1 \rightarrow S_0$ 111
V.1.2. Données sur la conversion interne $S_1 \rightarrow S_0$ du chromophore 112
V.2. Principes de l'analyse des dynamiques moléculaires de la protéine à l'état excité . 114
V.3. Temps moyen de premier passage et temps de vie non-radiatif 115
V.3.1. Évaluation directe du MFPT par simulation 117
V.3.2. Évaluation du MFPT à partir du nombre de FPT obtenus par simulation 119
V.4. Détermination de τ_{nr} de la GFP 120
V.4.1. Torsion du chromophore 120

V.4.2. Calcul de τ_{nr} . . . 121
V.5. Détermination de τ_{nr} de la YFP . . . 122
V.5.1. Torsion du chromophore . . . 122
V.5.2. Calcul de τ_{nr} . . . 123
V.6. Conclusion du chapitre . . . 125

VI. Conclusion générale et perspectives **127**

A. Acides Aminés **139**

B. Interpolation **140**

C. Simulation de dynamique moléculaire **142**
C.1. Les lois du mouvement de Newton . . . 142
C.2. Mécanique Statistique . . . 142
C.3. Des conditions aux limites . . . 143
C.4. Contrôle de la température et de la pression (thermostat et barostat) . . . 144
C.5. La méthode Particle Mesh Ewald . . . 144
C.6. Préparation de la structure initiale . . . 145
C.6.1. La structure cristalline . . . 145
C.6.2. Compléter la structure . . . 146
C.6.3. Minimisation et mise à l'équilibre du système . . . 147
C.6.4. Dynamique moléculaire : phase de production . . . 148
C.7. Champ de forces de Amber et de CHARMM . . . 149

D. NMA **151**
D.1. Les modes normaux de vibration . . . 151
D.2. Application sur la GFP . . . 152

I. Introduction générale

La protéine fluorescente verte (GFP) (Fig. I.1) été découverte au début des années 1960 dans la méduse *Aequorea victoria*. Il existe aujourd'hui une large palette de protéines fluorescentes (FPs), issues d'organismes marins (FPs sauvages) ou de mutagenèse (FPs mutants), dont la fluorescence s'étend du proche IR au proche UV [1, 2, 3]. Un grand nombre de FPs ont été identifiées dans des organismes marins appartenant à l'embranchement *Cnidaria*, e.g. méduses et corails. Des FPs ont également été découvertes dans d'autres organismes marins, e.g. crustacés [4] et céphalochordés [5], appartenant aux embranchements *Arthropoda* et *Chordata*, respectivement. Cela signifie que le gène de la famille de FP a existé dans un ancêtre commun à ces embranchements.

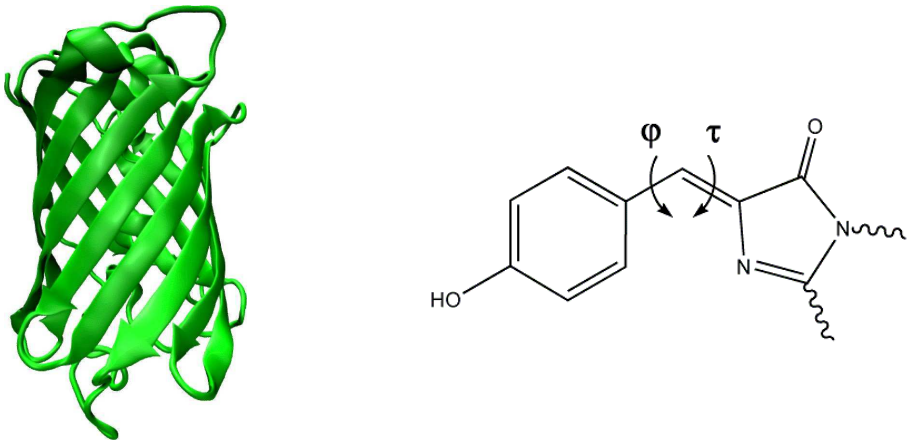


FIG. I.1.: La structure globale de la GFP (à gauche) et le chromophore (à droite) sous la forme neutre *cis*. Les angles dièdres autour des deux liaisons centrales sont appelés φ et τ. Les lignes ondulées représentent les deux jonctions covalentes avec le reste de la protéine.

L'utilité des FPs pour ces divers organismes n'est pas élucidée. Les hypothèses les plus répandues proposent que les FPs peuvent être impliquées dans la pigmentation de l'organisme porteur, la protection contre la lumière du soleil [6], la protection contre le stress oxydant [7], la photosynthèse [6] ou la photoreception [5].

Par contre, l'utilisation des FP au laboratoire a un intérêt majeur évident, ce qui explique les nombreux travaux qui ont permis de trouver ou fabriquer de nouvelles variantes. Ainsi les FPs sont des marqueurs et des senseurs très répandus en imagerie cellulaire [1, 8, 9, 10, 11]. Le chromophore, responsable de la fluorescence, est formé spontanément dans les cellules qui expriment le gène [12]. En tant que marqueurs, les FPs permettent d'étudier la division

cellulaire, les interactions protéine-protéine, la structure et le fonctionnement d'une cellule [13]. Leur fluorescence peut également être utilisée dans la microscopie à très haute résolution (nanoscopie) [14] ou pour sonder l'environnement afin de donner des mesures précises sur le pH [15, 16, 17], la concentration de Ca^{2+} [18, 19] et l'activité redox intracellulaire [20].

Cependant, l'utilisation expérimentale des FPs est souvent limitée par différents effets : dépendance au pH, à la force ionique, polymérisation, blanchiment *etc.* Il y a ainsi en permanence une recherche de nouveaux mutants permettant de limiter ces inconvénients [21]. Parallèlement aux techniques d'essais systématiques ('screening'), la recherche de l'origine de l'effet des mutations sur la fluorescence peut être utile (d'ailleurs des rationalisations, au moins partielles, de ce type sont également indispensables dans le cadre du 'screening'). Cette nécessité est la première motivation de ce travail : déterminer les acides aminés qui ont une influence sur la fluorescence et par quels mécanismes, afin de guider la recherche de nouveaux mutants.

Un autre aspect de l'intérêt théorique des FPs vient de leur mode de fonctionnement : grâce à la technologie laser il est possible d'obtenir des spectres de fluorescence résolus en temps avec une grande résolution temporelle (nanoseconde ou inférieure) et d'avoir ainsi une information qui atteint presque l'étape élémentaire des processus en jeux, ici le processus responsable de l'extinction de fluorescence et de la limitation du rendement quantique. Cette possibilité distingue les FPs des autres protéines, en particulier des enzymes, pour lesquelles la résolution temporelle, celle qu'on obtient par exemple avec la technique du 'stopped flow', est supérieure à la microseconde ce qui donne des informations utiles mais beaucoup plus globales que celles qu'on obtient avec les FPs. Il semble que seules les protéines à transfert d'électron permettent aussi d'approcher l'étape élémentaire.

Dans ces deux cas (FPs et transfert d'électron) on dispose ainsi d'informations permettant une comparaison directe avec les approches théoriques de la dynamique interne des protéines. De fait, le milieu protéique a de nombreux aspect très spécifiques ; en particulier, ce n'est ni un liquide (pas assez organisé), ni un cristal liquide (trop organisé), ni un polymère (pas d'information sur des monomères particuliers) *etc* ... On dispose déjà de nombreuses approches décrivant ce milieu. Les FPs (et les protéines à transfert d'électron) permettent de les approfondir de façon notable. Et on peut espérer que cet approfondissement sera transférable à la dynamique des protéines en général.

La fluorescence des FPs dans le domaine visible est émise par un groupement chimique, formé au sein de la protéine après le repliement. Ce groupement, appelé chromophore, provient de la cyclisation autocatalytique de trois acides aminés consécutifs. Il est situé à l'intérieur de la protéine et composé de deux cycles aromatiques liés entre eux par un groupement CH (Fig. I.1). Après l'excitation du chromophore par irradiation lumineuse il peut se désactiver par émission de fluorescence mais également par des processus non-radiatifs. Une torsion du chromophore peut aboutir à une désexcitation non-radiative et ainsi supprimer la fluorescence. En effet, la rotation de 90° autour d'une des deux liaisons centrales du chromophore fait tomber la force d'oscillateur de la transition radiative vers l'état fondamental à zéro [93, Table 1]. Le chromophore peut ensuite regagner l'état fondamental par conversion interne.

Au sein d'une protéine, les mouvements du chromophore sont contraints par des empêchement stériques, des interactions électrostatiques et des liaisons hydrogène. La désactivation non-

radiative de l'état excité du chromophore, liée aux torsions φ et τ, est donc limitée. Par conséquent, la fluorescence du chromophore est favorisée par les contraintes de la protéine [22]. La comparaison entre le rendement quantique (= nombre de photons émis / nombre de photons absorbés) de la GFP (0,8 [23]) et celui du chromophore isolé dans l'eau ($<10^{-3}$ [24]) illustre bien ce phénomène.

Malgré les contraintes qui peuvent être imposées par la protéine, dans une partie des cas le chromophore réussit à se tordre. Le chemin conformationnel qu'il emploie n'est pas bien caractérisé. Il peut s'agir d'une simple torsion autour de l'une ou de l'autre des liaisons du pont (φ ou τ) ou d'une concertation entre les deux. Afin de pouvoir proposer des améliorations structurales qui favoriseraient la fluorescence d'une protéine, il faut comprendre l'origine structurale et dynamique de la torsion du chromophore au sein de cette protéine.

On voit ainsi que le mécanisme non-radiatif de l'extinction de la fluorescence se situe à trois échelles de temps différentes : le changement d'état électronique lui-même qui est une étape non Born-Oppenheimer dont la durée est limitée par le processus de décohérence [25] à quelques femtosecondes ; la dynamique de torsion du chromophore qui prend quelques picosecondes ; et la dynamique de la protéine qui aboutit à libérer la torsion du chromophore au bout de quelques nanosecondes. On ne cherche pas ici à décrire en détails ces trois échelles de temps mais seulement les aspects de chacune qui interviennent dans la description du processus global.

Après un bref rappel des propriétés des FPs étudiées ici (chapitre II) on abordera la partie principale du présent travail qui a été consacrée à la description de la dynamique du chromophore à l'état excité et en interaction avec la protéine (les échelles de temps pico- et nanosecondes). La difficulté dans ce cas vient de ce qu'il est difficile d'utiliser les méthodes mixtes classique-quantiques (QM/MD) à cause du nombre d'atomes (quelques 10^4 pour la protéine avec une boîte d'eau) et de l'échelle de temps (quelques nanosecondes). Par ailleurs les champs de force standard intégrés dans les programmes de dynamique moléculaire ne sont pas non plus adaptés ici, d'une part, parce qu'on veut décrire un état excité et, d'autre part, parce que leur forme algébrique (une somme de termes correspondants aux différentes coordonnées internes) suppose implicitement un chemin de réaction où les différentes coordonnées telles que les torsions φ et τ peuvent se produire indépendamment alors même qu'il y a des évidences que ce n'est pas le cas [26]. Nous avons donc déterminé un potentiel de torsion à deux dimensions (2D), τ et φ, par des calculs de chimie quantique du premier état excité du chromophore. Par ailleurs, les champs de force implémentés dans les programmes de dynamique moléculaire (MD) standard n'étant pas adaptés à un potentiel de torsion 2D, nous avons implémenté un nouvel algorithme dans le programme de MD Amber. Ces résultats sont détaillés dans le chapitre III.

Les résultats des dynamiques obtenues à l'aide du potentiel 2D NADiA nous ont permis de développer un modèle cinétique de l'extinction de fluorescence directement comparable à la cinétique expérimentale. Ce modèle permet donc un test du potentiel utilisé. Par ailleurs le développement de ce modèle nécessite différents changements d'échelle : l'utilisation d'informations sur le processus non Born-Oppenheimer (transfert femtoseconde - nanoseconde) et l'extrapolation d'un petit nombre d'évènement observés (les torsions du chromophore) à un nombre suffisant pour permettre une statistique (traitement d'un 'évènement rare'). Un outil particulièrement efficace de ce point de vue sera d'évaluer les incertitudes sur les grandeurs

à transférer d'une échelle à l'autre. Le chapitre V est consacré à la construction du modèle cinétique permettant d'obtenir ce type de données, notamment dans le cas où les évènements étudiés sont rares par rapport au temps de simulation.

On peut aussi utiliser les résultats des dynamiques de façon plus qualitative en recherchant quels sont les mouvements de la protéine et/ou du chromophore qui aboutissent à une torsion en un temps environ 1000 fois plus long dans la protéine que dans l'eau ou dans le vide. Dans un premier temps on a cherché à identifier les acides aminés qui font obstacle à la torsion et les mouvements qui libèrent cette dernière. On a aussi cherché à relier ces mouvements d'acides aminés individuels aux mouvements collectifs de la protéine. Mais il apparaît aussi que l'effet des mouvements de la protéine sur la fluorescence ne peuvent pas être compris indépendamment de ceux du chromophore et que ces derniers ne consistent pas toujours en une simple torsion mais impliquent plusieurs degrés de liberté. Cette problématique est directement reliée à celle de la recherche de mutations permettant d'améliorer les propriétés physico-chimiques des FPs. D'un autre coté on verra qu'elle permet de mettre en évidence des aspects de la dynamique qui sont probablement transférables à d'autres types de protéines. Ces discussions font l'objet du chapitre IV.

II. Protéines fluorescentes

II.1. Historique

II.1.1. Découverte et caractérisation de la protéine fluorescente verte

La protéine fluorescente verte (GFP) est la première protéine fluorescente (FP) qui a été découverte au début des années 1960 lors d'une étude de la méduse *Aequorea victoria* (*Av*) menée par Osamu Shimomura [27]. La lumière verte produite par cet organisme marin vient d'un couplage de deux protéines. La première, l'aequorine [27], catalyse une réaction d'oxydation provoquant une émission de lumière bleue qui excite la deuxième protéine, la *Av*GFP, qui à son tour émet de la fluorescence verte [28, 29, 30].

Une avancée décisive pour les applications d'aujourd'hui a été réalisée par Douglas Prasher en 1992 lorsqu'il a réussi à isoler le code génétique de la protéine [31]. Cela a permis la reproduction de la *Av*GFP dans des nouveaux organismes. De nombreux groupes ont commencé à travailler sur l'expression du gène *in vivo* [32, 33, 34, 35, 29]. Un des premiers groupes, mené par Martin Chalfie, a exprimé la *Av*GFP dans la bactérie *Escherichia coli* et dans le ver *Caenorhabditis elegans* [12]. Ces résultats ont montré que le repliement de la protéine et la formation du chromophore ne dépendaient pas d'une enzyme ou d'un cofacteur spécifique à son hôte d'origine. Cependant, il a rapidement été montré que la formation du chromophore nécessite de l'oxygène [36, 37].

Shimomura et al. ont été les premiers à caractériser la structure du chromophore [38]. Il s'agit d'un hydroxybenzylideneimidazolinone (Fig. I.1) formé à partir de trois acides aminés Ser65-Tyr66-Gly67 [39, 31, 36, 40, 41]. La réaction se déroule en trois étapes (Fig. II.1) : 1) un cycle imidazolone est formé par une attaque nucléophile de l'azote peptidique de Gly67 sur le carbonyle de Ser65 [42] 2) le cycle est déshydraté pour former un cycle imidazolinone 3) la liaison $C_\alpha - C_\beta$ de Tyr66 est oxydée par du dioxygène [36, 37]. Un système conjugué, étendu sur deux cycles aromatiques, est ainsi formé.

Même si la *Av*GFP a été cristallisée pour la première fois en 1974 [30], la structure de la protéine n'a été résolue qu'en 1996, par deux groupes indépendants [43, 44]. La protéine contient 238 acides aminés. La structure cristalline a révélé que le chromophore est situé au centre de la protéine qui prend la forme d'un tonneau composé de 11 feuillets β (Fig. II.2). Le chromophore fait partie d'une hélice α centrale qui est partiellement déformée à cause de la formation du chromophore. La protéine protège le chromophore du solvant, mais aussi d'agents dénaturant et de changements de pH. Le chromophore est entouré par un grand nombre d'acides aminés chargés ou polaires et de molécules d'eau. Les résidus qui entourent le chromophore sont susceptibles de jouer un rôle dans la maturation du chromophore [45, 46] ou d'influencer l'absorption [47],

la fluorescence. Les deux résidus C- et N-terminaux sont situés à l'extérieur du tonneau. Ceci facilite le raccordement de la GFP à une autre protéine lors d'un marquage fluorescent.

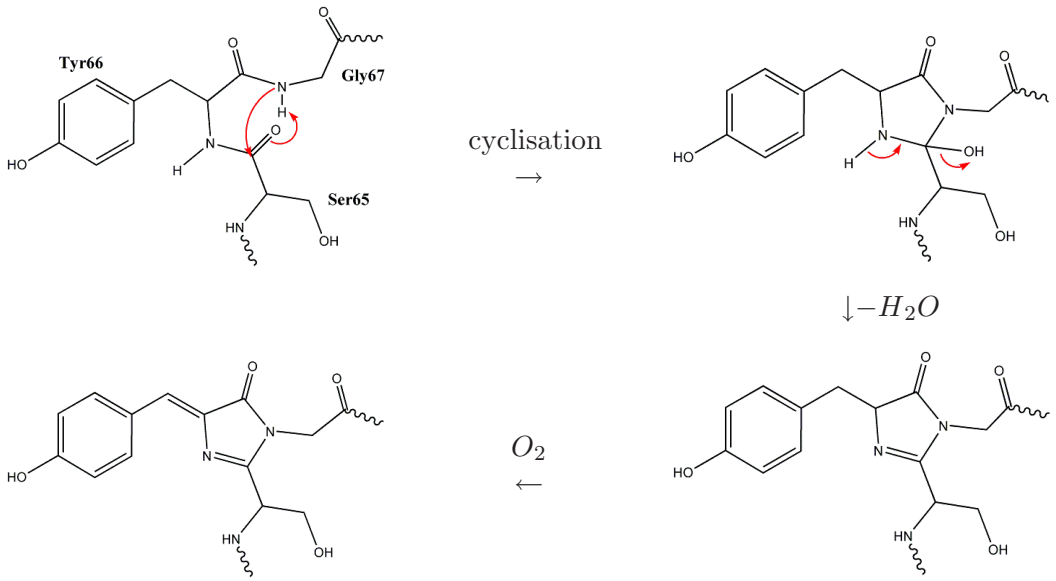


FIG. II.1.: Schéma du mécanisme de formation du chromophore de la *Av*GFP, proposé par Cubitt et al. [40].

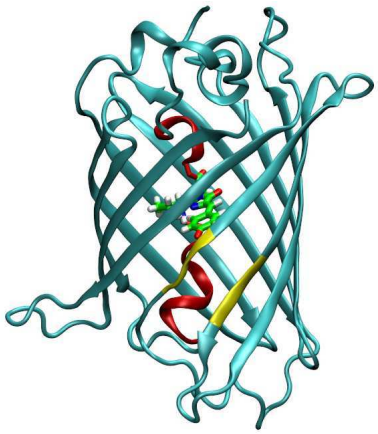


FIG. II.2.: Schéma de la structure cristalline de la GFP qui a la forme d'un tonneau composé de 11 feuillets β (en cyan). Le chromophore (en vert) fait partie d'une hélice α centrale (en rouge), placée à l'intérieur de la protéine. La structure de feuillets β comporte une irrégularité en face du chromophore ; une crevasse entre deux brins β (en jaune) où le réseau de liaisons hydrogène est rompu.

Dans une population de *Av*GFP, deux états de protonation du chromophore coexistent : une majorité sous la forme protonée (neutre) et une minorité sous la forme déprotonée (anionique) [48, 49]. La déprotonation se produit sur le groupement OH du phénol. Le spectre d'absorption de la *Av*GFP présente donc deux bandes [30], centrées autour de 395 nm et 475 nm, qui sont attribuées à la forme neutre et la forme anionique du chromophore, respectivement [36, 40, 48]. Quand l'anion est excité, la *Av*GFP émet une fluorescence verte

autour de 503 nm [36]. L'excitation du chromophore neutre conduit également à une fluorescence verte autour de 508 nm [36], ce qui est interprété comme une déprotonation du chromophore à l'état excité [49]. Le décalage de 5 nm entre les deux maxima de fluorescence peut venir de la différence d'environnement selon qu'il est adapté au chromophore neutre ou à sa forme anionique [49]. La déprotonation est favorisée par le fait que le phénol devient plus acide à l'état excité [48]. Le passage entre la forme neutre et la forme anionique du chromophore semble être réversible et réglé par un réseau de liaison hydrogène. Une chaîne de transfert de proton, à partir du groupement phénol du chromophore, via une molécule d'eau, le résidu Ser205 jusqu'à l'acide glutamique Glu222 à été proposée [48, 50].

La partie, très minoritaire, des chromophores qui restent sous la forme protonée à l'état excité produit une fluorescence bleu autour de 450 nm [51].

La fluorescence verte de la *Av*GFP présente un rendement quantique, Φ_{fluo}, de 0,8 [30, 23] et une durée de vie de fluorescence, τ_{fluo}, de 3,3 ns [52, 53]. Le déclin de fluorescence peut être décrit par une fonction mono-exponentielle avec une bonne précision.

II.1.2. La famille des protéines fluorescentes

L'utilisation très répandue des protéines fluorescentes (FPs) comme marqueurs et senseurs en imagerie cellulaire [1, 8, 9, 10, 11], a fortement promu la recherche de nouvelles FP. Il existe aujourd'hui un grand nombre de mutants et d'homologues de la *Av*GFP dont la structure globale (Fig. II.4) reste la même mais dont les caractéristiques photophysiques sont très diverses.

Une première génération de nouvelles protéines fluorescentes (FPs) a été conçue par mutagenèse dirigée de la *Av*GFP. Le spectre de couleurs a ainsi été étendu du bleu jusqu'au jaune [1]. Des mutations du chromophore, formé par la séquence tripeptidique Ser65-Tyr66-Gly67, ont eu lieu en position 65 et 66. Le remplacement de la sérine en position 65, par une cystéine, leucine, alanine ou thréonine favorise la forme anionique du chromophore et réduit ainsi le spectre d'absorption à une seule bande, e.g. S65T-GFP [43]. La mutation du résidu en position 66 (la tyrosine dans la *Av*GFP) modifie le système conjugué du chromophore. L'introduction d'une histidine ou d'une phenylalanine a donné lieu à des FPs bleu (BFPs) , e.g. EBFP [54] et Sirius [55], et l'introduction d'un tryptophane à des FPs cyan (CFPs), e.g. ECFP [36], Cerulean [56] et mTurquoise [57]. Le spectre d'absorption du chromophore peut également être modifié par la mutation de résidus voisins. L'introduction d'une tyrosine en position 203 crée un π-stacking entre le cycle phénol de Tyr203 et le cycle phénol/phénolate du chromophore (chaîne latérale de Tyr66). Cette interaction décale l'émission de fluorescence vers le jaune et a ainsi donné lieu à des YFPs [58] (yellow FP). Ces diverses mutations, impliquant des résidus plus ou moins proches du chromophore, ont été utilisées pour améliorer les propriétés de maturation du chromophore, de repliement de la protéine, de sensibilité de pH, d'oligomérisation etc.

Parmi les FPs issues de la mutagenèse de la *Av*GFP, les YFPs étaient pendant longtemps celles qui présentaient la fluorescence la plus décalée vers les grandes longueurs d'onde. Le spectre de couleur des FPs a été étendue jusqu'au rouge par la protéine DsRed [59, 60], découverte dans le corail *Discosoma*. Cet organisme marin appartient, comme *Aequorea victoria* (*Av*), à l'embranchement *Cnidaria* mais fait partie de la classe *Anthozoa*, contrairement à *Av* qui fait

partie de la classe *Hydrozoa*. La fluorescence rouge de la DsRed vient du fait que le système conjugué du chromophore a été étendu par une oxydation supplémentaire. La mutagenèse dirigée de la DsRed, visant notamment les résidus 65 et 66 du chromophore, a donné lieu à la série *mFruit [3]* où les FPs *mOrange*, *mStrawberry* et *mCherry,* qui fluorescent de l'orange jusqu'au rouge, sont parmi les plus populaires. Aujourd'hui, une multitude de mutants, dont la fluorescence s'étend du bleu au rouge, peuvent s'ajouter à la liste de FPs anthozoaires [61, 9]. La découverte des FPs dans des crustacés [62] et d'autres animaux de mer [5] a également contribué à la diversité de la famille de FPs.

Il existe également des FPs photo-activables où le chromophore peut être transformé d'une forme 'désactivée' (non-fluorescente) à une forme 'activée' (fluorescente) par irradiation lumineuse. La PA-GFP (PhotoActivatable Green Fluorescent Protein) [63], un mutant de la *Av*GFP, a été une des premières à être conçue. Avant la photo-activation, le chromophore est neutre (protoné) et une irradiation dans la bande d'absorption du chromophore anionique (504 nm) ne produit donc pas de fluorescence. Cette forme de la protéine est appelée non-fluorescente ou désactivée. Une irradiation par lumière violette (∼400 nm) 'active' le chromophore par deux mécanismes : elle excite le chromophore neutre et le déprotone, et elle décarboxyle le résidu Glu222 - un relais essentiel pour la déprotonation réversible du chromophore. Dans cette forme activée où le chromophore est déprotoné et stabilisé dans sa forme anionique, une irradiation à 504 nm produit de la fluorescence.

Les FPs où le chromophore peut être transformé d'une forme fluorescente à une autre par irradiation lumineuse sont appelées photo-convertibles. Dans la protéine *Kaede*, issue du corail anthozoaire *Trachyphyllia*, le chromophore absorbe et fluoresce initialement dans le vert. Une irradiation de lumière violette crée la rupture d'une liaison peptidique au début de la séquence tripeptidique du chromophore. Le système conjugué du chromophore est ainsi étendu sur une cycle aromatique supplémentaire - une histidine. Cette photo-conversion est irréversible et décale l'absorption et la fluorescence du chromophore vers le jaune.

Les FPs qui présentent une photo-activation réversible sont appelées *photo-switchable*. Dans la protéine Dronpa [64], issue du corail anthozoaire *Pectiniidae*, le chromophore est fluorescent ou 'activé' à l'état d'équilibre [65]. Le chromophore peut ensuite être désactivé par une irradiation lumineuse. Cette transformation est réversible et liée à une isomérisation du chromophore : *cis* (activé)↔*trans* (désactivé) [66, 67]. Dans la protéine Padron [68], issue d'une mutagenèse de Dronpa, le chromophore est non-fluorescent à l'état d'équilibre. Cette forme désactivée correspond à l'isomère *trans* du chromophore qui se trouve majoritairement sous la forme anionique. Une irradiation par lumière bleue active le chromophore par l'isomérisation *trans*→*cis*. La forme anionique du chromophore *cis* absorbe également dans le bleu et l'irradiation produit donc à la fois l'isomérisation du chromophore (en excitant la forme *trans*) et la fluorescence du chromophore *cis*. L'isomère *cis* présente un mélange de la forme anionique et la forme neutre du chromophore. L'irradiation dans la bande d'absorption du chromophore neutre, par lumière UV, désactive le chromophore par l'isomérisation *cis*→*trans*.

La grande famille des FPs comprend également des chromoprotéines ou FPs noires (*dark*) qui absorbent dans le visible mais qui ne fluorescent pas. Certaines chromoprotéines issues de coraux et d'organismes marins peuvent être rendues fluorescentes par un petit nombre de

mutations, e.g. aceGFP [69]. D'autres FPs noires, e.g. REACh (*Resonance Energy-Accepting Chromoproteins*) [70, 71], ont été dérivées de formes sauvages pour être utilisées comme accepteurs dans des techniques de transfert d'énergie par résonance, e.g. la microscopie FRET.

Les protéines qui sont fortement fluorescentes de la famille de FPs ont une durée de vie de fluorescence de quelques nanosecondes et un rendement quantique pouvant être proches de 1,0. Dans cette étude, sans d'autre précision, une FP renvoie à une FP fortement fluorescente.

II.2. Photophysique

II.2.1. Phénomènes photophysiques

Il est commode de visualiser l'ensemble des processus liés à l'absorption et à la fluorescence à l'aide d'un schéma analogue à celui de Perrin-Jablonski (Fig. II.3). Le schéma présenté ici montre d'abord des éléments de la structure du système : les courbes d'énergie potentielle de l'état électronique fondamental S_0 (en bleu) et de l'état électronique excité S_1 (en vert) et, d'autre part, des niveaux vibrationnels (traits horizontaux). Il montre aussi l'ensemble des processus mis en jeu (en rouge).

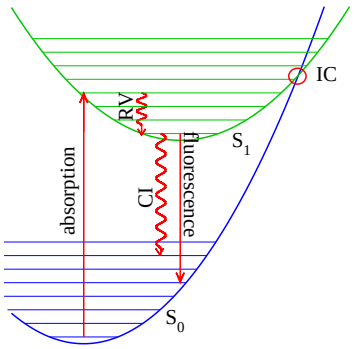


Fig. II.3.: Schéma des mécanismes d'absorption, de fluorescence et de relaxation non-radiative (d'après le diagramme de Jablonski [72]). Les lignes horizontales représentent des niveaux vibrationnels supposés.
S_0 = état fondamental
S_1 = état excité singulet
RV = relaxation vibrationnelle
CI = conversion interne
IC = intersection conique

L'absorption d'un photon excite le chromophore par une transition électronique verticale, selon le principe de Franck–Condon (FC). Le point d'arrivée à l'état S_1 (FC) correspond donc à la géométrie d'équilibre de l'état S_0 (une géométrie plane) et la configuration électronique de l'état S_1. La géométrie du chromophore de ce niveau vibrationnel excité de l'état S_1 est ramenée au niveau vibrationnel fondamental de ce même état par relaxation vibrationnelle. La fluorescence

est ensuite émise de cette géométrie, appelée FS (*fluorescent state*), le minimum de la surface d'énergie potentielle (PES) de l'état S_1 (règle de Kasha).

La désactivation énergétique d'une molécule est toujours un équilibre entre des processus radiatifs et non-radiatifs. L'émission de fluorescence, faisant intervenir un photon, est radiative alors que la conversion interne (CI) est un processus non-radiatif qui ramène le système de l'état électronique S_1 à l'état S_0 par un couplage des niveaux vibrationnels de l'état excité et de l'état fondamental. Quand les surfaces des deux états se rapprochent, ce couplage non-adiabatique devient plus grand et la conversion plus efficace. Dans le cas où les deux surfaces sont dégénerées on parle d'intersection conique (IC). Il serait également utile, de manière générale, de considérer les états triplets de spin et les notions associées de 'croisement intersystème' (conversion interne entre états de spin différents) et de phosphorescence (relaxation radiative triplet → singulet). Cependant, la phosphorescence se produit sur une échelle de temps plus longue que la fluorescence.

II.2.2. Caractérisation de la fluorescence

La fluorescence peut être décrite par plusieurs caractéristiques. Le spectre d'émission décrit l'intensité de fluorescence en fonction de la longueur d'onde. La longueur d'onde λ du photon émis est inversement proportionnelle à l'écart énergétique entre l'état initial et l'état final de la transition. La relaxation vibrationnelle sur la PES de l'état excité avant l'émission donne donc lieu à cette relation :

$$\lambda_{absorption} < \lambda_{fluorescence}$$

Une autre notion importante est celle de rendement quantique : rapport du nombre de photons émis au nombre de photons absorbés. On peut le relier aux constantes de vitesse des processus radiatifs et non-radiatifs à l'aide de la relation

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\left(k_r + k_{nr}\right) N(t)$$

où $N(t)$ est la population de chromophore encore fluorescent au temps t, k_r est la constante de vitesse radiative (fluorescence) et k_{nr} est la somme des constantes de vitesse des différents processus non-radiatifs (conversion interne). Il en résulte

$$\begin{aligned} N(t) &= N(0)\, exp\left(-\frac{t}{\tau_{fluo}}\right) \\ \tau_{fluo} &= \frac{1}{k_r + k_{nr}} \end{aligned} \tag{II.1}$$

où τ_{fluo} est la durée de vie de fluorescence.

On en déduit le rendement quantique

$$\begin{aligned} \phi_{fluo} &= \frac{1}{N(0)} \int_{0,\infty} k_r \, N(t) \, \mathrm{d}t \\ &= \frac{k_r}{k_r + k_{nr}} \end{aligned}$$

La probabilité d'émission de fluorescence ($S_1 \rightarrow S_0$) peut être caractérisée par la force d'oscillateur de la transition f_{01} (paramètre sans dimension), définie comme

$$f_{01} = \frac{p_{01}}{q_{01}}$$

où p_{01} est la probabilité de la transition $S_1 \rightarrow S_0$ dans la molécule et q_{01} la probabilité de la transition $S_1 \rightarrow S_0$ dans un oscillateur harmonique où l'écart en énergie $S_1 - S_0$ est le même que dans la molécule. Plus la force d'oscillateur est grande, plus la transition est probable. Une force d'oscillateur très faible (<0,1) indique une transition quasi-interdite. La force d'oscillateur est liée au moment dipolaire de transition μ_{01} qui représente le changement du moment dipolaire de la molécule pendant la transition (dipôle de transition).

La durée de vie de fluorescence et le rendement quantique peuvent être tirés des expériences. La force d'oscillateur de la transition f_{01} correspondant à la fluorescence verte du chromophore peut également en être déduite. Cependant, pour savoir comment f_{01} varie avec la conformation du chromophore, des calculs théoriques sont nécessaires.

II.2.3. Photophysique des protéines fluorescentes

Dans le cas des protéines fluorescentes, le schéma (fig. II.3) issu du diagramme de Perrin-Jablonski omet un aspect important du processus. Le minimum de la géométrie plane de l'état S_1 (FS) est lié à d'autres minima sur la PES. Ils sont atteints par une torsion de 90° (**Tw**ist) autour d'une des deux liaisons centrales (angle dièdre φ ou τ) du chromophore (Fig. II.4). Les minima sont nommés TwP ($\tau = 0°$,$\varphi = \pm 90°$) et TwI ($\tau = \pm 90°$,$\varphi = 0°$) et correspondent à des géométries où le cycle **P**hénolate ou **I**midazolinone est perpendiculaire au reste du chromophore
.

Un schéma plus réaliste du diagramme de Perrin-Jablonski, issu de données numériques de la littérature [73], est présenté sur la figure II.5.

La principale différence vient de la spécificité des surfaces d'énergie potentielle des états S_0 et S_1 le long de la coordonnée de déformation géométrique considérée sur la figure (τ ou φ). L'absorption a lieu dans la géométrie d'équilibre de l'état S_0 - une géométrie plane. La fluorescence a lieu dans la géométrie plane (FS) car la force d'oscillateur est grande. La conversion interne, par contre, n'est pas susceptible de s'y produire car la différence d'énergie est trop grande (*ca.* 3 eV). En revanche, dans les géométries perpendiculaires (TwP et TwI), la conversion interne est favorisée par l'écart énergétique $S_1 - S_0$ plus petit (*ca.* 0,5 eV) et la fluorescence est défavorisée par une diminution de la force d'oscillateur.

La durée de vie de la géométrie plane à l'état S_1 (FS) est donc limitée par deux processus : l'émission de fluorescence (désactivation radiative) et la torsion du chromophore S_1 (l'extinction de fluorescence). La désactivation radiative du FS est caractérisée par la constante de vitesse k_r ou son inverse, la durée de vie radiative τ_r :

$$\tau_r = \frac{\tau_{fluo}}{\phi_{fluo}} \tag{II.2}$$

De la même façon, la désactivation non-radiative est caractérisée par la constante de vitesse k_{nr} ou son inverse, la durée de vie non-radiative τ_{nr}, tirées de l'équation II.1.

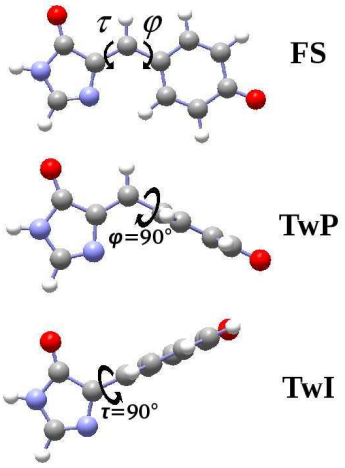


FIG. II.4.: Schéma du modèle du chromophore anionique de la GFP - p-hydroxybenzylidene-imidazolinone (HBI) - dans sa forme isomère cis, dans trois différentes géométries ; la plane (FS) ($\tau = \varphi = 0^\circ$), le TwP ($\tau = 0^\circ, \varphi = \pm 90^\circ$) et le TwI ($\tau = \pm 90^\circ, \varphi = 0^\circ$).

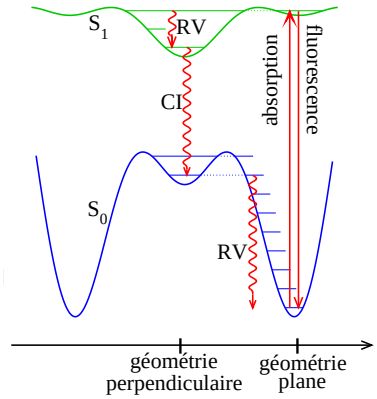


FIG. II.5.: Schéma des mécanismes d'absorption, de fluorescence et de relaxation non-radiative dans le chromophore de la GFP (d'après des calculs *ab initio* de la littérature [73]). Les lignes horizontales représentent des niveaux vibrationnels supposés ; les lignes horizontales pointillées montrent la zone classiquement interdite de ces niveaux.
S_0 = état fondamental
S_1 = état excité singulet
RV = relaxation vibrationnelle
CI = conversion interne

II.2.4. L'influence de l'environnement

Les propriétés de fluorescence du chromophore peuvent drastiquement changer avec l'environnement. Au sein d'une protéine, le chromophore peut être fortement fluorescent, dans le cas de la GFP : $\phi_{fluo} = 0,64$ [23] et $\tau_{fluo} = 3\,ns$ [75]. Or, lorsque la protéine est dénaturée [76] (dépliée) ou quand le chromophore de la GFP se trouve dans l'eau, il est non-fluorescent. Par exemple, dans l'eau : $\phi_{fluo} < 10^{-3}$ [24] et $\tau_{fluo} = 0,17\,ps$ (il y a en fait deux composantes : $\tau_1 = 0,2\,ps$ et $\tau_2 = 1,1\,ps$ [74]). De plus, la fluorescence du chromophore en solution peut être récupérée par un refroidissement du système (77 K) [24]. La variation de fluorescence avec l'environnement vient d'une désactivation non-radiative plus ou moins efficace. En effet, les interactions du chromophore avec l'environnement et la dynamique de l'environnement contrôlent la flexibilité du chromophore. Il atteint donc plus ou moins rapidement les géométries perpendiculaires où la conversion interne est la plus probable. La diversité des rendements quantiques des FPs contenant des chromophores identiques [51] montre le rôle important de l'architecture de la protéine dans le contrôle du déclin non-radiatif du chromophore.

II.2.5. Mesures expérimentales

Des mesures expérimentales de la durée de vie de fluorescence τ_{fluo} et le rendement quantique de fluorescence ϕ_{fluo} permettent de déduire une valeur 'expérimentale' de τ_r (eq. II.2) et τ_{nr} (eq. II.1). La durée de vie non-radiative τ_{nr} dépend fortement de l'environnement et peut difficilement être extrapolée d'un mutant à un autre. La durée de vie radiative τ_r est, au contraire, une propriété caractéristique d'un chromophore donné, qui ne dépend que faiblement de l'environnement [77]. Elle devrait donc être similaire dans des mutants de la GFP qui ne diffèrent que par quelques mutations et qui ne modifient pas le système conjugué du chromophore. On peut noter que les mutants S65T-GFP et YFP, qui diffèrent par 4 mutations, ont des durées de vie radiative proches : 4,7 ns [23, 75] et 6,0 ns [78, 1], respectivement.

II.3. Protéines étudiées

Dans cette étude nous avons étudié trois membres de la famille des protéines fluorescentes (FPs) : la S65T-GFP - un mutant vert de la *Av*GFP, la YFP - un mutant jaune de la *Av*GFP et Padron - une FP photoactivable.

II.3.1. La GFP

La protéine S65T-GFP [79] diffère de la *GFP* sauvage (*Av*GFP) par une seule mutation ; la sérine 65 a été remplacée par une thréonine. Le chromophore est donc formé par le tripeptide Thr65-Tyr66-Gly67 (Fig. II.6), au lieu de Ser65-Tyr66-Gly67.

Le spectre d'absorption présente une bande à 394 nm, correspondant à la forme neutre, et une bande nettement plus importante à 488 nm [79], correspondant à la forme anionique du chromophore. Dans cette FP, la forme anionique du chromophore est donc prédominante, contrairement à la *Av*GFP dans laquelle l'anion coexiste avec la forme neutre [48]. L'irradiation dans la bande du chromophore anionique produit une fluorescence à 510 nm [79]. La fluorescence verte de la S65T-GFP présente des caractéristiques proches de celles de la *Av*GFP : une durée de vie de fluorescence de 3 ns [75], et un rendement quantique de 0,64 [23] à pH 8,0.

La structure de la protéine a été déterminée par diffraction de rayons X (RX) et elle est déposée sous le nom de 1EMA [43] dans la *protein data bank* [80]. La protéine a été cristallisée à pH 8,0-8,4 alors que le pKa du chromophore au sein de la protéine est proche de 6 [81]. La structure cristalline 1EMA représente donc uniquement la forme anionique et fluorescente du chromophore.

Dans l'étude de la S65T-GFP, nous nous intéressons à la dynamique de la forme fluorescente du chromophore au sein de la protéine. La structure 1EMA constitue un bon point de départ pour la dynamique moléculaire puisque l'environnement protéique est adapté pour la forme anionque du chromophore.

Dans cette étude, le nom GFP renvoie à la S65T-GFP.

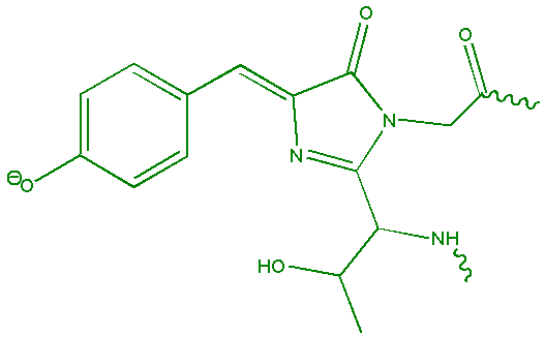


FIG. II.6.: Le chromophore de la protéine S65T-GFP.

II.3.2. La YFP

La protéine fluorescente jaune (YFP) [58] a été obtenue par mutagenèse et diffère de la *Av*GFP par quatre mutations : T203Y, S65G, V68L et S72A. Le chromophore est donc formé par le tripéptide Gly65-Tyr66-Gly67 (Fig. II.7). La mutation T203Y introduit une tyrosine en π-stacking avec le phénolate du chromophore ce qui a pour effet de déplacer la fluorescence vers le jaune. La structure de la protéine a été déterminée par diffraction de rayons X (RX) et elle est déposée sous le nom de 1YFP [58] dans la *protein data bank* [80].

La forme neutre du chromophore absorbe à 392 nm dans la YFP et la forme anionique à 514 nm [58]. La forme anionique est responsable de la fluorescence qui est maximale à 528 nm [58]. Le pKa du chromophore au sein de la protéine a été déterminé à 7,0 [58]. La structure 1YFP est cristallisée à pH 7,5 et représente donc un mélange entre la forme neutre (protonée) et anionique (déprotonée) du chromophore. Cependant, l'arrangement des résidus voisins du chromophore suggère une forte présence de la forme anionique. Ce point est davantage détaillé dans la section IV.2.

Le déclin de fluorescence peut être décrit par une mono-exponentielle avec une durée de vie de 3,68 ns [78]. La YFP a un rendement quantique de 0,61 [1].

II.3.3. Padron

La protéine photoactivable Padron a été obtenue par mutagenèse de Dronpa [64], une FP photoactivable issue du corail anthozoaire *Pectiniidae*. Les deux protéines présentent des photo-activations inverses. Chez Dronpa l'activation est dite négative, i.e. l'irradiation qui induit

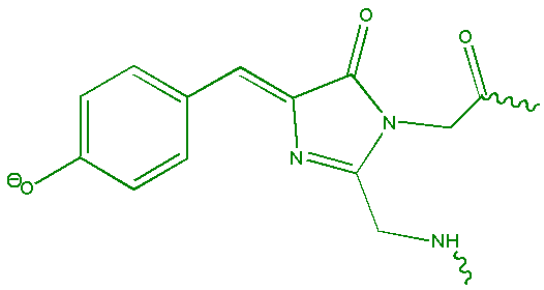


FIG. II.7.: Le chromophore de la protéine YFP en π-stacking avec le résidu 203.

la fluorescence désactive le chromophore, alors que chez Padron l'activation est dite positive, i.e. l'irradiation qui induit la fluorescence active le chromophore. Dans les deux cas, la photo-activation est liée à une isomérisation *trans*→*cis* du chromophore. Le chromophore est formé par le tripéptide Cys-Tyr-Gly. Il est montré sous la forme *cis* et *trans* sur la figure II.8.

Ici la forme protonée (neutre) du chromophore sera appelée A et la forme déprotonée (anionique) du chromophore B.

Padron a été rapportée par Andresen *et al.* [68]. Le chromophore se trouve sous la forme désactivée *trans*, quasiment non-fluorescente (0,7% du signal de fluorescence de la forme activée), à l'état d'équilibre à un pH physiologique. Le spectre d'absorption du chromophore *trans* présente une seule bande à 505 nm qui correspond à la forme déprotonée du chromophore. Une irradiation dans cette bande B_{trans} induit l'isomérisation *trans*→*cis*. Le spectre d'absorption du chromophore *cis* révèle un mélange de la forme protonée A_{cis}, bande à 396 nm, et de la forme déprotonée B_{cis}, bande à 503 nm, du chromophore. Une irradiation dans la bande de B_{cis} produit une fluorescence verte (522 nm) avec un rendement quantique de 0,64. Une irradiation dans la bande de A_{cis} induit l'isomérisation *cis*→*trans*.

Brakemann *et al.* [82] ont déterminé la structure cristalline de la forme désactivée (chromophore *trans*) et de la forme activée (chromophore *cis*) de Padron par diffraction de rayons X. Leur variante de Padron, appelée Padron0.9, diffère de celle de Andresen *et al.* par deux mutations introduites pour favoriser la cristallisation. Les spectres d'absorption et de fluorescence des deux variantes sont très similaires. Le cristal de la forme activée *cis* de la protéine est obtenu par une irradiation du cristal de la forme désactivée *trans* à température ambiante. Les deux formes du cristal sont ensuite gelées à 100 K pour préserver leur structure. L'isomérisation du chromophore déplace très peu le cycle imidazolinone et ne demande aucun réarrangement important des résidus voisins.

La structure cristalline de la variante initiale de Padron [68] a été déterminée par diffraction de rayons X par Faro *et al* [83]. Ils ont caractérisé la forme désactivée et activée de la protéine ainsi que plusieurs intermédiaires de la photo-activation. Leur résultats montrent que l'isomérisation du chromophore, *trans*→*cis*, peut se produire même dans un environnement protéique refroidi à 100 K, et sans changement d'état de protonation du chromophore. Une irradiation à 100 K convertit la forme déprotonée *trans* du chromophore, B_{trans}, en un premier intermédiaire I_{cis} qui correspond à la forme déprotonée *cis* du chromophore. Une relaxation dans le noir du cristal, transforme I_{cis} en un deuxième intermédiaire qui se convertit en AB_{cis}, un mélange de B_{cis} et A_{cis}, dès lors que la température passe au-dessus de la température de transition vitreuse. Le

rendement de l'isomérisation *trans*→*cis* à 100 K est faible ($\sim 5.10^{-6}$) par rapport au rendement à température ambiante ($\sim 2.10^{-4}$), mais montre la possibilité du chromophore de s'isomériser même à basse température.

Les structures cristallines de B_{trans}, I_{cis} et AB_{cis} sont déposées dans la *protein data bank [80]* sous les noms 3ZUF, 3ZUL, 3ZUJ, respectivement.

La structure I_{cis} correspond à l'état de la protéine directement après l'isomérisation. L'état de protonation du chromophore n'a pas changé et l'environnement protéique ne s'est pas encore relaxé autour de la nouvelle conformation du chromophore. La comparaison entre les structures cristallines de B_{trans} et I_{cis} peut donc mettre en évidence des déplacements de résidus voisins liés à l'isomérisation du chromophore. Cette analyse est détaillée dans le chapitre IV.3.

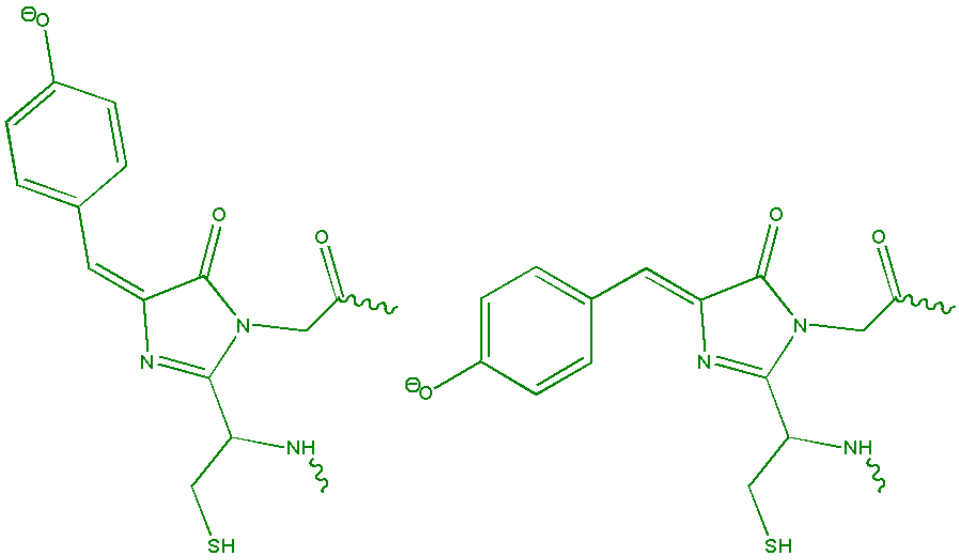


FIG. II.8.: Le chromophore de la protéine Padron, sous la forme isomérique *trans* (à gauche) et *cis* (à droite).

III. Potentiel du chromophore - NADiA

III.1. Contexte

Des simulations de dynamique classique du chromophore de la GFP à l'état excité (S_1) *in vacuo*, basée sur un Hamiltonian semi-empirique [73] ou sur des études de dynamique *ab initio* [84, 85], ont montré que l'extinction de fluorescence est due à une torsion autour de l'une ou de l'autre des deux liaisons centrales du chromophore (angles dièdres τ et φ, voir Fig. II.4) et que ces torsions se produisent en quelques picosecondes après l'excitation. Cela indique que la durée de vie de fluorescence de quelques nanosecondes observée dans des protéines fluorescentes (FPs) résulte d'un empêchement de la dynamique de torsion par l'environnement protéique. Il est évident que cet obstacle n'est pas un verrouillage massif des deux torsions puisque dans la plupart des FPs le volume de la cavité autour du chromophore peut accommoder des géométries tordues [22, 86].

Les contraintes imposées par la protéine ont été étudiées plusieurs fois. Virshup *et al.* [85] ont étudié la dynamique du chromophore neutre à l'état S_1 au sein de la GFP par la mécanique quantique/dynamique moléculaire (QM/MD) semi-empirique (FOMO-CI). Ils montrent qu'aucune torsion du chromophore ne se produit dans la protéine à l'échelle de quelques picosecondes de simulation. Grâce à des simulations MD biaisées (*umbrella sampling*), ils identifient des grandes barrières de torsion de φ (>20 kcal/mol) et τ (7 kcal/mol) au sein de la protéine, par contre, ils ne fournissent aucune explication moléculaire de ces contraintes. Maddalo *et al.* [86] et Megley *et al.* [22] montrent, par une étude statique et une étude dynamique, respectivement, qu'en absence d'un potentiel de torsion interne du chromophore, l'environnement protéique de la GFP et de divers mutants tend à imposer une conformation non-planaire au chromophore. Le plus souvent, cette conformation correspond à une torsion concerté entre τ et φ de 20° à 40° avec une corrélation négative : $\varphi < 0°$ et $\tau > 0°$. Moors et al. [87] ont étudié l'isomérisation cis→trans à l'état fondamental (S_0) dans la protéine photoactivable Dronpa par *parallel tempering* MD et ont pu identifier des résidus bloquant cette transformation très coûteuse en énergie. Vallverdu *et al.* [88] ont déterminé le potentiel de force moyenne de φ, par *umbrella sampling* MD, à l'état S_1 du chromophore au sein de la GFP. Dans le cas ou τ vaut 0° et φ varie de 0°→90°, ils identifient que deux résidus et une molécule d'eau, placés au-dessus et en dessous du chromophore, sont à l'origine du blocage qui limite la torsion φ, dans un sens ou dans l'autre.

Les méthodes QM/MD sont parfaitement adaptées à l'étude de la désactivation non-radiative du chromophore au sein de la protéine. Elles ont effectivement été utilisées dans l'étude du déclin non-radiatif dans les protéines photo-activables asFP595 [89] et Dronpa [90], mais seulement dans les états noirs où ce déclin se fait à l'échelle de la picoseconde. Dans ces deux études, la dynamique du chromophore est basée sur des calculs CASSCF effectués au fur et à mesure que

la simulations avance (*on-the-fly*). Ces méthodes, très coûteuses en temps de calcul, ne peuvent être menées sur des échelles de temps de la nanoseconde. Même à l'échelle de la picoseconde on doit trouver un compromis entre la précision et le temps de simulation : 20 ps (Dronpa) et 1 ps (asFP595). En effet, leurs calculs CAS(6,6) (espace actif défini par 6 électrons et 6 orbitales) ne prennent pas en compte la corrélation dynamique (CASPT2) qui peut modifier la surface d'énergie potentielle et ainsi la dynamique [91, 92].

Si on se tourne vers la dynamique moléculaire (MD) avec des champs de forces classiques, des simulations de plusieurs nanosecondes ne posent pas de problème, mais elle reposent sur la détermination d'une expression analytique de la surface d'énergie potentielle de l'état S_1 du chromophore. La structure électronique du chromophore dans l'état S_1 a été étudiée à l'aide de calculs quantiques de haut niveau mais seulement sur un petit nombre de géométries [26, 93, 94]. Ces résultats révèlent néanmoins le couplage entre les angles dièdres τ et φ : les géométries perpendiculaire ($\tau = 90^\circ$ ou $\varphi = 90^\circ$) sont plus stables que les géométries planes alors que les torsions simultanées de 90° des angles τ et φ, déstabilisent le chromophore S_1 de 5 kcal/mol par rapport à la géométrie plane. Un potentiel couplant les angles τ et φ a déjà été proposé [95], mais il est basé sur un petit nombre de valeurs d'énergie *ab initio* et ne reproduit pas avec une bonne précision la stabilisation des géométries perpendiculaires à l'état S_1. Les champs de forces des programmes MD standard, basé sur des potentiels additifs ne peuvent rendre compte de cette caractéristique de la PES.

Pour permettre des simulations MD à l'échelle de temps de plusieurs nanosecondes, nous avons élaboré un champ de forces de l'état S_1 du chromophore. Pour cela, nous avons d'une part construit un potentiel analytique à deux dimensions (2D), τ et φ, de type non-additif et d'autre part déterminé des charges ponctuelles de l'état S_1 ainsi que des paramètres de déformation du cycle imidazolinone.

Nous avons développé le potentiel 2D NADiA (Non-Additive Dihedral Angle potentiel) en trois étapes : 1) des calculs *ab initio* précis à l'état S_1 répartis sur une grille serrée de géométrie ; 2) l'interpolation de la PES par une expression algébrique dépendant de τ et φ ; 3) l'implémentation d'un nouvel algorithme, permettant l'utilisation d'un potentiel 2D, dans le module parallélisé PMEMD de la suite Amber [96].

III.2. Calculs *ab initio*

III.2.1. La méthode

La PES du premier état singulet excité S_1 du chromophore anionique a été obtenue par des calculs quantiques. Nous avons d'abord optimisé les géométries par la méthode CASSCF (complete active space multiconfiguration self-consistent field) [97] avec la base 6-31G* et ensuite calculé l'énergie à un plus haut niveau de calcul : CASPT2 (complete active space second-order perturbation theory). Dans CASSCF, la matrice d'interaction de configuration (CI) comprend toutes les configurations possibles qui peuvent être créées par l'espace actif CAS(m,n), défini par m électrons et n orbitales, et qui donnent un état singulet de spin.

Une simple minimisation de l'énergie S_1 peut créer des problèmes de convergence dûs à une inversion (flip-flop) entre l'énergie de l'état fondamental S_0 et l'énergie S_1. Ce phénomène vient

du fait que l'énergie S_0 augmente quand on optimise les orbitales pour l'état S_1. Afin d'éviter ce problème nous avons utilisé un CASSCF moyenné (state averaged), SA2-CASSCF, où la moyenne des énergies S_0 et S_1 est minimisée.

L'espace actif complet du chromophore contient 16 électrons et 14 orbitales π. L'emploi de cet espace est très lourd en temps de calcul et non concevable pour l'étude des nombreuses géométries impliquées dans le développement de NADiA. Nous avons donc plutôt cherché à utilisé le meilleur espace "minimal".

Plusieurs études théoriques [26, 98], où l'énergie d'absorption est calculée et comparée avec l'expérience, montrent que l'état fluorescent du chromophore est obtenue par une excitation de nature $\pi \rightarrow \pi*$. Cette excitation correspond principalement à la transition HOMO $\rightarrow$ LUMO, sauf si on utilise une base très diffuse qui fait baisser l'énergie des états Rydberg.

On note d'abord que les torsions autour des liaisons CA2-CB2 et CB2-CG2 (l'angle dièdre τ et φ, respectivement) (Fig. III.1) impliquent chacune la rupture de la conjugaison du système. Une représentation minimale de la torsion implique donc de pouvoir construire des orbitales π liantes et anti-liantes sur chacune des liaisons concernées : comme les deux liaisons sont adjacentes, on peut construire les quatre orbitales nécessaires à partir de trois orbitales de type orbitale atomique.

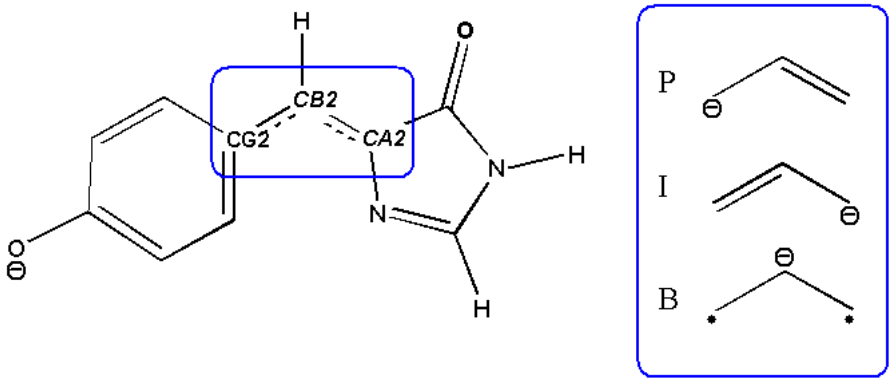


FIG. III.1.: A gauche : le chromophore anionique avec les noms des trois atomes centraux affichés. A droite : trois structures mésomères d'un ion allylique. L'analogie avec le chromophore peut être faite si les deux carbones terminaux sont remplacés par les deux cycles du chromophore.

Par ailleurs, il a été montré [95, 99] que la fonction d'onde électronique du chromophore anionique peut être décrite à l'aide de trois structures mésomères analogue à celles de l'anion allylique (Fig. III.1). Cette description contient deux structures P et I présentant une double liaison soit sur la liaison CA2-CB2 soit sur la liaison CB2-CG2 et une troisième structure B avec deux électrons non-appariés ; un sur l'atome CA2 et un sur l'atome CG2. Ce genre de structures mésomères peuvent être trouvées dans des systèmes avec trois orbitales et quatre électrons [100] ou dans des composés plus grands comme des colorant de Brooker [99]. Sauf si une des liaisons est allongée [100] la troisième structure B a un poids plus petit dans les premiers états électroniques S_0 et S_1 que les structures P and I [95, Tab. V, VI, VII]. Cependant, la structure B est nécessaire pour coupler P et I.

Finalement, un choix raisonnable sera donc l'espace actif CAS(4,3) qui représente les orbitales HOMO-1, HOMO et LUMO et les quatre électrons π correspondant. Nous avons ainsi optimisé la géométrie de l'état S_1 au niveau SA-CAS(4,3) pour un ensemble de conformations du chromophore définies par les valeurs fixées de τ et φ (Fig. III.2).

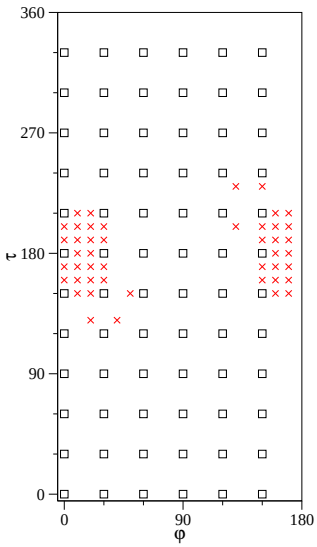


FIG. III.2.: Schéma de la grille (τ, φ) recouvrant le domaine de torsion du chromophore le long la période complète des angles dièdres τ et φ. Les torsions τ et φ sont définies ici comme l'angle dièdre C2-CA2-CB2-CG2 et CA2-CB2-CG2-CD1 (Fig. III.3), respectivement, tel que $\tau = 180^\circ$ et $\varphi = 0^\circ$ dans la géométrie plane de l'isomère cis (FS). Les points de calcul appartenant à la grille commune entre les deux isomères (incrément de 30°) sont représentés par des carrés (72 points) et ceux appartenant à la grille spécifique à l'isomère cis sont représentés par des croix (46 points).

Des calculs CASSCF moyennés sur trois états (SA3), S_0, S_1et S_2, ont été également rapporté dans la littérature. Les résultats diffèrent de ceux des SA2, même au niveau CASPT2. Cette différence peut venir du fait que les orbitales qui minimisent l'énergie de l'état S_2, impliquant principalement la structure B (diradical) et peut-être des orbitales Rydberg, ne sont pas forcément les même que les orbitales donnant un couplage correcte entre les structure P et I.

La méthode CASPT2 utilise les fonctions d'ondes CASSCF comme états de référence pour une perturbation de second ordre de l'espace actif. Ce traitement perturbatif donne une estimation de la corrélation dynamique, non inclue dans CASSCF. De plus, le traitement PT2 compense en grande partie la taille réduite de l'espace actif et le fait que l'on utilise un CAS moyenné.

Finalement, nos résultats au niveau de CASPT2 sont en très bon accord avec les calculs d'un plus haut niveau de calcul : CASSCF(12,11)-PT2 (Tab. III.1). Nos calculs reproduisent bien les profondeurs relatives des deux minima TwP et TwI ainsi que la hauteur du maxima HT. L'effet du traitement PT2 est le plus grand dans les calculs où l'espace actif est petit. On note également que les calculs CASSCF moyennés sur trois états (SA3) diffèrent fortement des calculs SA2. La méthode SA2-CAS(4,3)-PT2 est donc la meilleure approche "minimale".

Les calculs CASSCF et CASSCF//CASPT2 ont été effectués en utilisant les programmes Gaussian [101] et MOLPRO [102], respectivement.

TAB. III.1.: Energies relatives (en kcal/mol) de géométries critiques de la PES de l'état S_1 du chromophore anionique (HBI ou le dérivé diméthylique HBDI selon la référence). L'énergie de référence est celle de la géométrie plane (FS) de l'isomère cis. Les définitions de TwP et TwI sont données sur la figure II.4
a : cette étude **b** : Martin *et al.* [26]. **c** : Polyakov *et al.* [94]. **d** : Olsen *et al.* [93]. **e** : Olsen *et al.* [84].

	TwP	TwI	HT
SA2-CAS(4,3)[a]	-14,7	-18,9	-9,2
SA2-CAS(4,3)-PT2[a]	-6,0	-9,7	4,8
SA2-CAS(12,11)[b]	-8,6	-11,3	0,1
SA2-CAS(12,11)-PT2[b]	-7,2	-8,9	4,9
SA2-CAS(12,11)[c]	-7	-13	
SA3-CAS(4,3)[d]	-21,5	-28,2	
SA3-CAS(4,3)-PT2[d]	0	-7,2	
SA2-CAS(2,2)[e]	-30,0	-34,1	
SA2-CAS(2,2)-PT2[e]	-8,0	-11,1	

III.2.2. Les géométries

Les conformations étudiées sont distribuées sur une grille de points qui recouvre complètement la PES. Les torsions τ et φ sont d'abord incrémentées par 30° tout le long de leur période de 360° et 180°, respectivement. Cette grille de points inclut les deux isomères cis ($\tau = 180°, \varphi = 0°$) et trans ($\tau = 0°, \varphi = 0°$). Pour affiner la description de la région autour des deux géométries planes (cis et trans), nous avons utilisé une grille plus fine (incrément de 10°). Ces grilles s'étendent jusqu'à ±30° de chaque géométrie plane. La description des barrières de torsion a également été améliorée par six points supplémentaires pour chaque isomère (Fig. III.2).

A partir de cette PES, connue point par point, nous avons établi deux potentiels séparés, $V^{cis}_{\mathrm{NADiA}}(\tau, \varphi)$ et $V^{trans}_{\mathrm{NADiA}}(\tau, \varphi)$, qui décrivent le paysage énergétique autour de l'isomère cis et trans, respectivement. Chaque potentiel est interpolé sur la grille commune (incrément de 30°) ainsi que sur les points supplémentaires dans la région de la géométrie plane cis ou trans. Les points de la grille commune garantissent la périodicité de la PES et la cohérence entre les deux potentiels alors que les points supplémentaires améliorent la description dans la région de la géométrie plane spécifique au potentiel.

III.3. Interpolation de la surface d'énergie potentielle

III.3.1. Définitions des angles dièdres τ et φ

Chaque torsion peut être décrite par quatre angles dièdres qui, dans une géométrie complètement plane, différeraient uniquement par 0° ou 180°. Dans le développement de NADiA, pour la description de τ et φ, nous avons choisi de n'utiliser que les angles dièdres où l'hydrogène HB2 (Fig. III.3), considéré comme trop flexible, n'intervient pas. Nous obtenons donc deux angles dièdres par torsion avec les définitions suivantes : τ_N=N2-CA2-CB2-CG2, τ_C=C2-CA2-CB2-CG2, φ_1=CA2-CB2-CG2-CD1 et φ_2=CA2-CB2-CG2-CD2. Dans la géométries plane de l'isomère *cis* (Fig. III.3), on a donc : τ_N=0°, τ_C=180°, φ_1=0° et φ_2=180°.

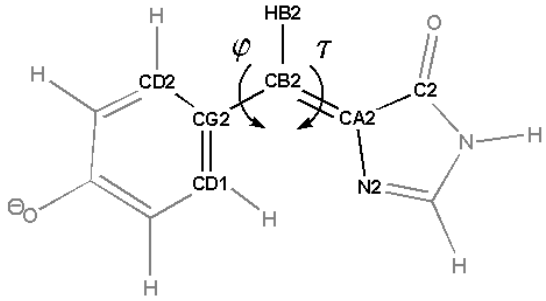


FIG. III.3.: Le HBI (p-hydroxybenzylidène imidazolinone) - le modèle du chromophore de la GFP utilisé pour les calculs SA-CAS(4,3)//PT2.

Afin d'imposer des valeurs spécifiques de τ et φ dans le calcul d'optimisation de géométrie de l'état S_1, nous avons contraint les angles dièdres τ_C et φ_1. Les six angles restant sont laissés libres. Dans certaines géométries optimisées, un ou plusieurs des carbones CA2, CB2 et CG2 (Fig. III.3) sont pyramidalisés. Par conséquent, la différence entre la valeur de l'angle dièdre contraint, e.g. τ_C, et celle de l'angle qui a été laissé libre, e.g. τ_N, n'est pas toujours égale à 180°. On cherche à déterminer un potentiel global des torsions τ et φ et non seulement un potentiel spécifique à τ_C et φ_1. Nous avons donc choisi d'associer l'énergie de chaque calcul aux moyennes $\tilde{\tau}$ et $\tilde{\varphi}$ (eq. III.1), entre τ_N et τ_C et entre φ_1 et φ_2, respectivement. Nous considérons que ces valeurs de τ et φ décrivent au mieux la géométrie obtenue.

$$\begin{aligned} \tilde{\tau} &= \tfrac{1}{2}(\tau_C + \tau_N^{\pi}) \\ \tilde{\varphi} &= \tfrac{1}{2}(\varphi_1 + \varphi_2^{\pi}) \end{aligned} \tag{III.1}$$

où τ_N^{π} et φ_2^{π} sont obtenu à partir de τ_N et φ_2 par un déphasage de 180° :

$$\begin{aligned} \tau_N^{\pi} &= \tau_N \pm 180^{\circ} \\ \varphi_2^{\pi} &= \varphi_2 \pm 180^{\circ} \end{aligned}$$

Un exemple de ce calcul est donné dans le tableau III.2.

angles dièdres de τ	
τ_C	240,0°
τ_N	74,8°
τ_N^π	254,8°
$\tilde{\tau}$	**247,4°**

angles dièdres de φ	
φ_1	30,0°
φ_2	201,8°
φ_2^π	21,8°
$\tilde{\varphi}$	**25,9°**

TAB. III.2.: Exemple du calcul des valeurs moyenne de τ et φ dans l'interpolation de l'énergie potentielle. Les valeurs de $\tau_C = 240^\circ$ et $\varphi_1 = 30^\circ$ sont imposées dans l'optimisation de géométrie.

Les valeurs $\tilde{\tau}$ et $\tilde{\varphi}$ qui seront utilisées dans l'interpolation des calculs CASPT2 diffèrent donc légèrement des valeurs τ_C et φ_1 imposées dans l'optimisation de géométrie, comme indique la figure III.4

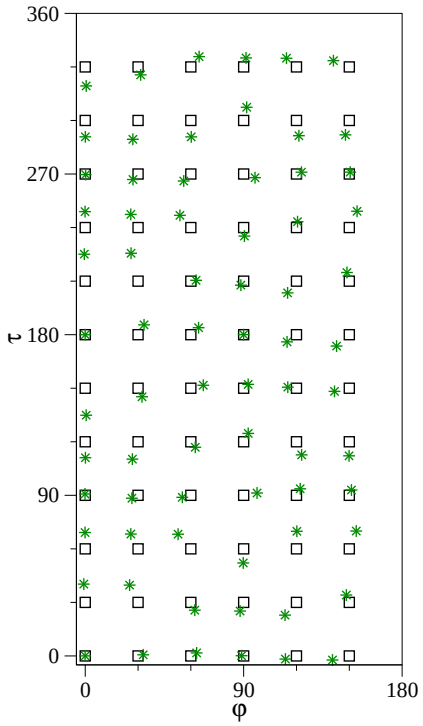


FIG. III.4.: Schéma de la grille commune entre les deux isomères (incrément de 30°). Les valeurs de τ_C et φ_1, contraintes dans l'optimisation de géométrie, sont représentés par des carrés noirs et les valeurs de $\tilde{\tau}$ et $\tilde{\varphi}$, résultant du calcul de l'équation III.1, sont représentées par des étoiles vertes.

III.3.2. Procédure d'interpolation

L'espression de l'énergie potentielle utilisée dans les programmes de MD comporte un terme provenant de chacun des angles dièdres τ et φ mais aussi des termes provenant des interactions non-liées (van der Waals et électrostatiques). Nous avons choisi d'ajuster le potentiel $V_{\mathrm{NADiA}}(\tau,\varphi)$ directement sur les énergies CASPT2 et de supprimer les interactions non-liées, internes au chromophore et qui dépendent de τ et φ de l'expression du potentiel (voir description détaillée dans section III.4.2).

Une procédure itérative est utilisée pour interpoler les énergies CASPT2 (au total 118 points) par deux expressions analytiques (pour les conformations *cis* et *trans*, respectivement). L'expression est constituée d'une somme de fonctions $g(\tau,\varphi)$ qui sont des produits de fonctions sinus et cosinus dépendant de multiples de τ et φ :

$$\begin{aligned} V_{NADiA}(\tau,\varphi) &= \sum k_i g_i(\tau,\varphi) \\ g_i(\tau,\varphi) &= \begin{cases} cos(n\omega_1) \\ sin(n\omega_1) \\ cos(n\omega_1)cos(m\omega_2) \\ sin(n\omega_1)sin(m\omega_2) \\ cos(n\omega_1)sin(m\omega_2) \end{cases} \end{aligned} \tag{III.2}$$

où ω_1 et ω_2 représentent τ ou φ et les entiers n et m varient entre 1 et 10. Afin de préserver la périodicité de 180° en φ, seuls les multiples pairs de φ sont retenus. Les paramètres k_i sont ajustés en minimisant l'écart quadratique moyen χ^2 entre l'interpolation ($V_{\mathrm{NADiA}}(\tau,\varphi)$) et les points de calcul CASPT2 (E_{CASPT2}) (eq. III.3).

$$\chi^2 = \frac{1}{n}\sum_{i=1,n}(V_{NADiA}(\tau_i,\varphi_i) - E_{CASPT2,\,i})^2 \tag{III.3}$$

Les fonctions $g(\tau,\varphi)$ sont choisies de façon itérative. La procédure débute avec une expression initiale contenant deux termes :

$$V^{initial}_{NADiA}(\tau,\varphi) = k_1 cos(2\varphi) + k_2 cos(2\varphi)\cdot cos(\tau)$$

Il est apparu *a posteriori* qu'un terme seul de $cos(\tau)$ n'améliore pas l'interpolation.

L'interpolation est ensuite améliorée par l'ajout de nouvelles fonctions $g(\tau,\varphi)$ à la somme initiale, en ajustant les paramètres k_i à chaque itération. Les détails de cet ajustement sont données dans l'annexe B.

Les nouvelles fonctions sont sélectionnées dans une base de 200 variantes de $g(\tau,\varphi)$ (l'eq. III.2). La sélection d'une nouvelle fonction est contrôlée par deux critères : elle doit pouvoir baisser le χ^2 le plus possible et augmenter l'oscillation de surface le moins possible. Dans le cas où plusieurs $g(\tau,\varphi)$ présentent des recouvrements proches, la fonction qui contient les plus petites multiplicités des angles τ et φ est sélectionnée. Ce choix a pour but de minimiser l'oscillation de surface de l'interpolation.

La procédure est poursuivie jusqu'à ce que χ^2 passe sous un seuil prédéfini. Cette procédure itérative est systématique et peut être appliquée à d'autres problématiques d'un petit nombre de variables. L'interpolation de la PES est contrainte implicitement par la composition de la base de candidats.

Les résultats des expressions analytiques $V^{cis}_{\mathrm{NADiA}}(\tau,\varphi)$ et $V^{trans}_{\mathrm{NADiA}}(\tau,\varphi)$, montrés dans le tableau III.3, reproduisent les valeurs d'énergie CASSCF//CASPT2 avec un écart quadratique moyen relatif (rrsmd) de 7%. Ils contiennent 30 et 27 termes, respectivement.

TAB. III.3.: Les termes des potentiels NADiA en kcal/mol.$V_{NADiA}(\tau,\varphi) = \sum k_i g_i(\tau,\varphi)$

$V(\tau,\varphi)^{cis}_{NADiA}$		$V(\tau,\varphi)^{trans}_{NADiA}$	
k	$g(\tau,\varphi)$	k	$g(\tau,\varphi)$
-0,81	1	-0,74	1
-0,26	$cos(2\varphi)$	-0,30	$cos(2\varphi)$
0,52	$cos(2\tau)$	0,62	$cos(2\tau)$
-0,46	$cos(4\varphi)$	-0,73	$cos(4\varphi)$
-1,06	$cos(4\tau)$	-1,30	$cos(4\tau)$
0,18	$cos(5\tau)$	0,28	$cos(5\tau)$
0,48	$cos(6\tau)$	0,45	$cos(6\tau)$
0,24	$cos(2\varphi) \times cos(\tau)$	0,26	$cos(2\varphi) \times cos(\tau)$
5,47	$cos(2\varphi) \times cos(2\tau)$	5,40	$cos(2\varphi) \times cos(2\tau)$
-2,28	$cos(2\varphi) \times cos(4\tau)$	-1,79	$cos(2\varphi) \times cos(4\tau)$
0,64	$cos(2\varphi) \times cos(8\tau)$	0,28	$cos(2\varphi) \times cos(8\tau)$
-1,25	$cos(4\varphi) \times cos(2\tau)$	-0,95	$cos(4\varphi) \times cos(2\tau)$
-0,27	$cos(6\varphi) \times cos(4\tau)$	-0,20	$cos(6\varphi) \times cos(4\tau)$
-1,09	$sin(2\varphi) \times sin(2\tau)$	-0,99	$sin(2\varphi) \times sin(2\tau)$
-0,33	$sin(2\varphi) \times sin(3\tau)$	-0,23	$sin(2\varphi) \times sin(3\tau)$
0,75	$sin(2\varphi) \times sin(4\tau)$	0,40	$sin(2\varphi) \times sin(4\tau)$
0,43	$sin(2\varphi) \times sin(7\tau)$	0,20	$sin(2\varphi) \times sin(7\tau)$
0,32	$sin(4\varphi) \times sin(3\tau)$	0,19	$sin(4\varphi) \times sin(3\tau)$
1,11	$sin(4\varphi) \times sin(4\tau)$	1,04	$sin(4\varphi) \times sin(4\tau)$
0,15	$cos(6\varphi) \times cos(2\tau)$	-0,50	$cos(4\varphi) \times cos(6\tau)$
-0,11	$cos(6\varphi) \times cos(8\tau)$	0,18	$cos(8\varphi) \times cos(8\tau)$
-0,23	$cos(6\varphi) \times cos(10\tau)$	0,19	$cos(8\varphi) \times cos(10\tau)$
0,29	$cos(10\varphi) \times cos(8\tau)$	-0,64	$sin(2\varphi) \times sin(6\tau)$
0,48	$sin(4\varphi) \times sin(2\tau)$	-0,21	$sin(8\varphi) \times sin(2\tau)$
0,18	$sin(4\varphi) \times sin(7\tau)$	-0,25	$sin(10\varphi) \times sin(10\tau)$
-0,27	$sin(10\varphi) \times sin(5\tau)$	-0,20	$cos(7\tau)$
-0,32	$sin(10\varphi) \times sin(10\tau)$	0,12	$cos(10\varphi)$
0,11	$sin(4\varphi) \times cos(1\tau)$		
0,15	$cos(2\varphi) \times cos(5\tau)$		
-0,32	$cos(8\tau)$		

III.3.3. La surface d'énergie potentielle

Les résultats de l'interpolation de $V^{cis}_{\mathrm{NADiA}}(\tau,\varphi)$ et $V^{trans}_{\mathrm{NADiA}}(\tau,\varphi)$ sont présentés graphiquement dans les figures III.5 et III.6. Dans chaque potentiel la géométrie plane (*cis* ou *trans*) est entouré par quatre minima : TwP$^+$ ($\tau = 0^\circ$,$\varphi = +90^\circ$), TwP$^-$ ($\tau = 0^\circ$,$\varphi = -90^\circ$), TwI$^+$ ($\tau = +90^\circ$,$\varphi = 0^\circ$) et TwI$^-$ ($\tau = -90^\circ$,$\varphi = 0^\circ$). Pour des raisons de symétrie, les géométries TwP$^+$et TwP$^-$sont identiques et TwI$^+$et TwI$^-$sont très similaires.

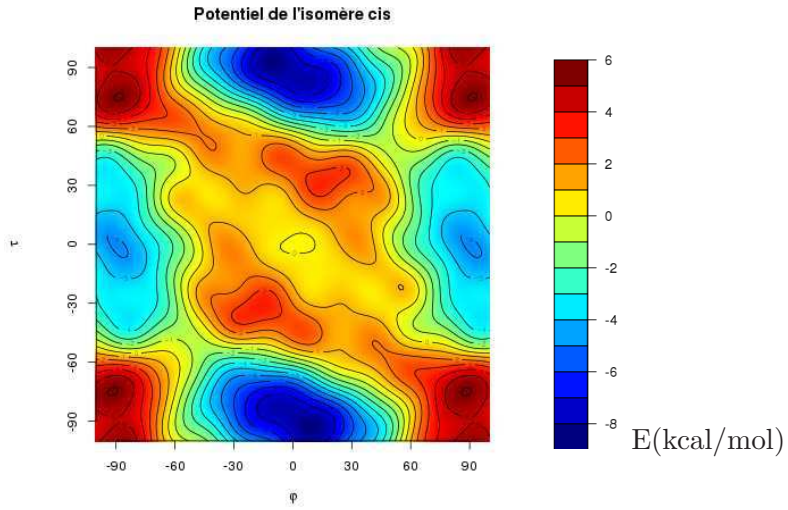


FIG. III.5.: Représentation graphique du potentiel $V^{cis}_{\mathrm{NADiA}}(\tau,\varphi)$. Les valeurs ($\tau = 0^\circ$,$\varphi = 0^\circ$) sont attribuées à la géométrie plane *cis*.

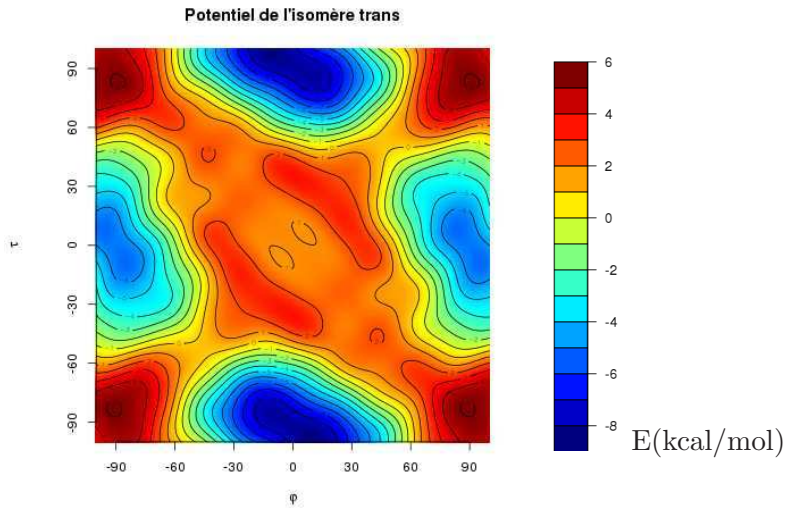


FIG. III.6.: Représentation graphique du potentiel $V^{trans}_{\mathrm{NADiA}}(\tau,\varphi)$. Les valeurs ($\tau = 0^\circ$,$\varphi = 0^\circ$) sont attribuées à la géométrie plane *trans*.

On peut constater que les PES sont lisses malgré le fait que $V^{cis}_{\mathrm{NADiA}}(\tau,\varphi)$ et $V^{trans}_{\mathrm{NADiA}}(\tau,\varphi)$ contiennent des fonction de multiples élevés de τ et φ, e.g. $sin(10\varphi) \times sin(10\tau)$.

La géométrie plane de l'isomère cis (FS) se situe dans le fond d'une vallée étendue et assez plate avec deux sorties étroites. Les minima $TwP^{+/-}$ et $TwI^{+/-}$ sont situés 6 kcal/mol et 10 kcal/mol en dessous de FS, respectivement. Pour atteindre les sorties de la vallée de FS, sans passer par des barrières d'une rotaion 'propre' de τ ou de φ , le chromophore doit effectuer une torsion concertée de τ et φ en sens opposés (Fig. III.7) - un mouvement de type *hula twist* [103] - et franchir une petite barrière (Fig. III.8). La barrière de $TwP^{+/-}$, de 0,6 kcal/mol, est située à $\varphi = \pm 27^\circ$, $\tau = \mp 18^\circ$, et la barrière de $TwI^{+/-}$, de 1,3 kcal/mol, est située à $\varphi = \pm 27^\circ$, $\tau = \mp 38^\circ$. La barrière de $TwI^{+/-}$ est plus étroite que la barrière de $TwP^{+/-}$.

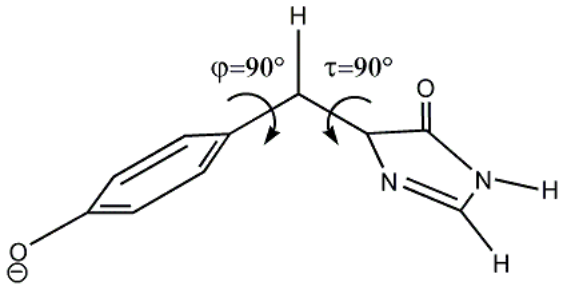


FIG. III.7.: Une torsion concertée de τ et φ, i.e. le *hula twist* [103].

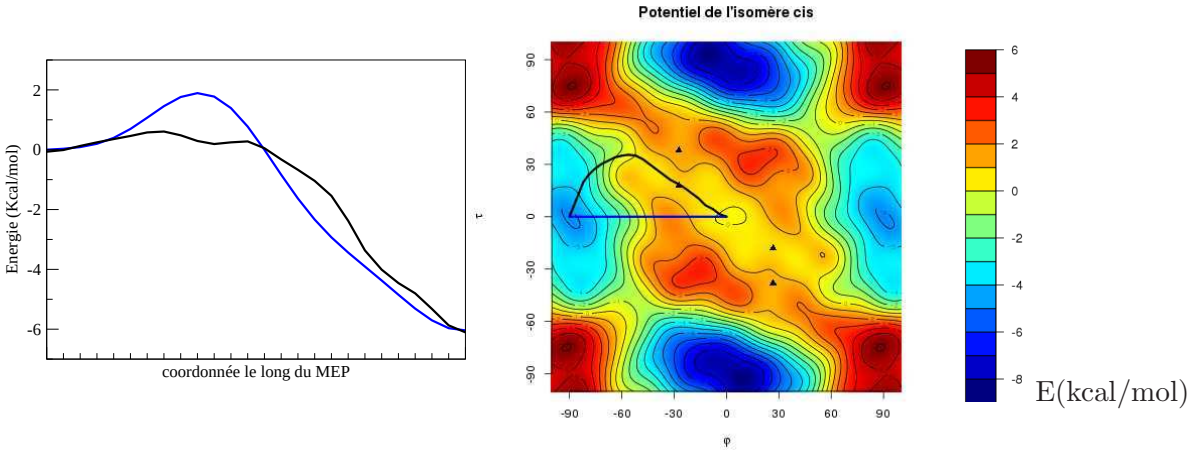


FIG. III.8.: Représentation graphique du potentiel $V^{cis}_{NADiA}(\tau, \varphi)$ où deux chemins de la torsion $\varphi = 0^\circ \rightarrow \varphi = 90^\circ$ du chromophore sont indiqués : le chemin d'énergie minimale (en noir) ou le chemin où τ est bloqué à 0° (en bleu). Les maxima des barrières énergétiques sont indiqués par des triangles sur la PES.

La géométrie plane de l'isomère trans se trouve environ 1 kcal/mol au-dessus de celle de l'isomère cis. Elle se situe également dans une vallée étendue mais avec des sorties plus larges que pour celle de l'isomère cis. Le chromophore doit effectuer un mouvement de type *hula twist* [103] pour quitter la vallée. Les chemin d'énergie minimum (MEP) pour atteindre $TwP^{+/-}$ passe par une petite barrière 0,6 kcal/mol, située à $\varphi = \pm 19^\circ$, $\tau = \mp 29^\circ$. Pour atteindre $TwI^{+/-}$ le chromophore doit passer par la barrière de $TwP^{+/-}$ ainsi qu'une barrière supplémentaire, de 0,2 kcal/mol, située à $\varphi = \pm 27^\circ$, $\tau = \mp 41^\circ$. La barrière de $TwI^{+/-}$ est plus étroite que la barrière de $TwP^{+/-}$.

La PES présente également des maxima situés 5 kcal/mol au-dessus de FS. Ces maxima, nommés HT (hula twist), sont atteints par une torsion concertée de τ et φ de $\pm 90^{\circ}$, dans le même sens ou en sens inverse (Fig. III.7). Les passages entre FS et les géométries HT suivent une crête étroite avec des pentes raides de chaque côté menant directement aux minima TwP et TwI. Ces résultats pourraient expliquer les difficultés rapportées par Martin et al. [26] dans la recherche d'un MEP entre FS et HT. Ils indiquent également une faible probabilité pour que le chromophore visite cette zone au cours d'une dynamique.

III.4. Implémentation de NADiA dans Amber

III.4.1. Algorithme du potentiel NADiA

Dans le champ de forces de Amber, l'énergie potentielle associée à la torsion autour d'une liaison peut être définie par un potentiel générique ou spécifique. Dans le cas d'un potentiel générique, les quatre angles dièdres d'une torsion ω autour d'une double liaison sont traités par le même potentiel. Un potentiel spécifique est, au contraire, adapté et appliqué à un seul angle dièdre. Dans le cas du chromophore, nous avons choisi de traiter les angles dièdres de τ et φ qui impliquent l'hydrogène HB2 (Fig. III.9) par un potentiel égal à zéro et les angles dièdres restant, τ_C, τ_N, φ_1 et φ_2 (définitions données dans section III.3.1 et rappelée dans Fig. III.9 ci-dessous) par le potentiel $V_{\mathbf{NADiA}}(\varphi, \tau)$. La position de HB2 est ainsi uniquement définie par les autres termes du champ de forces, i.e. les angles de valence et l'angle dièdre impropre qui décrit la pyramidalisation de CB2.

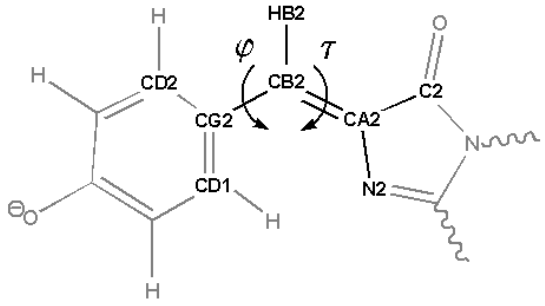


FIG. III.9.: Schéma des atomes impliqués dans les angles dièdres τ et φ. La valeur de τ est définie par la moyenne de τ_C=C2-CA2-CB2-CG2 et τ_N=N2-CA2-CB2-CG2. La valeur de φ est définie par la moyenne de φ_1=CA2-CB2-CG2-CD1 et φ_2=CA2-CB2-CG2-CD2.

Dans l'interpolation de la surface d'énergie potentielle nous devons associer l'énergie CASPT2 à une seule valeur de τ et une seule valeur de φ. Dans un champ de forces tel que Amber, l'expression standard de l'énergie potentielle de torsion autour d'une double liaison est constituée de quatre termes de type $R_1 - X - X - R_2$ (avec R_1 et R_2 des substituants des atomes X). Par analogie, nous avons choisi d'introduire le potentiel couplé en τ et φ de la manière suivante dans Amber :

$$U(\varphi, \tau) = \frac{1}{4}\left[V_{\mathbf{NADiA}}(\varphi_1, \tau_C) + V_{\mathbf{NADiA}}(\varphi_2, \tau_C) + V_{\mathbf{NADiA}}(\varphi_1, \tau_N) + V_{\mathbf{NADiA}}(\varphi_2, \tau_N)\right]$$

Nous avons voulu rendre l'algorithme qui traite le calcul de l'énergie et de la dérivée de V_{NADiA} par rapport à τ ou φ souple pour permettre le traitement d'autres potentiels, conçus pour

différentes formes (neutre/anionique), états (S_0/S_1) ou types (GFP/CFP) du chromophore. L'implémentation est donc adaptée à un potentiel de la forme :

$$V_{\mathbf{NADiA}}(\varphi,\tau) = C + \sum_i k_i g_i \tag{III.4}$$

où C est une constante, g une fonction trigonométrique et k le coefficient correspondant. La souplesse vient de la grande diversité de la fonction g qui est choisie parmi les termes suivants : $sin(n\omega_1)$, $cos(n\omega_1)$, $cos(n\omega_1)cos(m\omega_2)$, $sin(n\omega_1)sin(m\omega_2)$ ou $cos(n\omega_1)sin(m\omega_2)$. Les angles dièdres ω_1 et ω_2 représentent τ ou φ et la multiplicité est déterminée par les entiers n et m.

Un nouveau module a été créé pour le calcul de $V_{\mathbf{NADiA}}(\varphi,\tau)$ et des forces qui en dérivent. Les forces sont opposées au gradient du potentiel (eq. III.5) et calculées en coordonnées cartésiennes $\overrightarrow{r}$.

$$\overrightarrow{f} = -\overrightarrow{\nabla} V(\omega) \tag{III.5}$$

L'angle dièdre ω est calculé à partir des coordonnées cartésiennes des atomes composé par les atomes i, j, k, et l (Fig. III.10) par l'intermédiaire de la formule suivante :

$$cos(\omega) = \overrightarrow{n_{P_1}} \cdot \overrightarrow{n_{P_2}} \tag{III.6}$$

où $\overrightarrow{n_{P_1}}$ et $\overrightarrow{n_{P_2}}$ sont des vecteurs unitaires orthogonaux à deux plans P_1 et P_2 (Fig. III.10).

Un atome donné, intervenant dans l'angle dièdre ω, est soumis suivant la cordonnée cartésienne r, à la force $f(\omega)$, que l'on peut exprimer de la façon suivante :

$$f(\omega) = -\frac{\mathrm{d}V(\omega)}{\mathrm{d}\omega}\frac{\mathrm{d}\omega}{\mathrm{d}cos(\omega)}\frac{\mathrm{d}cos(\omega)}{\mathrm{d}r} \tag{III.7}$$

$$= \frac{1}{sin(\omega)}\frac{\mathrm{d}V(\omega)}{\mathrm{d}\omega}\frac{\mathrm{d}cos(\omega)}{\mathrm{d}r}$$

Le potentiel $V_{\mathbf{NADiA}}(\varphi,\tau)$ dépend des deux angles τ et φ et la force suivant la coordonnée r impliquée dans ces torsions est donc donnée par

$$f(\varphi,\tau) = \frac{1}{sin(\varphi)}\frac{\partial V_{\mathbf{NADiA}}(\varphi,\tau)}{\partial\varphi}\frac{\mathrm{d}cos(\varphi)}{\mathrm{d}r} + \frac{1}{sin(\tau)}\frac{\partial V_{\mathbf{NADiA}}(\varphi,\tau)}{\partial\tau}\frac{\mathrm{d}cos(\tau)}{\mathrm{d}r} \tag{III.8}$$

$$= f_\varphi(\varphi,\tau) + f_\tau(\varphi,\tau)$$

Ici, seul les termes de $\partial V_{\mathbf{NADiA}}(\varphi,\tau)/\partial\varphi$ et $\partial V_{\mathbf{NADiA}}(\varphi,\tau)/\partial\tau$ dépendent de l'expression du potentiel NADiA. Le calcul des autres termes est identique au traitement du potentiel de torsion standard dans Amber.

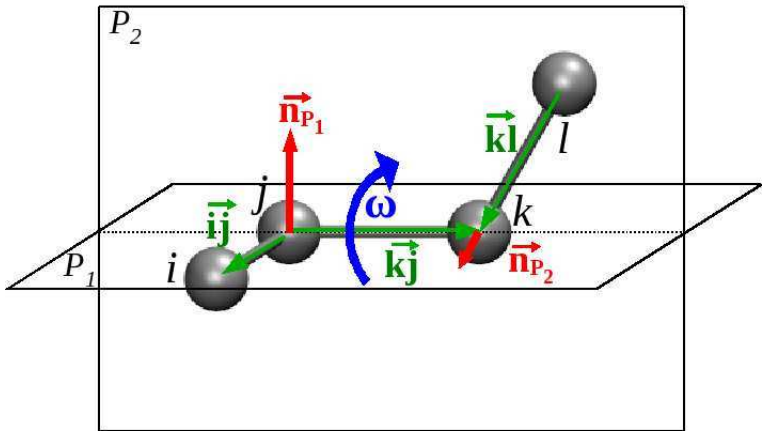


FIG. III.10.: Schéma présentant les différents vecteurs intervenant dans le calcul d'un angle dièdre composé par les atomes i, j, k, et l. Les vecteurs inter-atomes $\overrightarrow{ij}$, $\overrightarrow{kj}$, $\overrightarrow{jl}$ (en vert) sont exprimés en coordonnées cartésiennes et définissent deux plans P_1 ($\overrightarrow{ij}$ et $\overrightarrow{kj}$) et P_2 ($\overrightarrow{kj}$ et $\overrightarrow{jl}$). Les vecteurs normaux (en rouge) à ces plans sont nommés $\overrightarrow{n_{P_1}}$ et $\overrightarrow{n_{P_2}}$. L'angle dièdre est l'angle entre les deux plans et il est donc égal à l'angle entre $\overrightarrow{n_{P_1}}$ et $\overrightarrow{n_{P_2}}$.

En analogie avec les énergies, la contribution de chaque angle dièdre, τ_C, τ_N, φ_1 et φ_2, est prise en compte par une somme dans le calcul des forces.

On peut noter que les atomes figurant dans un seul angle dièdre, C2, N2, CD1 et CD2, subissent chacun une force comportant deux termes alors que les atomes figurant dans tous les angles dièdres, CA2, CB2 et CG2, subissent une force comportant huit termes. Ceci est montré dans l'exemple suivant :

$$f(\varphi,\tau)_{tot}^{x_{\mathrm{N2}}} = \frac{1}{4}\left[f_\tau(\varphi_1,\tau_N) + f_\tau(\varphi_2,\tau_N)\right] \tag{III.9}$$

$$f(\varphi,\tau)_{tot}^{x_{\mathrm{CA2}}} = \frac{1}{4}\Big[f(\varphi_1,\tau_C) + f(\varphi_2,\tau_C) + f(\varphi_1,\tau_N) + f(\varphi_2,\tau_N)\Big] \tag{III.10}$$

où $f(\varphi,\tau)_{tot}^{x_a}$ représente la force exercée sur la coordonnée x de l'atome a. La force associée à un couple (φ,τ) contient deux termes (la dérivée partielle de τ et φ) qui sont définis dans eq. III.8).

III.4.2. Traitement des interactions entre atomes non liés

Les interactions électrostatiques et van der Waals intra-chromophore qui varient avec τ et φ sont implicitement prises en compte dans $V_{\mathbf{NADiA}}(\varphi,\tau)$. Elles doivent donc être enlevées du traitement de Amber où elles sont calculées automatiquement. Pour cela un nouveau test a été introduit dans les modules qui traitent ces interactions non liées à longue et à moyenne (interactions 1-4) portée. Il compare les paires d'atomes traités dans les modules avec une liste prédéfinie correspondant aux interactions à supprimer. Les contributions de paires de la liste ne sont pas calculées et ne contribuent donc pas à la somme globale. Cette liste contient toutes les interactions entre un atome du cycle imidazolinone et un atomes du cycle phénolate, et également toutes les interactions entre un atome du pont CH (CB2 ou HB2, voir Fig.III.9) et un atomes de chacun des deux cycles. On peut noter que la suppression de ces termes ne modifie pas

la sommation de particle mesh Ewald (PME) - la méthode utilisée pour traiter les interactions électrostatiques (voir Annexe C.5).

III.4.3. Lecture et partage de données

Des nouveaux sous-programmes ont été construits pour la lecture et la diffusion des données utilisées par le potentiel NADiA. Des nouveaux mots clé, introduits dans le fichier d'entrée de la dynamique, permettent d'activer ou de désactiver le potentiel NADiA et le traitement des interactions non-liées. Un nouveau fichier d'entrée est également créé pour définir les paramètres du potentiel, les numéros des atomes impliqués dans τ et φ ainsi que les numéros des paires d'atomes impliquées dans les interactions non-liées qui varient avec τ et φ. Afin d'identifier une erreur dans les données (un terme du potentiel qui n'est pas conforme à la définition de l'algorithme NADiA ou le manque d'une donnée dans les fichiers d'entrée), le code imprime un diagnostic dans le fichier de sortie et arrête ensuite le programme principal. Dans le cas de *pmemd*, un programme MD parallélisé de Amber, les données d'entrée sont lues par un seul processus et doivent donc être transmises aux autres processus par une diffusion générale (MPI_BCAST()).

III.5. Simulations dans le vide

Une première étude sur la dynamique du chromophore en absence de la protéine a été effectuée en appliquant le potentiel de torsion 2D dans des simulations de dynamique moléculaire (MD) du chromophore isolé (Fig. III.11) à énergie constante. Le modèle de chromophore utilisé comporte la partie conjuguée saturée par deux groupements méthyle liés aux atomes N3 et C1 (modèle HBDI Fig. III.11). Dans cette étude, nous avons effectué 1000 simulations de 10 ps, toutes partant de différentes géométries planes de l'isomère cis. Les vitesses initiales, engendrées au hasard, donnent des énergies cinétiques initiales très proches (un écart type d'environ 1,5 kcal/mol). Les conformations initiales sont générées par une simulation à l'état fondamental où les angles τ et φ sont décrits par des oscillateurs harmoniques. On voit sur la figure III.12 que la distribution de ces géométries est centrée autour de $\tau = 0^\circ$, $\varphi = 0^\circ$. Après 0,5 ps de dynamique les conformations sont distribuées sur la PES de l'état excité et remplissent la région de basse énergie de FS. Le fait que la distribution change peu entre t=0,5 ps et t=1 ps indique qu'un équilibre est atteint rapidement. On note aussi que seulement une petite partie des trajectoires ont conduit le chromophore en dehors de la région de FS.

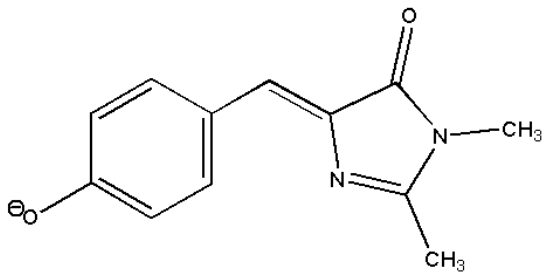


FIG. III.11.: Le HBDI (4'-hydroxybenzylidène-2,3-dimethylimidazolinone) - un modèle du chromophore de la GFP où la partie conjuguée est saturée par deux groupements méthyle liés aux atomes N3 et C1.

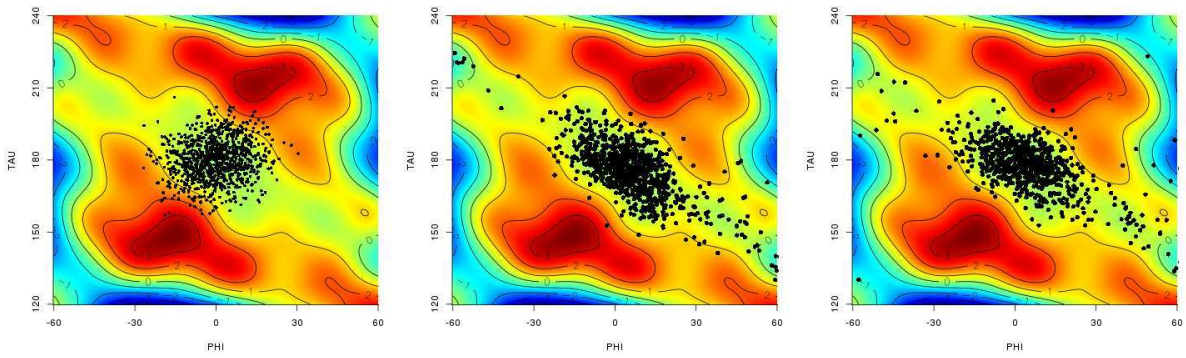


FIG. III.12.: Schéma du paysage énergétique de NADiA autour de la géométrie plane (FS) de l'isomère cis. Les conformations initiales des 1000 simulations à l'état excité sont représentées en noir dans le premier graphique et leur évolution après 0,5 ps et 1,0 ps dans le deuxième et troisième graphique, respectivement.

Dans toutes les simulations, le chromophore finit par quitter la géométrie plane (FS) et visite les autres minima sur la PES (TwP et TwI) (Fig. III.13). Le chromophore quitte la région de FS par une torsion de 90° en φ dans environ 95% des cas et par une torsion de 90° en τ dans les simulations restantes. La préférence pour la torsion en φ est cohérente avec une barrière plus faible et plus large que pour la torsion en τ. Les maxima de type HT ne sont jamais visités dans les simulations.

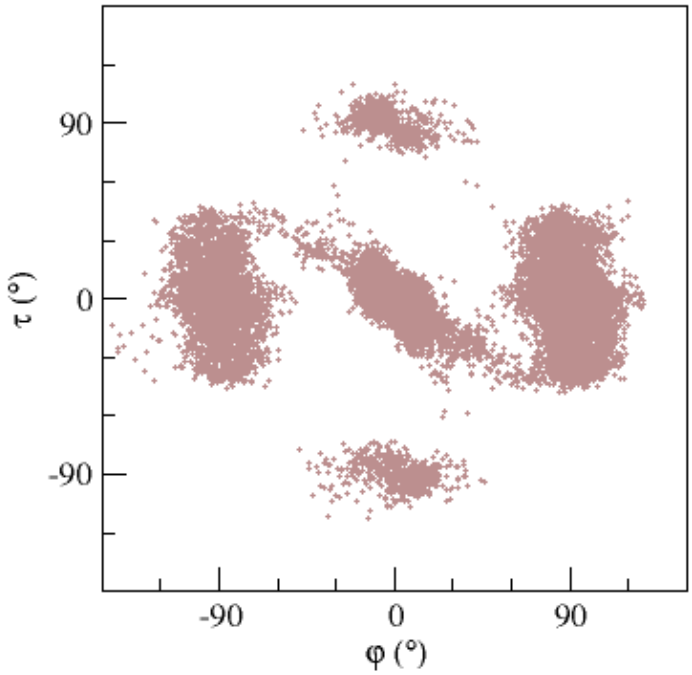


FIG. III.13.: Corrélation entre l'angle τ et l'angle φ au cours de 200 simulations du chromophore dans le vide.

On peut caractériser la cinétique de torsion par une durée de vie τ de la géométrie plane en utilisant le temps de premier passage (*first passage time* : FPT) à une géométrie perpendiculaire

(TwP ou TwI) observé dans chaque simulation. La distribution de FPT est montrée sur la figure III.15. Cette distribution doit correspondre à la loi exponentielle :

$$\rho(t) = A\, exp(-\frac{t}{\tau}) \tag{III.11}$$

où la durée de vie τ est égale au temps moyen de premier passage (MFPT). On obtient ainsi la valeur $\tau = 2,6\, ps$.

Le fait que le chromophore ne quitte pas immédiatement la région de la géométrie plane (FS) peut être expliqué par les sorties de la vallée qui jouent le rôle d'un goulet d'étranglement. Le chromophore peut s'échapper par les sorties dans un temps très court uniquement si les vitesses initiales le dirigent directement à travers ces passages étroits. Dans le cas inverse le chromophore va parcourir la vallée d'une façon aléatoire jusqu'à ce que la trajectoire finisse par atteindre une sortie.

Il n'existe pas de résultats expérimentaux sur la fluorescence du chromophore dans le vide, mais seulement en solution. Dans le cas de l'eau, le déclin de fluorescence indique une désactivation non-radiative à l'échelle de la picoseconde [74] - le même ordre de grandeur que notre résultat.

Une étude par dynamique *ab initio* SA2-CAS(2,2) montre un MFPT d'environ 0,25 ps [84]. La différence avec notre résultat (2,6 ps) peut venir des conditions initiales (énergie cinétique etc.) et du fait que les forces qui dirigent le chromophore vers les géométries perpendiculaires sont plus fortes avec le niveau de calcul CAS qu'avec le niveau CASPT2 (voir tab. III.1).

III.6. Analyse d'incertitude

Nous avons chercher à évaluer la précision du potentiel et la sensibilité du résultat (le MFPT) à travers deux analyses.

Dans un premier temps, nous avons cherché à estimer la précision de l'interpolation dans la région qui est visitée par le chromophore. Pour cela, nous avons évalué l'écart quadratique moyen (rmsd) entre le potentiel $V^{cis}_{\mathrm{NADiA}}(\tau, \varphi)$ et les valeurs d'énergie CASPT2, en pondérant les points de la grille par la fréquence à laquelle le chromophore passe à proximité au cours de la dynamique dans le vide.

Quand on considère toute la PES (calcul non-pondéré), le rmsd vaut 0,05 kcal/mol. Quand les points CASPT2 sont pondérés le rmsd vaut 2.10^{-4} kcal/mol dans la région de la géométrie plane et 6.10^{-4} kcal/mol dans les régions des géométries perpendiculaires. La limite entre les deux régions est définit par $|\tau|$ *ou* $|\varphi| = 60^{\circ}$ (la géométrie plane cis correspondant à $\tau = \varphi = 0^{\circ}$). On voit ainsi que la description de la PES dans la région visitée par le chromophore est encore plus précise que la description globale. Cette baisse de rmsd peut être expliquée par le fait que la différence entre la valeur de $V^{cis}_{\mathrm{NADiA}}(\tau, \varphi)$ et l'énergie CASPT2 est particulièrement faible dans les minima de la géométrie plane et des géométries perpendiculaires (Fig. III.14).

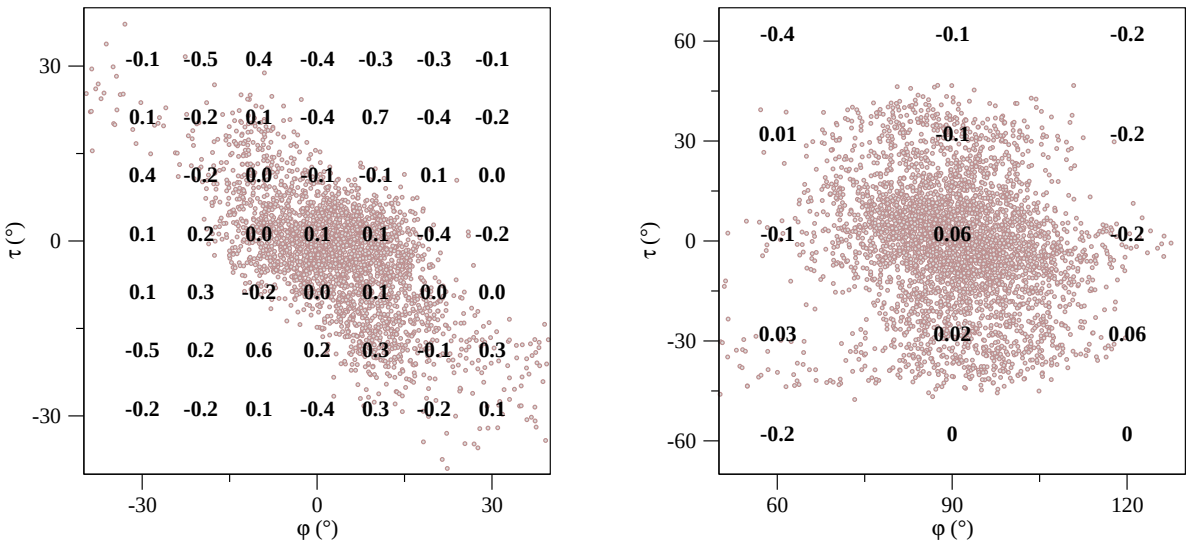


FIG. III.14.: Différence entre $V^{cis}_{\mathrm{NADiA}}(\tau, \varphi)$ et l'énergie CASPT2 sur les points de la grille en noir (valeurs en kcal/mol) et régions visitées au cours de la dynamique du chromophore dans le vide (Fig. III.13) en cercles marrons.

Dans un deuxième temps, nous avons analysé la sensibilité des prédictions produites par simulation MD par rapport au potentiel, en faisant varier les paramètres k_i de ce potentiel. Les variations sont aléatoires et leurs amplitudes sont ajustées afin de refléter l'incertitude sur les paramètres venant de l'imprécision de l'interpolation : un paramètre modifié k_i est engendré

selon une distribution gaussienne centrée sur sa valeur initiale avec un écart type σ_i (eq. III.12, [104, eq. 15.6.4]).

$$\sigma_i^2 = \frac{\chi^2}{C_{ii}} \tag{III.12}$$

où χ désigne et le rmsd de l'interpolation et C_{ii} l'élément diagonal de la matrice de covariance A des différentes fonctions $g(\tau, \varphi)$ (décrites dans section III.3.2) où p désigne l'indice des points de calcul CASPT2 (n=118) :

$$A_{ij} = \sum_{p=1,n} g_i(\tau_p, \varphi_p) g_j(\tau_p, \varphi_p) \tag{III.13}$$

Pour tenir compte de la corrélation entre les paramètres, nous avons appliqué l'équation III.13 à la combinaison linéaire de paramètres qui diagonalisent la matrice de covariance. Par ailleurs, nous avons introduit deux contraintes pour que la géométrie plane reste un minimum d'énergie.

Deux séries de 1000 simulations sont effectuées, l'une avec le potentiel initial et l'autre avec des potentiels modifiés. Les deux distributions des FPT sont ensuite comparées (Fig. III.15.). La modification des paramètres ne cause qu'un petit déplacement de la moyenne, 2,6 ps à 2,8 ps et aucun déplacement la médiane (2,0 ps), ni de l'intervalle de confiance à 50%, centré sur la médiane des deux distributions, qui vaut 2,6 ps (3,6-1,0 ps) pour les deux séries.

Nous concluons que le potentiel $V_{NADiA}(\tau, \varphi)$ est suffisamment robuste pour bien décrire la dynamique torsionelle du chromophore.

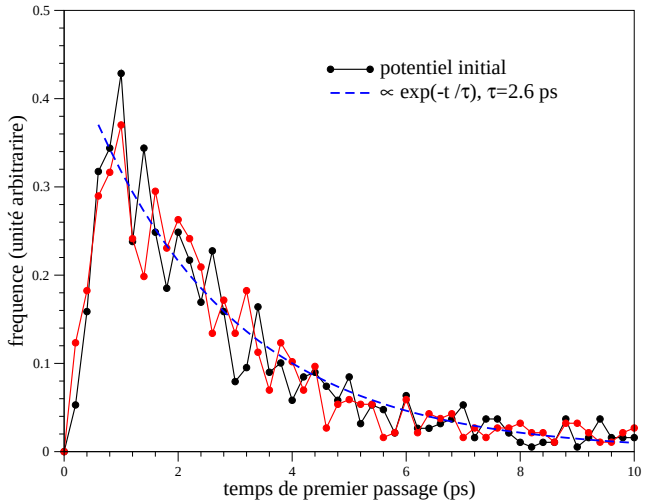


FIG. III.15.: Deux distributions de temps de premier passage (FPT) à une géométrie perpendiculaire (TwP ou TwI), basées sur 1000 simulations du chromophore dans le vide. Celle obtenue avec le potentiel initial ($V_{NADiA}(\tau, \varphi)$) est affichée en noir et celle obtenue avec des potentiels modifiés en rouge. Le déclin de FPT du potentiel initial est interpolé par une exponentielle (en bleu).

III.7. Potentiel de l'imidazolinone

III.7.1. Procédure

Au sein de la protéine, la torsion du chromophore est en général fortement inhibée par son

environnement. Cependant, le potentiel interne du chromophore peut, à partir d'une géométrie critique, induire une force importante tendant au contraire à amener le chromophore vers une géométrie perpendiculaire. L'affrontement de ces deux forces opposées peut entraîner des déformations des deux cycles du chromophore. La partie la plus susceptible d'être affectée est le cycle imidazolinone car il est lié au reste de la protéine par des liaisons covalentes et subit donc des contraintes supplémentaires. Nous avons déterminé un potentiel de torsions spécifique pour ce cycle, plutôt que d'utiliser des paramètres du champ de force (ff) de Amber. Le potentiel final est limité à cinq angles dièdres et quatre angles dièdres impropres (Fig. III.16). Les autres paramètres du champ de force de l'imidazolinone n'ont pas été remis en cause.

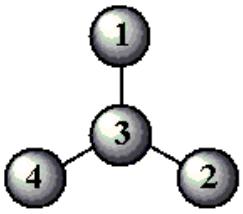


FIG. III.16.: L'angle dièdre impropre 1-2-3-4, où l'angle de pyramidalisation de l'atome 3, est l'angle entre le plan défini par les atomes 1, 2 et 3, et le plan défini par les atomes 2, 3 et 4.

Nous faisons l'approximation que le potentiel de torsion de l'imidazolinone change peu entre l'état fondamental (S_0) et l'état excité (S_1).

Pour déterminer le nouveau champ de force nous avons pris comme référence les résultats d'une simulation QM/MD de l'état S_0, combinant théorie de la fonctionnelle de la densité (DFT) et dynamique moléculaire (MD) du chromophore anionique à 300K. Afin de reproduire la flexibilité du cycle imidazolinone, nous avons calibré les termes de torsion du champ de forces de façon à reproduire au mieux les fluctuations de la géométrie du cycle dans la simulation QM/MD, au cours d'une dynamique moléculaire classique du chromophore à l'état fondamental.

III.7.2. Simulation DFT/MD

La simulation DFT/MD du chromophore à l'état S_0 a été effectuée par Aurélien de la Lande avec une version locale du programme *deMon2k* [105]. Le calcul quantique était traité par la fonctionnelle PBE (échange et corrélation) et une base double zeta avec polarisation. La température du système était maintenue à 300 K par un thermostat Nosé-Hoover avec une fréquence de couplage de 100 cm^{-1}. Le temps de simulation était de 10 ps avec un pas de temps de 0,5 fs. Les premières 1,5 ps, où la température n'est pas stable, ont été exclues de la calibration du potentiel. Le modèle de chromophore utilisé comporte la partie conjuguée saturée par deux groupements méthyle liés aux atomes N3 et C1 (modèle HBDI Fig. III.11).

Au cours de la simulation, les valeurs moyennes des angles dièdres sont toutes proches de 0° ou 180°, avec un écart maximum de 1°. Les fluctuations (c'est-à-dire les écarts quadratiques moyens σ) sont données dans le tableau III.4 (colonne DFT/MD). Les fluctuations des angles dièdres *intra* (ne faisant intervenir que les cinq atomes du cycle) (Fig. III.17) sont dans l'intervalle 3,8°-5,6° alors que celles des angles dièdres *extra* (faisant intervenir au moins un

atome hors cycle) sont nettement plus grandes, dans l'intervalle 8,0°-16,6°. On note entre autres que la fluctuation de CA3-N3-C1-N2 (13,1°) est presque trois plus grande que celle de C2-N3-C1-N2 (5,0°) (Fig. III.17), ce qui est en accord avec les grandes fluctuations de l'angle de pyramidalisation de N3 (15,1°).

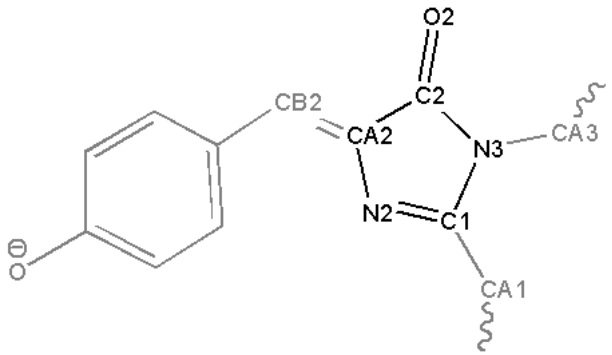


FIG. III.17.: Schéma des atomes du cycle imidazolinone (en noir) et le reste du chromophore (en gris).

III.7.3. Simulation MD avec un champ de forces standard

Nous avons effectué une simulation MD classique du chromophore à l'état S_0 de 10 ps à énergie constante et des vitesses initiales générées suivant une distribution de Boltzmann à 300K. Le champ de force (ff) était de type standard, construit par analogie avec le champ de forces de Amber. Dans ce champ de forces, la même constante de force est attribuée à tous les angles dièdres d'une liaison donnée (2 ou 4 angles suivant les cas). Pour les angles dièdres impropres décrivant la pyramidalisation des atomes CA2, C1, N3 et C2, nous avons choisi des constantes de force analogues à celle des résidus standard dans le champ de force de Amber : 1,1 kcal/mol pour les premiers et 10,5 kcal/mol pour C2. Les valeurs d'équilibre correspondent à un cycle plan.

On trouve ainsi les fluctuations des angles dièdres *intra* dans l'intervalle [5,1°-7,5°] et celles des angles dièdres *extra* dans l'intervalle [5,4°-10,5°] (Tab. III.4, colonne $\mathrm{MD}_{\mathrm{ff\,standard}}$). Comparées aux valeurs des simulations DFT/MD, les fluctuations des angles dièdres *intra* sont trop fortes et celles des angles dièdres *extra* trop faibles. En particulier, les fluctuations de CA3-N3-C1-N2 et de C2-N3-C1-N2 sont presque égales : 7,9° et 7,3°, respectivement, alors qu'elles sont très différentes dans les simulation DFT/MD.

TAB. III.4.: Les constantes de force, k, données en kcal/mol, et les fluctuations σ, données en degrés, des angles dièdres de l'imidazolinone. Les résultats sont issus d'une simulation DFT/MD et de trois simulations MD classiques. Le champ de force (ff) utilisé dans chaque simulation est un ff de type standard (ff standard), une version intermédiaire (ff inter) ou la version finale (ff final) de la calibration du potentiel, respectivement.

angle dièdre	DFT/MD	$MD_{ff\ standard}$		$MD_{ff\ inter}$		$MD_{ff\ final}$	
	σ	k	σ	k	σ	k	σ
intra							
C2-N3-C1-N2	**5,0**	1,7	**7,3**	5,0	**7,3**	8,0	**4,7**
C2-CA2-N2-C1	**4,5**	2,4	**6,3**	5,0	**7,6**	12,0	**4,4**
N3-C1-N2-CA2	**3,8**	5,0	**7,5**	5,0	**7,6**	16,0	**4,0**
N3-C2-CA2-N2	**5,6**	5,375	**5,1**	5,0	**7,3**	5,5	**5,0**
CA2-C2-N3-C1	**5,5**	1,45	**5,5**	5,0	**6,7**	5,5	**4,9**
extra							
CA3-N3-C1-N2	**13,1**	1,7	**7,9**	0	**16,4**	0	**14,0**
CA3-N3-C1-CA1	**16,6**	1,7	**10,5**	0	**19,3**	0	**16,4**
CA3-N3-C2-O2	**12,2**	1,45	**7,8**	0	**17,1**	0	**14,5**
CA3-N3-C2-CA2	**12,0**	1,45	**7,3**	0	**16,4**	0	**13,9**
CA1-C1-N2-CA2	**8,1**	5,0	**8,2**	0	**11,8**	0	**8,5**
CA1-C1-N3-C2	**9,0**	1,7	**8,6**	0	**11,7**	0	**9,2**
CB2-CA2-N2-C1	**9,0**	2,4	**6,8**	0	**11,9**	0	**9,1**
CB2-CA2-C2-N3	**9,2**	5,375	**5,8**	0	**11,3**	0	**8,9**
CB2-CA2-C2-O2	**10,7**	5,375	**6,1**	0	**12,1**	0	**9,9**
O2-C2-N3-C1	**7,5**	1,45	**5,4**	0	**7,7**	0	**6,0**
O2-C2-CA2-N2	**8,0**	5,375	**5,4**	0	**8,7**	0	**6,4**
impropre							
pyr. de CA2	**7,3**	1,1	**6,6**	6,0	**7,6**	6,0	**6,9**
pyr. de C2	**4,4**	10,5	**5,6**	30,0	**4,6**	30,0	**4,4**
pyr. de N3	**15,1**	1,1	**8,6**	4,0	**12,8**	4,0	**11,9**
pyr. de C1	**7,9**	1,1	**9,8**	10,0	**8,5**	10,0	**7,7**

III.7.4. Calibration du potentiel

Nous avons à représenter deux types d'effets venant de la structure électronique de l'imidazolinone : l'interaction entre liaisons vicinales (de type de celles qui sont responsables de la barrière de rotation de l'éthane) et le changement d'hybridation d'un atome (e.g. la barrière d'inversion de l'ammoniac). Dans le champ de forces de type standard les paramètres

de torsion contribuent aux deux effets. Pour simplifier l'optimisation des paramètres, nous avons développé un nouveau potentiel dans lequel on sépare les déformations du cycle (angles dièdres *intra*) et les changements d'hybridation, qui sont représentés uniquement par les angles dièdres impropres. Les fluctuations des angles dièdres *extra* sont donc uniquement déterminées par les angles impropres et par le reste du champ de force. Ce modèle ne compte que neuf paramètres (les constantes de force de cinq angles dièdres *intra* et quatre angles dièdres impropres), comme le champ de forces standard.

Des calculs exploratoires nous ont montré que les fluctuations des angles dièdres impropres sont moins sensibles aux variations des constantes de force des angles dièdres (*intra*) que l'inverse, i.e. les fluctuations des angles dièdres aux variations des constantes de force des angles dièdres impropres. Ce qui nous a conduit à adopter une démarche en deux temps : calibrer les constantes de force des angles dièdres impropres sur les fluctuations des angles dièdres impropres de la simulations DFT/MD, puis celles des angles dièdres *intra* sur les fluctuations des angles dièdres *intra*.

Nous avons ainsi commencé par un champ de force initial où toutes les constantes des angles dièdres *intra* valent 5,0 kcal/mol et celles des angles impropres valent 0,0 kcal/mol. La calibration des constantes des angles dièdres impropres décrivant la pyramidalisation des atomes CA2, C2, N3, C1 aboutit aux valeurs 6, 30, 4 et 10 kcal/mol, respectivement (Tab. III.4, colonne $\mathrm{MD_{ff\,inter}}$). Ces valeurs reproduisent les fluctuations avec une différence relative de 4 % à 10 %, sauf les fluctuations de N3 qui sont 15 % trop petites. Une constante de force plus faible sur l'angle impropre de N3 donnait des fluctuations beaucoup trop grandes des angles dièdres *extra*.

Dans la deuxième étape de la calibration, les constantes de force des angles dièdres *intra* sont ajustées en gardant les constantes de force des angles dièdres impropres fixes (Tab. III.4, colonne $\mathrm{MD_{ff\,final}}$). La différence relative entre les fluctuations DFT/MD et les fluctuations MD du potentiel final devient alors inférieure à 10 % pour la majorité des angles, à part quatre exceptions : O2-C2-CA2-N2 (fluctuations 20 % trop faibles), CA2-C2-N3-CA3 (fluctuations 15 % trop fortes), O2-C2-N3-CA3 (fluctuations 20 % trop fortes) et l'angle impropre de N3 (fluctuations 20 % trop faibles). Le rapport entre les fluctuations de C2-N3-C1-N2 et de CA3-N3-C1-N2 (4,7° et 14,0°) est en accord avec les valeurs DFT/MD.

III.8. Autres paramètres du champ de force

III.8.1. Acide aminé non-standard

Le chromophore n'est pas considéré comme un résidu standard dans le champ de force Amber 1999 (ff99). Il faut donc créer un champ de force spécifique pour l'état fondamental (S_0) et l'état excité (S_1) du chromophore.

Nous devons donc optimiser des charges ponctuelles pour l'état S_0 et S_1 du chromophore. Ces calculs ont été effectués sur le chromophore de la GFP, YFP et Padron et aussi sur le modèle du chromophore HBDI (Fig. III.11), utilisé dans les simulations dans le vide, sous sa forme isomère *cis* et *trans* (section III.8.4). Nous devons également définir les paramètres des termes concernant les longueurs des liaisons, les angles de valence, les angles dièdres, les termes électrostatiques et les termes de van der Waals.

Le chromophore étant le résultat d'une condensation de trois acides aminés standard, il contient des parties pour lesquelles les paramètres sont directement disponibles dans le champ de force de Amber. C'est le cas des chaînes latérales des résidus 65 et 67 (groupements R sur la figure III.18) et du cycle phénolate du résidu 66. Les valeurs d'équilibre des distances, des angles et des torsions dans ces parties de type standard sont définies par des paramètres du champs de force de Amber.

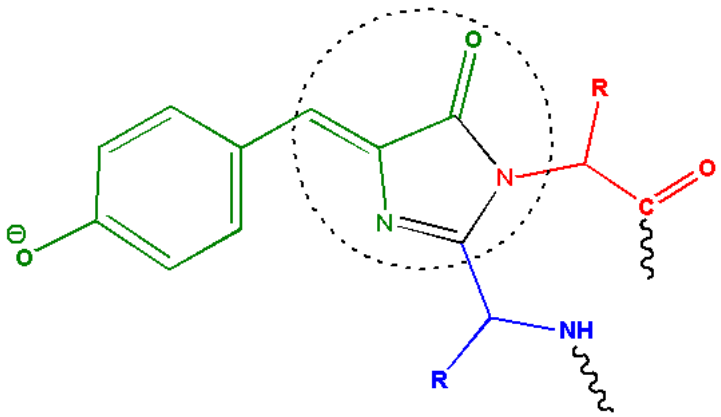


FIG. III.18.: Schéma du chromophore anionique. Les atomes issus des résidus 65, 66, et 67 sont affichés en bleu, vert et rouge respectivement. Nous avons créé des nouveaux types d'atomes pour les six atomes de type non-standard (partie entourée par un cercle).

Pour décrire la partie non-standard du chromophore nous avons créé six nouveaux type d'atome (Fig. III.18). Toutes les coordonnées (liaison, angle de valence, angle dièdre et angle impropre) qui contiennent un de ces types ont dû être paramétrisées. Leurs valeurs d'équilibre ont été déterminées à l'état S_0 et S_1 par un calcul d'optimisation de géométrie de la géométrie plane du chromophore au niveau B3LYP/6-31G* et CIS/6-31G*, respectivement. Les constantes de force ont été définies en analogie avec le champ de forces de Amber des types d'atomes ressemblants (cycle aromatique, double liaison etc.). Le développement du potentiel de torsion du cycle imidazolinone est décrit dans la section III.7. Les valeurs d'équilibre des angles des liaisons centrales du chromophore ont demandé une définition particulière qui est décrite dans la section III.8.3.

III.8.2. Angles dièdres

Le potentiel de torsion des angles dièdres τ et φ a été établi pour les états S_0 et S_1 par des méthodes différentes. Le développement du potentiel de l'état S_1 NADiA est décrit dans la section III.3. Les potentiels de τ et φ de l'état S_0 sont utilisés dans le but d'étudier la stabilité du réseau de liaisons H autour du chromophore et d'engendrer un grand nombre de géométries de départ pour des simulations à l'état S_1. Nous pouvons donc nous limiter à une description de la région autour de la géométrie plane. Nous avons utilisé des potentiels sous la forme classique du champ de force de Amber - des potentiels additif où chaque angle dièdre est décrit par une fonction cosinus :

$$U_{torsion} = k(1 + cos(n\omega - \gamma) \tag{III.14}$$

où k représente la constante de force, ω l'angle dièdre τ ou φ, et γ la phase.

Pour déterminer les constantes de force associées aux deux torsions τ et φ, nous avons effectué des calculs B3LYP/6-31G* pour des angles variant de 0° à 40°, par pas de 10°, avec optimisation de géométrie. Puis nous avons ajusté les constantes de force sur les valeurs obtenues. Nous avons obtenu ainsi : k_τ=10,0 kcal/mol et k_φ=10,0 kcal/mol (k_τ correspond à l'énergie totale associée aux différents angles dièdres associé à la torsion τ, et il en est de même pour k_φ). Ces deux valeurs sont identiques. Reuter *et al.* [132], ont trouvé par contre des valeurs qui différent entre elles d'un facteur 1,5 : k_τ=3,9 kcal/mol et k_φ=2,7 kcal/mol (correspond aux angles dièdres individuels). Ces auteurs ont utilisé des valeurs des angles jusqu'à 90° alors que nous n'utilisons que des valeurs jusqu'à 40°, ce qui pourrait être une origine de la différence entre leurs résultats et les nôtres.

III.8.3. Angles de valence

L'énergie potentielle d'un angle de valence est donnée par l'expression (eq. III.15)

$$U_{angle} = K_\theta(\theta - \theta_{eq})^2 \tag{III.15}$$

où θ_{eq} est l'angle d'équilibre et K_θ la constante de force. Les valeurs d'équilibre sont des valeurs uniques et indépendantes des autres paramètres, notamment de τ et φ. Or, les calculs quantiques (SA-2-CAS(4,3)) montrent que les angles du pont du chromophore varient fortement avec τ et φ. C'est essentiellement l'angle CA2-CB2-CG2 qui varie (Fig. III.3), il est maximal dans les géométries planes (127° en cis et 130° en trans) et minimal lorsque τ et φ vaut 90° (125° en TwI et 123° en TwP). Il faut choisir une valeur d'équilibre qui corresponde au mieux à ce que l'on veut étudier. Le choix d'angle d'équilibre présente deux limites. Premièrement, un angle trop petit cause un clash stérique entre les deux cycles du chromophore dans la conformation plane. On peut noter que c'est pour cette raison que l'angle diffère entre l'isomère cis et trans. Deuxièmement, un angle trop grand rend le chromophore plus volumineux et cela peut défavoriser les torsions du chromophore au sein de la protéine. Les angles du pont du chromophore jouent également un rôle important dans l'emplacement de l'oxygène du phénol. Sa position est essentielle dans la dynamique car l'atome crée plusieurs liaisons H qui donnent un ancrage du chromophore. Le choix de valeur d'angle d'équilibre est donc important mais contraint.

La torsion simultanée de τ et φ qui aboutit à la structure appelée papillon, résulte en une grande modification des angles du pont ($\sim$ 15°). Des calculs quantiques montrent cependant que le papillon se trouve 5 kcal/mol au-dessus de la géométrie plane et qu'il est peu visité. Au contraire, une torsion de 90° en τ ou en φ sont, respectivement, 10 kcal/mol et 6 kcal/mol plus stable que la géométrie plane. On peut s'attendre à ce que la torsion simultanée de τ et φ soit un événement rare par rapport à une torsion simple en τ ou en φ. Nous pouvons donc exclure la géométrie du papillon de l'optimisation des angles d'équilibre.

Dans cette étude, nous voulons étudier la dynamique du chromophore dans la protéine à l'état excité à partir d'une conformation plane jusqu'au moment où il se tord. Nous pouvons définir deux jeux d'angle d'équilibre différents, où chacun est adapté pour un isomère plan (*cis* ou *trans*) et les conformations perpendiculaires en τ ou φ accessibles pour cet isomère. L'angle

d'équilibre de CA2-CB2-CG2 peut ensuite être défini comme la moyenne entre l'angle dans la conformation plane et l'angle dans les conformations perpendiculaires. Cette moyenne diffère uniquement de quelques degrés de l'angle optimal des conformations perpendiculaires et ne devrait pas influencer la torsion du chromophore. Cet angle d'équilibre est néanmoins quelques degrés plus petit que l'angle optimal de la conformation plane et crée un *clash* stérique entre les deux cycles du chromophore. Comme décrit dans la section III.3.2 ce problème est contourné en utilisant le potentiel NADiA.

La constante d'équilibre K_θ a été définie comme 63 kcal/mol en analogie avec des acides aminés standards. On aurait pu diminuer sa valeur afin de permettre une variation plus souple entre les différentes géométries du chromophore. Cependant, une souplesse exagérée du chromophore peut également induire des interactions peu probables avec l'environnement protéique et ainsi fausser la dynamique.

III.8.4. Charges effectives

L'interaction électrostatique entre deux molécules dépend de leurs densités électroniques. Dans un champ de forces, la densité électronique est le plus souvent décrite par un ensemble de charges effectives ponctuelles localisées sur les atomes. Il existe différents niveaux de complexité puisque la charge effective peut être fixe ou variable, i.e. polarisable, et décrite par une ou plusieurs charges ponctuelles centrées sur l'atome, sur les liaisons ou en dehors.

Dans le champ de force *ff99* de Amber, les charges atomiques sont déterminées par la méthode RESP (Restrained ElectroStatic Potential fit) [106]. Dans cette méthode, les charges sont ajustées de façon à ce qu'elles reproduisent au mieux le potentiel électrostatique créé par la molécule. Ce potentiel électrostatique est établi par un calcul ab initio, au niveau HF/6-31G* sur une grille de points autour la molécule. Dans le processus d'ajustement, des contraintes hyperboliques sont introduites sur les charges des atomes lourds, avec une valeur cible égale à 0. Ces contraintes permettent entre autres d'éviter la présence de charges de grande amplitude, sans signification physique.

Nous utilisons ici une méthode développée au laboratoire [107, 108] très proche de la méthode RESP. Les charges sont ajustées sur le potentiel électrostatique calculé au niveau RHF/6-31G* pour l'état fondamental et CIS/6-31G* pour l'état excité, avec la possibilité d'ajouter toutes sortes de contraintes harmoniques.

Le modèle de chromophore utilisé comporte les parties qui le lient à la protéine, les liaisons pendantes créées par la coupure des liaisons peptidiques étant saturées par des groupements COH et NH_2, appelés groupements terminaux (Fig. III.19). La procédure d'ajustement est faite sur le chromophore dans la géométrie où sa partie conjuguée est plane. Nous avons également calculé des charges pour le chromophore dans le vide, et utilisé pour cela le modèle où la partie conjuguée est saturée par des groupements méthyle (Fig. III.19).

Les contraintes introduites dans l'ajustement sont les suivantes. La charge totale de chaque groupement terminal, CH_3 dans le modèle HBDI et COH et NH_2 dans les autres chromophores, est contrainte à zéro. Pour éviter des discontinuités dans la chaîne peptidique, les charges des

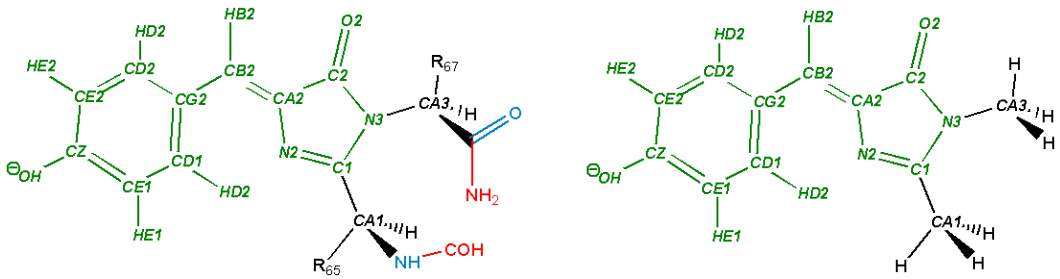


FIG. III.19.: A gauche : schéma de la géométrie utilisée dans le calcul de charges pour le chromophore de la GFP, la YFP et Padron. Les liaisons pendantes créées par la coupure des liaisons peptidiques (en bleu) sont saturées par des groupements terminaux (en rouge). Les groupements R_{65} et R_{67} représentent les chaînes latérales de Thr65 et Gly67 dans la GFP, Gly65 et Gly67 dans la YFP et Cys65 et Gly67 dans Padron.
A droite ; HBDI (4'-hydroxybenzylidène-2,3-dimethylimidazolinone) - un modèle du chromophore de la GFP, utilisé dans les simulations dans le vide.

atomes qui font partie de cette chaîne, qui sont de type standard, sont contraintes aux valeurs du champ de force de Amber. Des contraintes harmoniques, centrées sur les valeurs du champ de forces de Amber, sont imposées sur les atomes des chaînes latérales des résidus 65 et 67. Une contrainte harmonique légère, centrée sur zéro, est imposée sur chaque atome pour éviter des grandes valeurs non-physiques.

Les charges S_0 et S_1 des atomes communs du chromophore de la GFP, de Padron et du modèle HBDI du chromophore (les charges de la YFP dans cette partie étant identiques à celles de la GFP) sont présentés dans le tableau III.6. Une comparaison entre les charges des chaînes latérales des résidus 65 et 67 du champ de forces de Amber et les charges issues des calculs sur le chromophore de la GFP, de la YFP et Padron est présentée dans le tableau III.7, III.8 et III.9, respectivement.

TAB. III.5.: Les charges S_0 de la partie commune entre le chromophore anionique de la GFP (*cis*), de la YFP (*cis*), de Padron (*trans*) et le modèle HBDI du chromophore (*cis* et *trans*) (Fig. III.19). Nous avons utilisé les charges de la GFP pour décrire cette partie du chromophore dans la YFP (*cis*).

atome	GFP *cis*	Padron *trans*	HBDI *cis*	HBDI *trans*
N2	-0,4102	-0,5903	-0,3914	-0,6585
C1	0,1318	0,4785	0,2318	0,4378
N3	0,0683	-0,2000	-0,0491	-0,2048
C2	0,3785	0,3774	0,4051	0,4637
O2	-0,5789	-0,4719	-0,6158	-0,5606
CA2	-0,0307	0,0316	-0,0986	0,0217
CB2	-0,1238	-0,1405	-0,0714	-0,0575
HB2	0,1293	0,1360	0,1311	0,1377
CG2	0,0034	-0,0365	-0,0609	-0,1494
CD1	-0,1225	-0,0560	-0,0629	-0,0315
HD1	0,1457	0,1076	0,1086	0,1088
CE1	-0,4186	-0,4328	-0,4581	-0,4589
HE1	0,1416	0,1406	0,1424	0,1406
CZ	0,7126	0,7314	0,7237	0,7327
OH	-0,7032	-0,7070	-0,7252	-0,7252
CE2	-0,4186	-0,4328	-0,4581	-0,4589
HE2	0,1416	0,1406	0,1424	0,1406
CD2	-0,1225	-0,0560	-0,0629	-0,0315
HD2	0,1457	0,1076	0,1086	0,1088
CA1	-0,0646	0,0118	-0,2699	-0,2866
CA3	-0,5831	-0,4422	-0,1408	-0,1294

TAB. III.6.: Les charges S_1 de la partie commune entre le chromophore anionique de la GFP (*cis*), de Padron (*trans*) et le modèle HBDI du chromophore (*cis* et *trans*) (Fig. III.19). Nous avons utilisé les charges de la GFP pour décrire cette partie du chromophore dans la YFP (*cis*).

atome	GFP *cis*	Padron *trans*	HBDI *cis*	HBDI *trans*
N2	-0,4976	-0,5958	-0,4829	-0.6819
C1	0,1226	0,3335	0,2678	0,4399
N3	0,0392	-0,0874	-0,0460	-0,2122
C2	0,4051	0,4430	0,4924	0,5082
O2	-0,5712	-0,6139	-0,6548	-0,6279
CA2	0,1596	0,2628	0,0523	0,1881
CB2	-0,3601	-0,5277	-0,3793	-0,4214
HB2	0,1346	0,1716	0,1303	0,1514
CG2	0,2211	0,31759	0,2529	0,1441
CD1	-0,2391	-0,2410	-0,2180	-0,1459
HD1	0,1665	0,1601	0,1551	0,1499
CE1	-0,3364	-0,3796	-0,3814	-0,4491
HE1	0,1372	0,1516	0,1494	0,1605
CZ	0,6510	0,7534	0,7029	0,7886
OH	-0,6894	-0,7618	-0,7567	-0,7792
CE2	-0,3364	-0,3796	-0,3814	-0,4491
HE2	0,1372	0,1516	0,1494	0,1605
CD2	-0,2391	-0,2410	-0,2180	-0,1459
HD2	0,1665	0,1601	0,1551	0,1499
CA1	-0,0652	0,0166	-0,2249	-0,2422
CA3	-0,5581	-0,6283	-0,1818	-0,0737

TAB. III.7.: Comparaison des charges des atomes appartenant aux résidus Thr65 et Gly67 dans le chromophore de la GFP et les valeurs cibles du calcul - les charges du champ de force de Amber.

Amber			GFP	
résidu	atome	charge	atome	charge
Thr65	N	-0,4157	N1	-0,4157
	H	0,2719	H1	0,2719
	CA	-0,0389	CA1	-0,06524
	HA	0,1007	HA1	0,09669
	CB	0,3654	CB1	0,37657
	HB	0,0043	HB1	0,00507
	CG2	-0,2438	CG1	-0,24065
	3×HG	0,0642	3×H	0,06024
	OG1	-0,6761	OG1	-0,66320
	HG1	0,4102	HG1	0,40244
Gly67	CA	-0,0252	CA3	-0,55807
	2×HA	0,0698	2×HA	0,25448
	C	0,5973	C3	0,5973
	O	-0,5679	O3	-0,5679

TAB. III.8.: Comparaison des charges des atomes appartenant aux résidus Gly65 et Gly67 dans le chromophore de la YFP et les valeurs cibles du calcul - les charges du champ de force de Amber.

Amber			YFP	
résidu	atome	charge	atome	charge
Gly65	N	-0,41570	N1	-0,4157
	H	0,27190	H1	0,2719
	CA	-0,02520	CA1	-0,06524
	2×HA	0,06980	2×HA	0,07882
Gly67	CA	-0,02520	CA3	-0,55807
	2×HA	0,06980	2×HA	0,25448
	C	0,59730	C3	0,5973
	O	-0,56790	O3	-0,5679

Tab. III.9.: Comparaison des charges des atomes appartenant aux résidus Gys65 et Gly67 dans le chromophore de Padron et les valeurs cibles du calcul - les charges du champ de force de Amber.

Amber			Padron	
résidu	atome	charge	atome	charge
Cys65	N	-0,4157	N1	-0,4157
	H	0,2719	H1	0,2719
	CA	0,0213	CA1	0,0166
	HA	0,1124	HA1	0,1481
	CB	-0,1231	CB1	-0,1553
	2×HB	0,1112	2×HB	0,1418
	SG	-0,3119	SG1	-0,3384
	HG	0,1933	HG1	0,1902
Gly67	CA	-0,0252	CA3	-0,6283
	2×HA	0,0698	2×HA	0,2605
	C	0,5973	C3	0,5973
	O	-0,5679	O3	-0,5679

III.9. Conclusion du chapitre

Nous avons déterminé un champ de forces complet décrivant la structure électronique du chromophore dans l'état excité S_1. Deux ensembles de degrés de liberté ont fait l'objet de développements spécifiques : le potentiel décrivant les angles dièdres couplés τ et φ et les déformations du cycle imidazolinone. La stratégie employée est aisément transposable à d'autre molécules et facilement intégrable dans la version d'Amber que nous avons modifié. Le champ de forces est l'outil indispensable pour mener des simulations de dynamique moléculaire du chromophore à l'état excité sur l'échelle de temps allant de la picoseconde à la nanoseconde.

IV. Mécanismes moléculaires de torsion du chromophore dans les protéines fluorescentes

IV.1. Analyse de la GFP

L'objectif de cette analyse est d'identifier les mouvements de la protéine qui permettent une torsion du chromophore pouvant mener à une désexcitation non-radiative. Nous avons dans un premier temps cherché à identifier les résidus susceptibles de bloquer la torsion du chromophore en analysant la structure cristalline et une dynamique de la GFP avec le chromophore à l'état fondamental. Nous avons ensuite étudié l'influence réciproque des mouvements de la protéine et du chromophore à l'état excité.

IV.1.1. L'environnement du chromophore

Dans la structure cristalline, le chromophore est entouré par un réseau de liaisons. D'une part, deux liaisons covalentes font la jonction avec la protéine, d'autre part, plusieurs liaisons hydrogène (liaisons H) qui sont déduites des distances entre atomes lourds : l'oxygène O2 de l'imidazolinone (Fig. IV.1) avec Arg96 et Gln94 et l'azote N2 avec la partie non-conjuguée partie du chromophore issue de Thr65 (Fig. IV.2). De plus, l'oxygène du phénolate est lié par trois liaisons H avec His148, Thr203 et une molécule d'eau (W1).

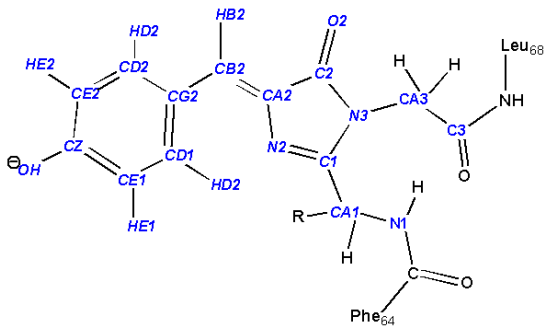


FIG. IV.1.: Le chromophore de la GFP (formé par le tripeptide Thr65-Tyr66-Gly67) avec les noms des atomes indiqués en bleu. La chaîne latérale de Thr65 est indiqué par un R. Le groupement carboxyle de Phe64 et le groupement NH de Leu68 sont également montrés sur le schéma.

Les mouvements du chromophore sont également contraints par des encombrements stériques : la chaîne latérale de Thr62 et le groupement méthyle de Thr203 qui s'étendent de chaque côté du

phénolate (Fig. IV.3) et un agrégat de cinq molécules d'eau placé du même côté que Thr203. Cet agrégat comporte une première couche de trois molécules d'eau (W2-W4, Fig. IV.2) qui forment des liaisons H entre elles mais également avec Glu222, Gln69 et Val68, et il comporte deux molécules supplémentaires à plus grande distance. Cet ensemble occupe une cavité de grande taille (le volume, mesuré par *pocket finder* [120], est égal à 138 $Å^3$) qui n'est pas en contact avec le solvant.

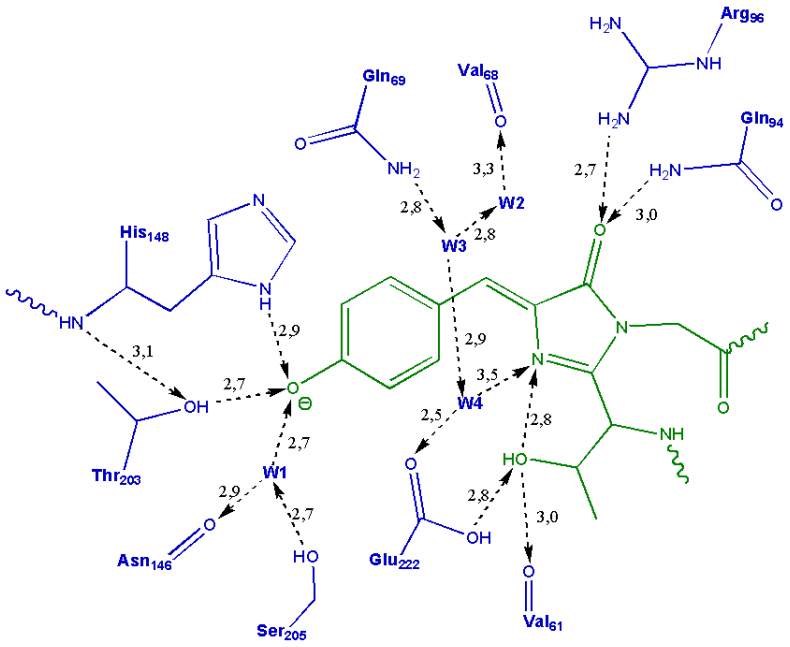


FIG. IV.2.: Le chromophore de la GFP (en vert) et une partie du réseau de liaisons hydrogène dans son environnement que l'on peut supposer à partir de la structure cristalline [43]. Les liaisons H sont indiquées par des flèches qui pointent du donneur vers l'accepteur de proton. Si le donneur a plusieurs possibilités de créer une liaison H, e.g. le groupement hydroxyle du chromophore (chaîne latérale de Thr65), chacune des liaisons possibles est indiquée. La lettre W représente une molécule d'eau. Les lignes ondulées représentent les deux jonctions covalentes avec le reste de la protéine. Pour chaque liaison H, le nombre indiqué est la distance en Ångström entre les atomes lourds, observée dans la structure cristalline.

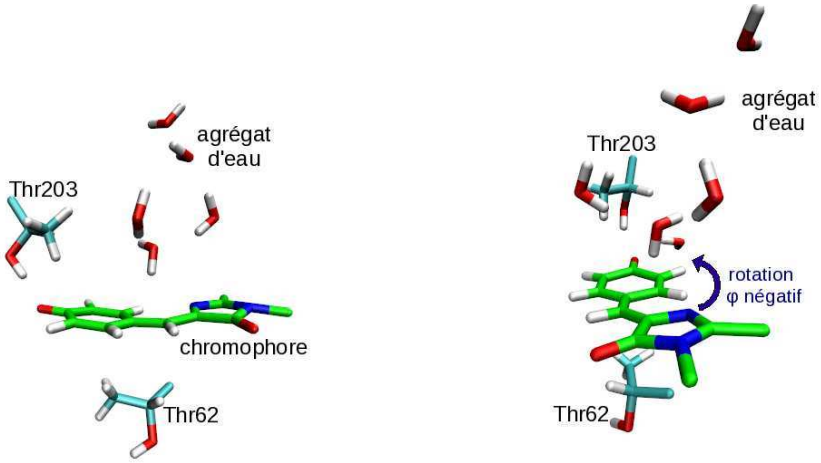


FIG. IV.3.: L'encombrement stérique dans le voisinage immédiat du chromophore. Le sens de la rotation en φ est indiqué sur la figure à droite.

IV.1.2. Dynamique moléculaire à l'état fondamental

Nous avons effectué une série de simulations de dynamique moléculaire (MD) de la GFP avec le chromophore dans sa forme anionique en partant des coordonnées de la structure cristalline [43].

Les acides aminés 'titrables' ont été choisis dans leurs état de protonation standard à pH neutre, excepté Glu222. L'état de protonation de ce résidu est sujet à controverse [132]. Il est souvent considéré comme l'accepteur final lors de la perte de proton du chromophore, donc protoné, ce qui est en accord avec la structure cristalline, mais en désaccord avec des études vibrationnelles. Nous avons choisi de nous appuyer sur la structure cristalline qui suggère que le groupement carboxyle de Glu222 forme une liaison H avec le groupement hydroxyle du chromophore (issu de la chaîne latérale de Thr65) (Fig. IV.2). Glu222 doit être le donneur de proton dans cette interaction puisque le proton du groupement hydroxyle du chromophore est engagé dans l'une ou l'autre des liaisons H suivantes : avec l'oxygène peptidique de Val61 ou avec l'azote N2 du chromophore. Nous avons donc choisi Glu222 protoné dans nos simulations.

Après la mise à l'équilibre du système (le protocole utilisé est décrit dans le section C.6), une simulation de 20 ns avec le chromophore à l'état fondamental (S_0) est effectuée. La structure globale de la protéine reste relativement stable : l'écart quadratique moyen (*root mean square deviation* - rmsd) des carbones α des brins β par rapport à la structure cristalline est de $0,8 \pm 0,1$ Å (Fig. IV.4).

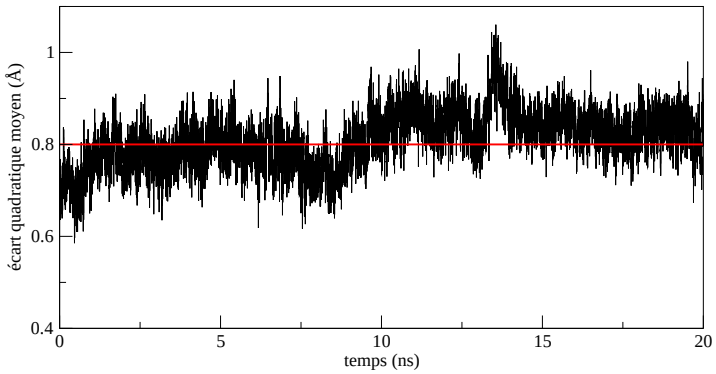


FIG. IV.4.: L'écart quadratique moyen entre les carbones α des brins β au cours de la mise en équilibre et de la simulation de 20 ns à l'état fondamental de la GFP, avec comme référence la structure cristalline.

Les valeurs moyennes des angles dièdres τ et φ sont nulles, avec des fluctuations de 7° pour chaque angle. La présence de chaque liaison H autour du chromophore au cours de la dynamique est montrée sur la figure IV.5. On voit que les interactions entre le chromophore et la protéine, identifiées dans la structure cristalline, sont présentes dans la dynamique à l'état fondamental à l'exception de la liaison H avec Gln94 qui est présente seulement 31% du temps. On remarque également que la liaison H entre Glu222 et le groupement hydroxyle du chromophore est stable. On observe aussi que les molécules de l'agrégat d'eau s'interchangent fréquemment tout le long de la simulation (Fig. IV.6).

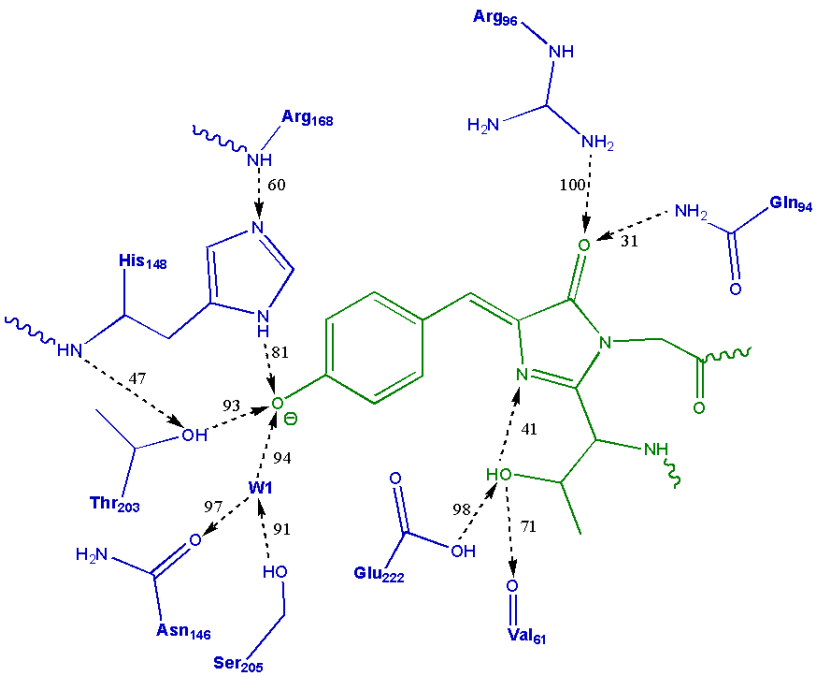


FIG. IV.5.: Le chromophore de la GFP (en vert) et une partie du réseau de liaisons hydrogène qui l'entoure. La présence des liaisons hydrogène au cours de la dynamique à l'état fondamental (indiquée en %) est définie par les critères suivants : la distance entre atomes lourds (X) doit être inférieure à 3,2 Å et l'angle X-H-X supérieur à 120°.

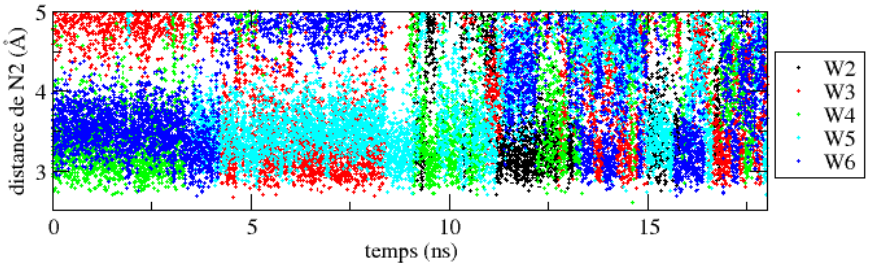


FIG. IV.6.: Les distances entre l'atome N2 et les cinq molécules d'eau de l'agrégat au cours de la dynamique à l'état fondamental. Les molécules d'eau sont nommées arbitrairement.

IV.1.3. Dynamique moléculaire à l'état excité

Nous avons effectué 50 simulations MD de 6 ns, donc un total de 300 ns, avec le chromophore à l'état excité (S_1). Les charges du chromophore, les paramètres d'équilibre (distance, angles, etc.) sont ceux de l'état excité et le potentiel de torsion de τ et φ est le potentiel NADiA. Les coordonnées et les vitesses intiales sont extraites de façon régulière, tout les 0,4 ns, de la simulation à l'état fondamental (S_0) décrite ci-dessus.

La figure IV.7 présente, superposées à la PES interne du chromophore, les géométries visitées au cours des 300 ns de simulation. A part la région de la géométrie plane (FS), le chromophore visite les régions TwP^-, TwP^+ et TwI^+ de façon très inégale et dissymétrique : le chromophore effectue un grand nombre de torsions vers la géométrie perpendiculaire TwP^- ($\varphi = -90^\circ$) et un petit nombre de torsions vers TwI^+ ($\tau = 90^\circ$) et TwP^+ ($\varphi = 90^\circ$) où seulement une partie de la région de TwI^+ est visitée. On n'observe aucune torsion vers TwI^- ($\tau = -90^\circ$) ou vers les maxima HT ($\tau = \pm 90^\circ$ et $\varphi = \pm 90^\circ$).

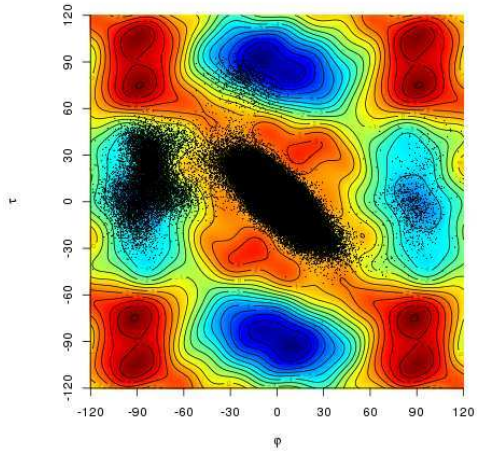


FIG. IV.7.: Corrélation entre l'angle τ et l'angle φ au cours des 300 ns de simulation superposé au potentiel de torsion interne du chromophore. Le chromophore visite majoritairement la région TwP^- et il effectue également un nombre minoritaire de torsions vers TwI^+ et TwP^+. A chaque torsion le chromophore quitte la région de la géométrie plane par une torsion concertée entre τ et φ avec une corrélation négative ($\tau > 0^\circ$ et $\varphi < 0^\circ$ ou bien $\tau < 0^\circ$ et $\varphi > 0^\circ$) conformément à la PES.

Le chromophore visite donc quatre régions : FS et trois des minima de géométries perpendiculaires. Il passe très peu de temps entre ces régions et on peut ainsi isoler les géométries correspondant à chacune entre elles. Ceci nous permet d'analyser certains paramètres de la dynamique dans la région FS, c'est à dire quand le chromophore n'a pas subi de torsion. Nous avons isolé les snapshots de cette région (définie par les critères $|\varphi| < 45^\circ$ et $|\tau| < 45^\circ$) et calculé les valeurs moyennes ainsi que les fluctuations de φ, τ, et des longueurs de liaison H du phénolate du chromophore avec His148 et Thr203. Les résultats sont montrés dans le tableau IV.1, comparé aux mêmes grandeurs dans la simulation à l'état fondamental.

TAB. IV.1.: Variation des angles φ et τ, et des longueurs de liaison H entre le phénolate du chromophore et His148 ou Thr203, dans la région de la géométrie plane (FS) du chromophore à l'état excité et à l'état fondamental.

	MD état excité	MD état fondamental
$\varphi(^\circ)$	2±13	0±7
$\tau(^\circ)$	1±12	0±7
longueur liaison H avec His148 (Å)	3,5±0,8	3,1±0,3
longueur liaison H avec Thr203 (Å)	3,2±0,8	2,9±0,4

On voit que, dans la région FS, le chromophore a des fluctuations de ses angles de torsion nettement plus grandes qu'à l'état fondamental, ce qui était prévisible d'après la forme de la PES. On voit également que les liaisons H avec His148 et Thr203 sont beaucoup plus faibles et fluctuantes qu'à l'état fondamental. Le chromophore à l'état excité explorant la large vallée de la région FS crée une déstabilisation de son environnement.

De plus, on observe qu'il y a des périodes de plusieurs nanosecondes où le chromophore reste dans la géométrie plane alors que le passage entre celle-ci et la géométrie perpendiculaire se fait en quelques picosecondes (Fig. IV.8), c'est-à-dire avec une dynamique analogue à celle observée dans le vide. Cela suggère que c'est l'environnement protéique du chromophore qui empêche la torsion mais qu'il y a certains mouvements spécifiques de la protéine (à l'échelle de la nanoseconde) qui modifie l'environnement et relâche les contraintes.

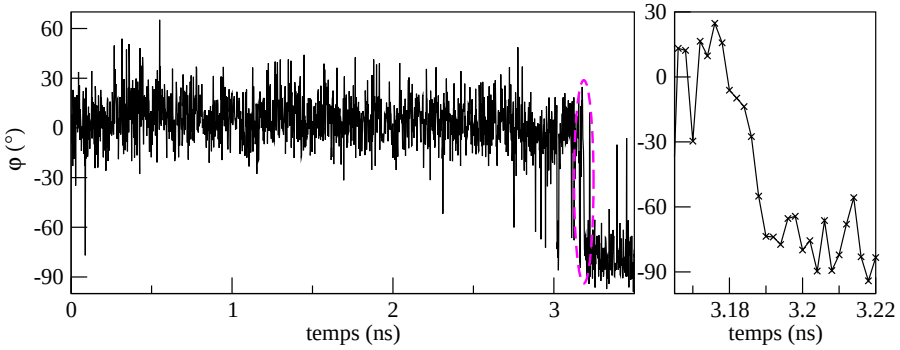


FIG. IV.8.: L'angle φ en fonction du temps au cours d'un simulation de la GFP avec le chromophore à l'état excité. La torsion à -90° entourée par une ellipse en magenta est montré en plus de détail sur le graphe à droite.

Nous avons dans un premier temps cherché à identifier ces mouvements dans le cas des torsions en φ.

IV.1.3.1. Analyse des torsions en φ

Les passages dans la région TwP^- (torsions en φ) sont très majoritairement associées à la perte de la liaison H entre le phénolate du chromophore et His148 (Fig. IV.9). L'examen des trajectoires montre que les rares cas de torsion en φ coexistant avec cette liaison H sont des torsions très courtes. On observe aussi que la torsion en φ est accompagnée par un léger changement de pyramidalisation des carbones *CA2* et *N3* (nomenclature sur Fig. IV.1) de l'imidazolinone et de l'angle dièdre *N1-CA1-C1-N2* (Fig. IV.10).

Cependant, on voit sur ces figures qu'aucun de ces évènements n'est une condition suffisante pour permettre un passage dans la région TwP^-.

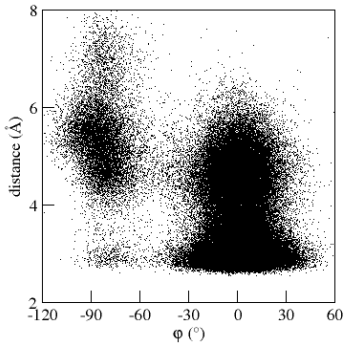


FIG. IV.9.: Corrélation entre l'angle dièdre φ et la longueur de liaison H entre le phénolate du chromophore et His148.

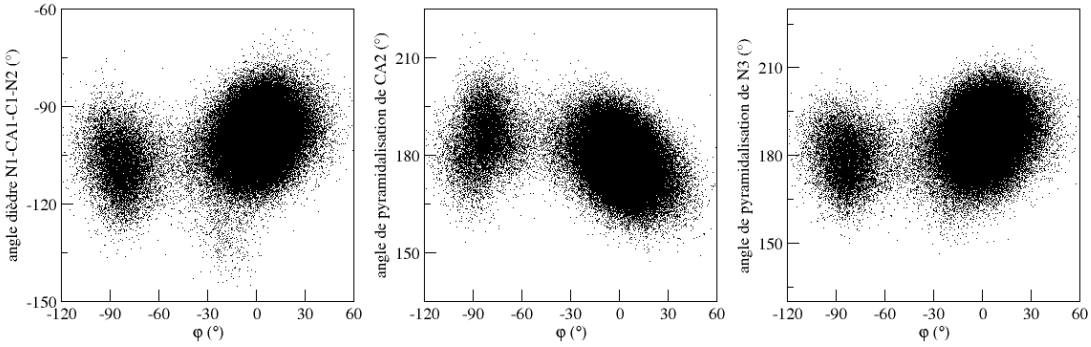


FIG. IV.10.: L'angle dièdre *N1-CA1-C1-N2* (à gauche), l'angle de pyramidalisation de l'atome *CA2* (au centre) et l'angle de pyramidalisation de l'atome *N3* (à droite) en fonction de l'angle φ.

Nous décrivons maintenant trois des simulations où une torsion vers TwP^- se produit. Nous les notons A, B et C.

Dans la simulation A, on observe une torsion durable en φ à t=3,2 ns (Fig. IV.11). Sur cette même figure on observe que la liaison H avec His148, très instable dès le début de la trajectoire, est presque continûment rompue à partir de t=1,3 ns, et celle avec la molécule d'eau W1 à partir de t=2,7 ns . Pendant l'intervalle de temps [2,7-3,2 ns] seules des torsions très brèves en φ ont lieu malgré l'absence de ces deux liaisons H. L'observation de la trajectoire par VMD [121] montre que l'agrégat de molécules d'eau joue le rôle déclenchant qui donne au chromophore l'espace nécessaire pour tourner de façon durable. Le chromophore reste ensuite en géométrie perpendiculaire pendant environ 2 ns.

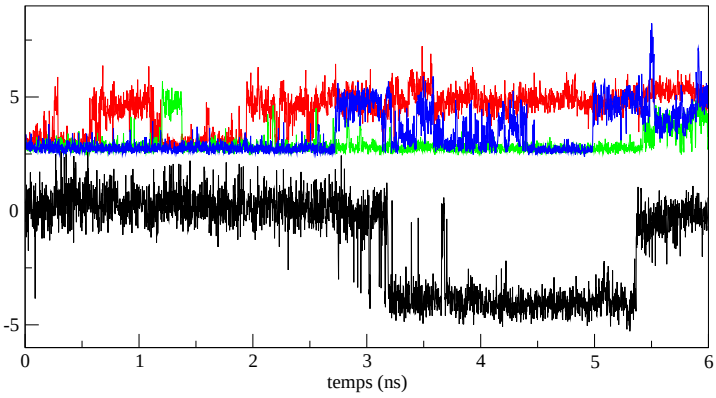


FIG. IV.11.: Simulation A : variation en fonction du temps de l'angle φ (divisé par 20, en degrés, couleur noir), longueurs de liaison H (Ångström) entre le chromophore et His148 (en rouge), W1 (en bleu) et Thr203 (en vert).

La figure IV.12 montre le chromophore, Thr203, His148, W1 et les 5 molécules de l'agrégat d'eau W2-W6 dans deux 'snapshots' consécutifs de la simulation (donc éloignés de 2 ps) : un premier juste avant la torsion et un deuxième juste après. Dans le premier snapshot, deux des molécules de l'agrégat (colorées en violet et jaune) se sont éloignées des autres. On remarque que la molécule d'eau violette a migré vers l'extrémité du phénolate et forme une liaison H avec son oxygène. Les trois molécules restantes ont donc plus de place et sont plus mobiles. La molécule d'eau (la verte) qui est au dessus de l'atome *CG2* du chromophore (nomenclature sur Fig. IV.1) est la plus proche du chromophore (distance de 4,4 Å ; sur la figure la molécule rouge semble plus proche mais en réalité elle est à l'arrière plan). Dans le deuxième snapshot, là où la torsion vient de commencer (φ=-58°), la molécule verte se trouve à une distance de 8,0 Å de *CG2*. La torsion en φ est accompagnée d'une faible torsion τ (+8°) et d'une pyramidalisation significative de l'atome *CA2* (-165°).

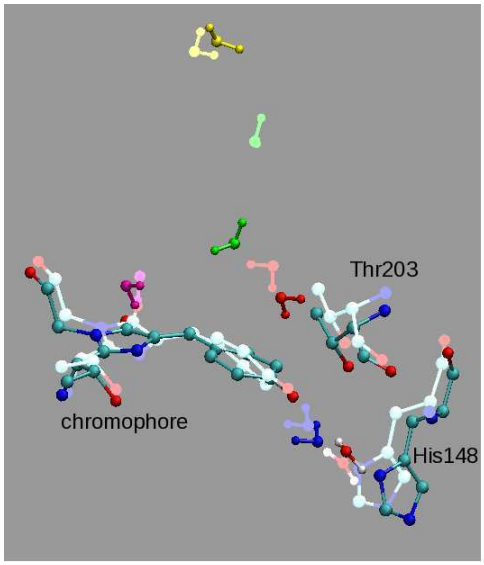


FIG. IV.12.: Simulation A : deux 'snapshots' consécutifs (donc éloignés de 2 ps), où le chromophore est plan (foncé) et perpendiculaire (clair) respectivement. Les cinq molécules de l'agrégat d'eau sont colorées en rouge, vert, violet, magenta et jaune. La molécule d'eau verte de l'agrégat, placée directement au-dessus du chromophore, s'est déplacée d'environ 4 Å entre les deux snapshots.

Dans la simulation B, on observe une torsion φ qui se produit sans qu'il y ait eu au préalable rupture de la liaison H avec His148 (Fig. IV.13). On observe cette rotation à t=4,9 ns et le chromophore reste ensuite environ 70 ps dans la géométrie perpendiculaire (ce passage est entouré par une ellipse en magenta). A l'échelle de la figure, il peut sembler que la torsion a lieu en même temps que les ruptures des liaisons H avec His148 et Thr203 ; mais un zoom sur ce passage montre que la torsion a lieu avant les ruptures de liaison.

L'observation de la trajectoire avec VMD montre que l'évènement précédant la torsion et donnant de la liberté au chromophore se passe au niveau de l'agrégat d'eau. Nous avons isolé 3 snapshots, numérotés 1, 2 et 3. Le premier est pris 16 ps avant la torsion, le deuxième 10 ps avant la torsion et le troisième juste après la torsion. Dans la figure IV.14 (à gauche), on compare les deux premiers snapshots. Dans le snapshot 1 (foncé) les molécules d'eau de l'agrégat représentées en vert et cyan sont proches du phénolate alors que dans le snapshot 2 (clair) elles se sont fortement éloignées et n'entravent plus la torsion du chromophore. Dans la figure IV.14 (à droite), on compare les snapshots 2 (clair) et 3 (foncé). Dans le snapshot 3, le chromophore a effectué une torsion jusqu'à φ =-67° et les molécules d'eau sont restées éloignées.

Cette torsion est accompagnée par une torsion τ (+28°), une pyramidalisation de *CA2* (-168°) et une pyramidalisation de *N3* (-171°). On voit également que les jonctions covalentes avec le reste de la protéine se sont légèrement déplacées. Au total, deux types de mouvements sont en jeu dans cette torsion : un mouvement de l'agrégat d'eau qui libère de la place et un mouvement de déformation du chromophore qui lui permet de se faufiler.

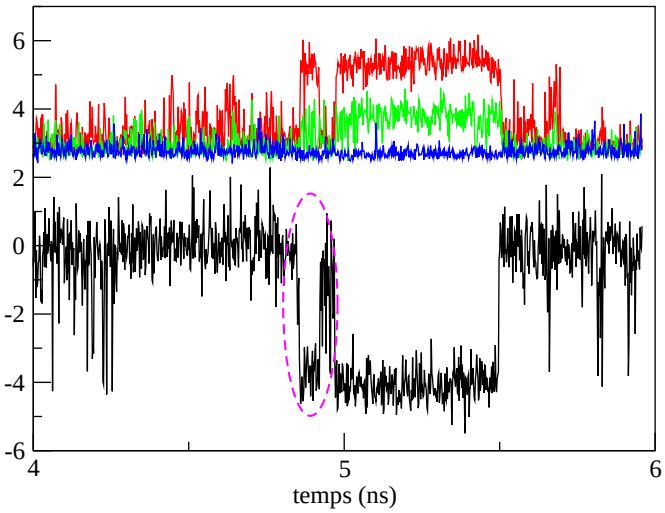


FIG. IV.13.: Simulation B : variation en fonction du temps de l'angle φ (divisé par 20, en degrés, couleur noir), longueurs de liaison H (Ångström) entre le chromophore et His148 (en rouge), W1 (en bleu) et Thr203 (en vert) dans les deux dernière nanosecondes de la simulations.

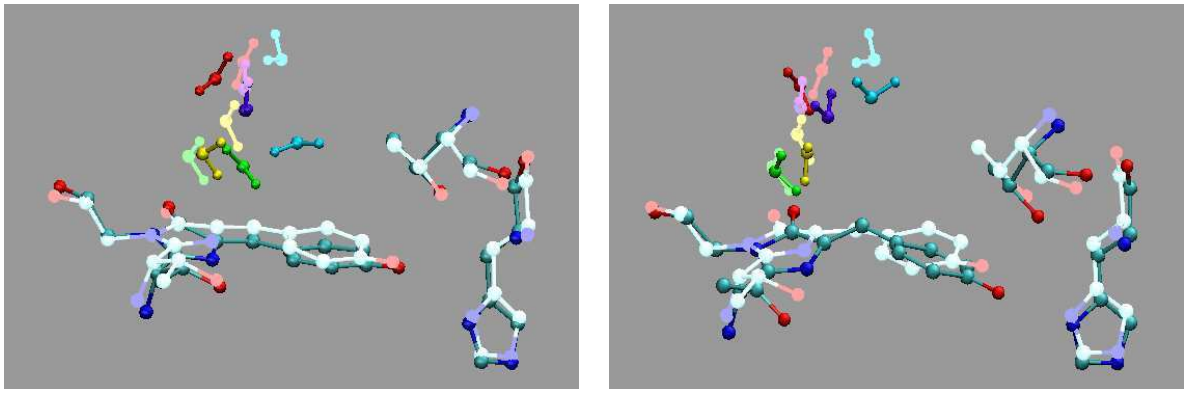


FIG. IV.14.: Simulation B : A gauche : deux 'snapshots' 16 ps (foncé) et 10 ps (clair) avant la torsion respectivement. On remarque que les molécules d'eau en vert et cyan de l'agrégat se sont éloignées du chromophore entre les deux snapshots. A droite : deux 'snapshots' 10 ps avant (clair) et juste après (foncé) la torsion respectivement. On remarque une petite torsion en τ et un déplacement du cycle imidazolinone, ainsi qu'une légère déformation des jonctions covalentes avec le reste de la protéine, entre les deux snapshots.

Dans la simulation C, on observe une torsion en φ qui se produit après une rupture préalable de la liaison H avec His148 (Fig. IV.15). On observe cette rotation à t=4,9 ns et le chromophore reste ensuite environ 30 ps dans la géométrie perpendiculaire (le passage est entouré par une ellipse en magenta). On voit sur la figure que la liaison H avec His148 est rompue ou instable depuis environ 0,1 ns avant la torsion, et que les liaisons H avec Thr203 et W248 sont majoritairement présentes. L'examen de la trajectoire avec VMD a mis en évidence que le déplacement du méthyle de Thr203 est le deuxième évènement, après la rupture de la liaison H avec His148, que permet la torsion.

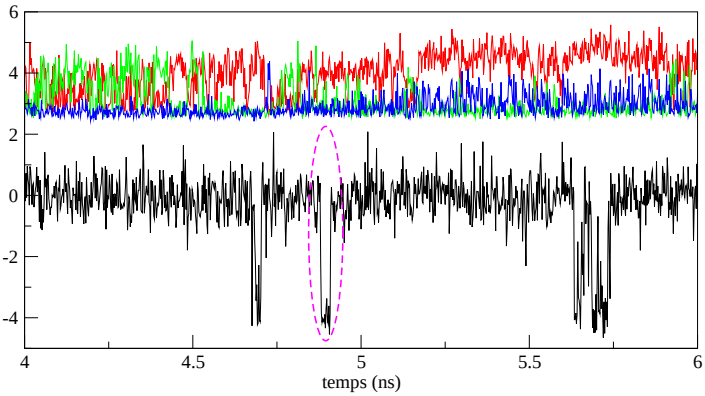


FIG. IV.15.: Simulation C : variation en fonction du temps de l'angle φ (en noir, divisé par 20, en degrés), longueurs de liaison H (Ångström) entre le chromophore et His148 (en rouge), W1 (en bleu) et Thr203 (en vert), dans les deux dernières nanosecondes de la simulation.

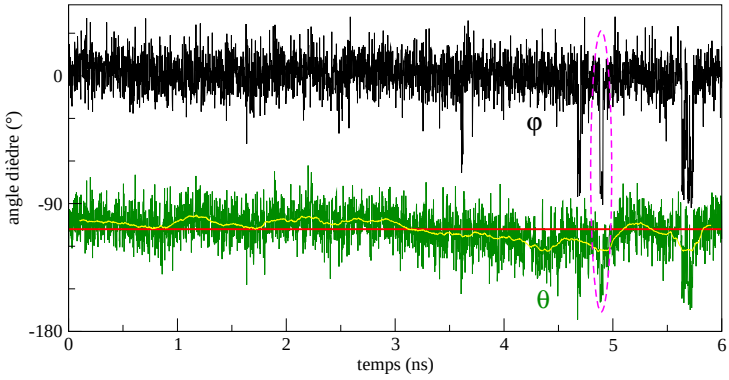


FIG. IV.16.: Simulation C : variation en fonction du temps de l'angle φ (en noir) et l'angle dièdre θ (*CA2-CB2-CG2*(chromophore)-*CG2*(Thr203)) (en vert). La moyenne de l'angle θ (en rouge) et la moyenne courante de θ (en jaune) sont également indiquées. La torsion du chromophore qui dure 30 ps (entourée par une ellipse en magenta) se produit quand l'angle θ est près de son minimum.

Nous avons caractérisé le déplacement du méthyle de Thr203 par un angle dièdre (*CA2-CB2-CG2*(chromophore)-*CG2*(Thr203)) que nous notons θ dans ce paragraphe. La valeur moyenne de θ le long des 6 ns de la trajectoire a une valeur moyenne de $-108^\circ \pm 14^\circ$ et prend des valeurs en moyenne significativement plus faibles dans les deux dernières ns (Fig. IV.16). Un zoom sur l'intervalle de temps de 0,1 ns autour de la torsion montre que les valeurs de θ sont de l'ordre de -130° dans les picosecondes précédent la torsion. Les images dans la figure IV.17 illustrent la correspondance entre les valeurs de θ et la position du méthyle de Thr203 par rapport au phénolate. Dans la première, où θ vaut -86°, le groupement méthyle est presque dans le plan médian du phénolate du chromophore (plan perpendiculaire au phénolate passant par *CG2-CZ*), alors que dans la deuxième, où θ vaut -128°, le groupement méthyle n'est plus au dessus du phénolate. Cette image correspond au snapshot précédant la torsion en φ. La troisième image contient une superposition du snapshot précédant la torsion (foncé) et du snapshot qui suit la

torsion (clair). Sur cette image, on voit que la position du groupement méthyle ne gêne pas la torsion du phénolate, à condition qu'elle se fasse dans le bon sens, c'est-à-dire vers TwP^- ($\varphi = -90^\circ$) et non vers TwP^+ ($\varphi = 90^\circ$). Le phénolate tourne (φ=-46°) en occupant l'espace laissé libre par le méthyle de Thr203. La torsion en φ est accompagnée d'une torsion τ=+20° et d'un léger déplacement de l'imidazolinone. L'examen des molécules de l'agrégat d'eau dans cette partie de la dynamique ne montre rien de particulier.

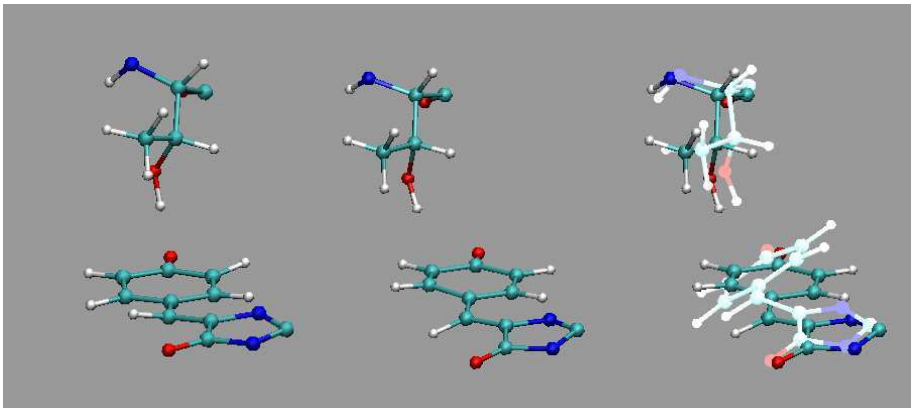


FIG. IV.17.: Simulation C : trois snapshots montrant les positions relatives du chromophore et du groupement méthyle de Thr203. Dans le premier, le méthyle est placé au-dessus du phénolate ; l'angle dièdre θ (*CA2*-*CB2*-*CG2*(chromophore)-*CG2*(Thr203)) vaut -86°. Dans le deuxième, le méthyle s'est déplacé (θ=-128°). Cette image est prise juste avant la torsion en φ. Le troisième snapshot est une superposition des géométries juste avant (foncé ; même géométrie que dans la deuxième image) et juste après (clair) la torsion en φ. La position du méthyle permet la torsion à condition qu'elle soit effectuée dans le bon sens (vers -90°). Dans cette image les deux géométries sont alignées sur les carbones α des feuillets β.

Au total, il ressort de la description de ces trois simulations, trois mouvements jouant un rôle déterminant dans la torsion en φ : la rupture de la liaison H entre le phénolate et His148, les mouvements des molécules de l'agrégat d'eau et le déplacement du groupement méthyle de Thr203.

La liaison H entre l'oxygène du phénolate et His148 ne s'oppose pas à une torsion φ pure puisque l'oxygène se trouve dans l'axe de rotation. En revanche, cette liaison H inhibe le mouvement du chromophore qui permet de sortir de la région FS puisqu'il implique une torsion corrélée entre φ et τ (Fig. IV.7) et donc un déplacement de l'oxygène du phénolate.

La rupture de la liaison H avec His148 est également observée, dans une moindre mesure, à l'état fondamental puisqu'elle n'existe que 81% du temps. Partant de l'hypothèse que les deux autres mouvements précédents sont également présents à l'état fondamental et amplifiés à l'état excité du fait de la déstabilisation de l'environnement du phénolate, nous les avons analysés dans la simulation à l'état fondamental. Dans cette simulation, le potentiel interne du chromophore maintient une géométrie plane et ne cherche pas à déformer l'environnement protéique par une torsion en φ ou τ. Elle devrait donc fournir une description des mouvements de la protéine qui se produisent à l'équilibre, et non des mouvements où peut intervenir la force motrice de torsion du chromophore.

IV.1.3.1.1. Retour sur l'analyse de la simulation à l'état fondamental .

Dans la simulation à l'état fondamental, la liaison H entre le phénolate du chromophore et His148 est nettement moins stable (81% d'existence) que les liaisons H avec Thr203 (93%) ou avec la molécule d'eau W1 (91%).

His148 est placé sur le brin β_{11}, et son cycle imidazole forme deux liaison H : l'une entre l'azote *ND1* et le chromophore et l'autre entre l'azote *NE2* et l'azote peptidique de Arg168, situé sur le brin β_1 (Fig. IV.18). Lors de la rupture de la liaison H entre le chromophore et His148, on observe que le brin β_{11}, portant His148, s'écarte des deux brins qui l'entourent, le brin β_{10} portant Thr203 et le brin β_1 portant Arg168 (Fig. IV.19). On observe aussi une augmentation systématique et de faible amplitude (au maximum égale à 0,3 Å) de la distance entre le brin β_{11} et l'hélice α centrale. Ce mouvement a été mis en évidence en étudiant les variations des distances entre les carbones α de Ser147, His148, Asn149 et les carbones α de Thr63, Phe64, Val68 et Gln69. Ces variations peuvent provenir d'un mouvement isolé du brin β_{11} s'écartant de l'hélice α et des brins β_1 et β_{10} ou d'un mouvement corrélé où les brins β_1, β_{10} et β_{11} s'écartent de l'hélice α par une respiration du tonneau.

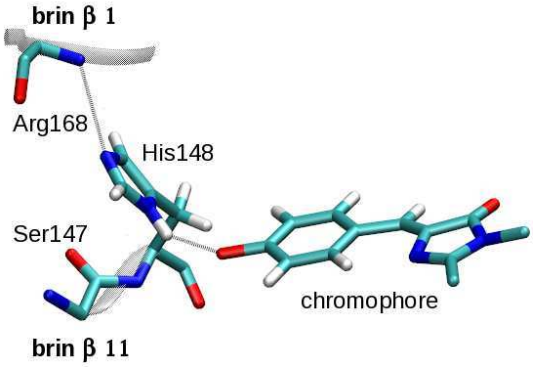


FIG. IV.18.: Réseau de liaisons H entre le chromophore, His148, Arg168 et Ser147, dans la structure cristalline de la GFP. L'oxygène du phénolate du chromophore est en liaison H avec l'azote *ND1* de His148. L'azote *NE2* de His148 est en liaison H avec l'azote peptidique de Arg168. Il n'y a pas de liaison H peptidique entre Arg168 et Ser147 (distance de $\sim 7,0$ Å entre les atomes lourds) puisque les deux résidus se trouvent dans la crevasse du tonneau β de la protéine (entre brin β_1 et β_{11}, voir Fig. II.2).

La figure IV.20 montre que la longueur de la liaison H entre His148 et Arg168 fluctue nettement moins que celle de la liaison H entre His148 et le chromophore, ce qui signifie que lorsque le brin β_{11} s'écarte, His148 a plutôt tendance à garder la liaison avec Arg168 et perdre celle avec le chromophore que l'inverse. La figure. IV.21 montre les déplacements de brins β lorsqu'on passe d'un snapshot comportant une liaison H entre le chromophore et His148 à un snapshot sans cette liaison H. La déformation des brins β et la réorientation de la chaîne latérale de His148 permettent de renforcer la liaison H avec Arg168. Cette réorientation est obtenu par une torsion autour de la chaîne latérale de His148, un mouvement qui se produit systématiquement quand His148 perd sa liaison H avec le chromophore (Fig. IV.22).

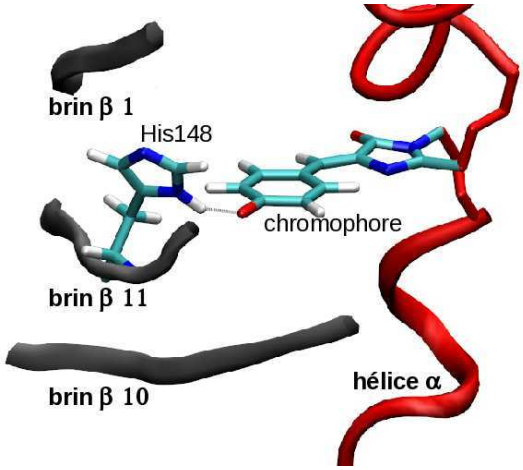


FIG. IV.19.: Schéma du chromophore situé sur l'hélice α centrale (en rouge) et les brins β (en gris) numéro 1 (en haut), 11 (au milieu) portant His148, et 10 (en bas).

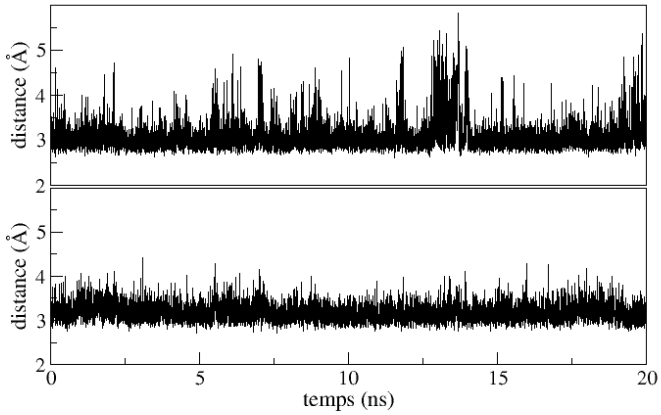


FIG. IV.20.: La longueur de liaison H entre l'azote *ND1* de His148 et l'oxygène *OH* du phénolate du chromophore (en haut), la distance entre l'azote *NE2* de His148 et l'azote *N* de Arg168 (en bas), en fonction du temps.

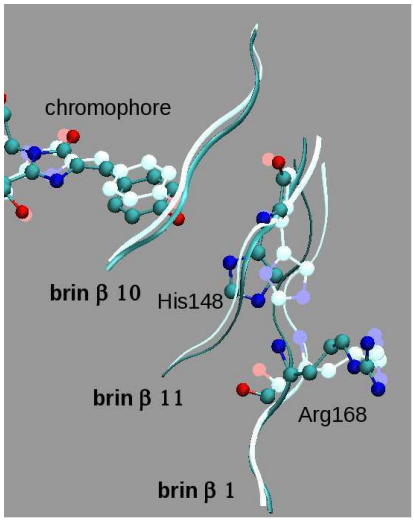


FIG. IV.21.: Le chromophore, His148 et Arg68. En foncé : un snapshot où la liaison H entre le chromophore et His148 existe. En clair : un snapshot où la liaison H est rompue. Les deux snapshots sont alignés sur les carbones α des feuillets β.

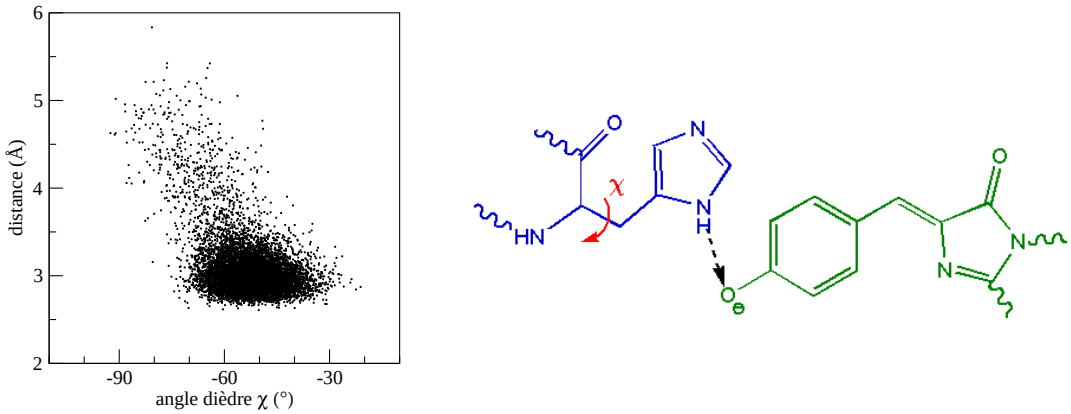


FIG. IV.22.: Corrélation entre la longueur de liaison H entre l'azote *ND1* de His148 et l'oxygène du phénolate du chromophore en fonction de l'angle dièdre χ (à gauche). L'angle χ correspond à la torsion autour de la liaison *CA-CB* de His148 (à droite). La rupture de la liaison H est associée à une diminution significative (13°) de l'angle χ.

La rupture de la liaison H entre le phénolate et His148 est donc lié à un déplacements des brins β en face du chromophore.

En ce qui concerne la position du groupement méthyle de Thr203 par rapport au plan du phénolate, nous avons tracé l'histogramme de l'angle dièdre θ (*CA2-CB2-CG2*(chromophore)-*CG2*(Thr203)) au cours de la simulation à l'état fondamental (Fig. IV.23). Cet angle présente de larges fluctuations. Une valeur de $\theta \lesssim -120°$ correspond à une position où le méthyle n'est plus au-dessus du plan du phénolate et où il ne gêne pas la rotation du cycle phénolate. Cette position est visitée environ 10% du temps dans la simulation à l'état fondamental.

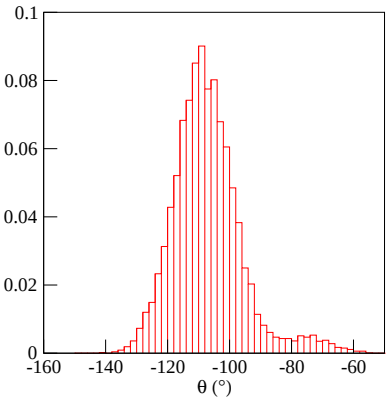


FIG. IV.23.: Distribution de l'angle dièdre θ (*CA2*-*CB2*-*CG2*(chromophore)-*CG2*(Thr203)) caractérisant la position du groupement méthyle de Thr203 par rapport au plan du phénolate, au cours de la simulation à l'état fondamental.

En ce qui concerne les mouvements des molécules de l'agrégat d'eau, on a déjà vu qu'elles s'échangent régulièrement au cours de la dynamique de l'état fondamental. Le figure IV.24 montre cet agrégat dans la structure cristalline, et à différents temps de la simulation à l'état fondamental. On remarque que l'agrégat devient nettement moins compact au bout de 12 ns.

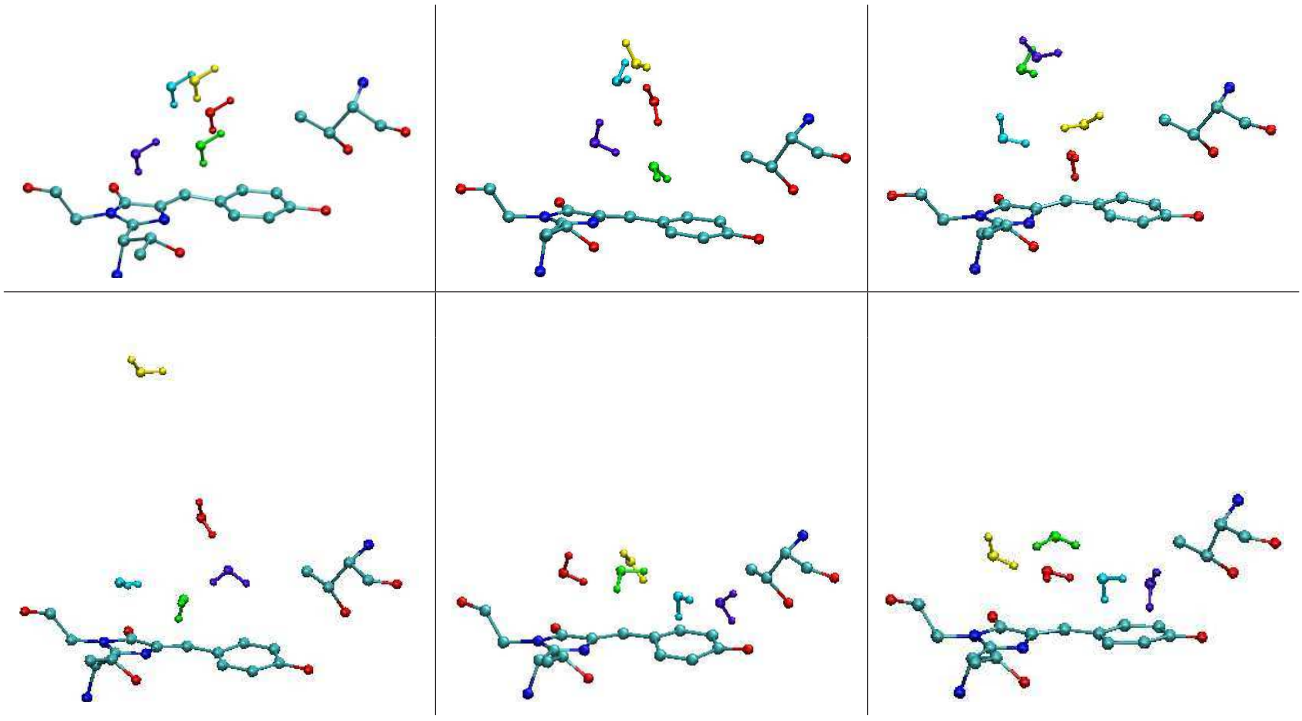


FIG. IV.24.: L'agrégat de molécules d'eau (en rouge, vert, violet, cyan et jaune) dans la structure cristalline (hydrogènes non-relaxés) et après 4 ns, 8 ns, 12 ns, 16 ns et 20 ns de simulation avec le chromophore à l'état fondamental, ainsi que le chromophore et Thr203. Dans les deux dernières structures, la molécule violette se trouve à l'avant de la figure (et non au dessus du phénolate).

IV.1.3.1.2. Conclusion sur les torsions en φ .

Différents mécanismes peuvent libérer la torsion en φ. Ils comprennent tous plusieurs mouvements spécifiques qui doivent coïncider. Une torsion est précédée par différents évènements qui se produisent dans différents ordres. Les mouvements spécifiques que nous avons identifiés sont la rupture de la liaison H entre le phénolate du chromophore et His148, le balancement du groupement méthyle de Thr203 et la levée de l'encombrement stérique dû à l'agrégat d'eau. Dans le cas où la liaison H avec His148 est maintenue, on peut se demander comment le chromophore arrive à quitter la région FS. On décrira un tel processus plus tard dans le cas des torsions en τ dans la GFP et dans la YFP.

Les mouvements spécifiques de la protéine liés à la torsion en φ sont présents dans la simulation à l'état fondamental et amplifiés à l'état excité par les forces internes du chromophore qui tendent à le faire tourner et déstabilisent son environnement. A partir du moment où l'un de ces évènements se produit, la torsion du chromophore est facilitée. S'il atteint la sortie de la région FS quand une ou plusieurs contraintes sont levées, ses forces internes peuvent être suffisamment fortes pour surmonter les contraintes restantes. La déformation du cycle imidazolinone l'aide à trouver son chemin.

IV.1.3.2. Analyse des torsions en τ

On voit sur la figure IV.25 (à gauche) que lors de chaque passage dans la région TwI^+ (torsions en τ), la liaison H entre le chromophore et Thr203 est rompue : les points de la figure pour lesquels $\tau \approx 90^\circ$ correspondent tous à une distance supérieure à 4 Å. La rupture de cette liaison est donc une condition nécessaire pour la torsion en τ. Par contre, ce n'est pas une condition suffisante puisqu'il y a de nombreux points à $\tau \approx 0^\circ$ avec une distance supérieure à 3,5 Å.

En ce qui concerne la liaison H avec His148, on voit sur la figure. IV.25 (à droite) que sa rupture n'est pas directement liée à la torsion en τ puisque les points à $\tau \approx 90^\circ$ correspondent à des valeurs de la distance qui s'étendent de 2,5 à 4 Å.

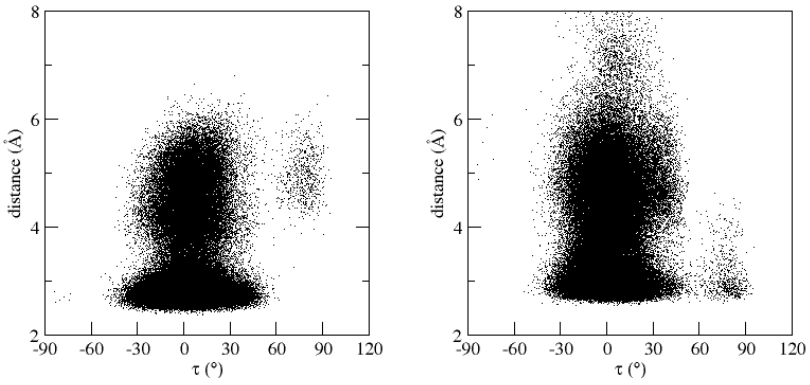


FIG. IV.25.: Corrélation entre l'angle τ et la distance entre l'oxygène du phénolate et Thr203 (à gauche) ou His148 (à droite).

Si on se limite aux cinq simulations où le passage dans la région de TwI^+ dure au moins 20 ps, il apparaît qu'il y a toujours un délai entre la rupture de la liaison H avec Thr203 et la torsion

en τ ce qui est une autre manifestation du fait que la rupture de cette liaison H n'est pas une condition suffisante.

On observe aussi que la torsion en τ est accompagnée par des mouvements de déformation du chromophore : une légère torsion en φ (corrélation négative), un changement de pyramidalisation de l'azote *N3* de l'imidazolinone et une modification de l'angle dièdre *N1-CA1-C1-N2* (nomenclature sur Fig. IV.1).

Nous donnons une description plus détaillée dans le cas d'une des cinq simulations où il y a un délai de 400 ps entre la rupture de la liaison H avec Thr203 et la torsion en τ (Fig. IV.26). On observe que suite à la rupture, le groupement hydroxyle de Thr203 s'éloigne de l'oxygène du phénolate du chromophore (distance 5 Å), par une rotation de la chaîne latérale, et se rapproche d'une des molécules de l'agrégat d'eau. La rotation du cycle phénolate est donc bloquée par la chaîne latérale entière de Thr203, et non seulement par le méthyle. La torsion en τ est accompagnée d'une variation de 20° de l'angle de pyramidalisation de l'azote *N3* du cycle imidazolinone : l'angle change de 190° (valeur moyenne dans l'état S_0) à 170° (Fig. IV.27). L'angle dièdre *N1-CA1-C1-N2* varie également de 20°. En revanche, ni l'angle dièdre *C3-CA3-N3-C2*, ni l'angle de pyramidalisation du carbone *C1*, ne varient.

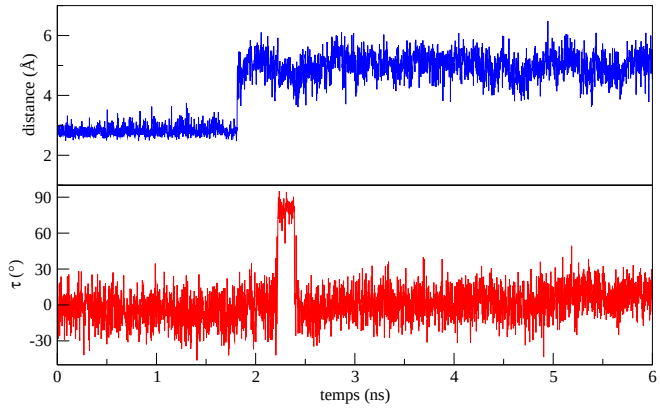


FIG. IV.26.: Une simulation dans laquelle le chromophore effectue une torsion en τ, l'amenant vers la région TwI^+, après 2,2 ns (en rouge). A 1,8 ns, la liaison H entre le phénolate et Thr203 (en bleu) est rompue.

La torsion en τ déplace le phénolate vers Thr62, i.e. dans le sens opposé de Thr203, et atteint une région où $\tau = 80 \pm 7°$ et $\varphi = -19° \pm 9°$. La chaîne latérale de Thr62 s'écarte en libérant de la place pour la géométrie perpendiculaire du chromophore.

Les liaisons H entre le phénolate et His148 et la molécule d'eau sont maintenues pendant toute la durée du passage dans la région de TwI^+.

De la même façon que pour les torsions en φ, nous avons cherché les circonstances de la rupture de la liaison H avec Thr203 dans la simulation à l'état fondamental.

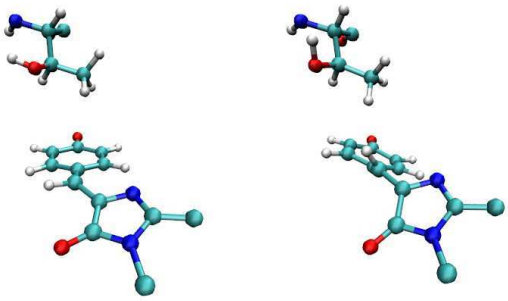


FIG. IV.27.: Deux snapshots de la simulation de la GFP à l'état excité, juste avant et juste après la torsion en τ. La liaison H entre Thr203 et le chromophore est rompue et la chaîne latérale entière de Thr203 est placée au-dessus du phénolate. Dans le premier snapshot, le chromophore est à la sortie de la région FS vers la région TwI^+ ($\tau = 40^\circ$, $\varphi = -21^\circ$) et dans le deuxième il a effectué une torsion en τ ($\tau = 73^\circ$, $\varphi = -35^\circ$).

IV.1.3.2.1. Retour sur l'analyse de la simulation à l'état fondamental .

La liaison H entre l'oxygène du phénolate et l'oxygène du groupement hydroxyle de Thr203 est presque permanente à l'état fondamental (93% d'existence). La distance moyenne entre atomes lourds vaut $2,85 \pm 0,25$ Å. La rupture de la liaison n'est observée que rarement pendant une durée maximale de 200 ps (Fig. IV.28, en haut). Dans le cas de la rupture la plus longue, le groupement hydroxyle de Thr203 se déplace par une rotation de la chaîne latérale pour occuper l'espace le long du cycle phénolate du chromophore. Le groupement hydroxyle de Thr203 perd ainsi ses deux liaisons H avec le phénolate du chromophore et l'azote peptidique N de His148 (Fig. IV.29). Simultanément, la liaison H entre l'oxygène peptidique de Thr203 et l'azote peptidique de His148 se renforce très nettement (Fig. IV.28, en bas). La chaîne latérale de Thr203 est partiellement stabilisée dans sa nouvelle position, par une liaison H entre l'oxygène hydroxyle et une molécule de l'agrégat d'eau, et probablement par des interactions de type van der Waals entre le groupe méthyle et le noyau aromatique du phénolate.

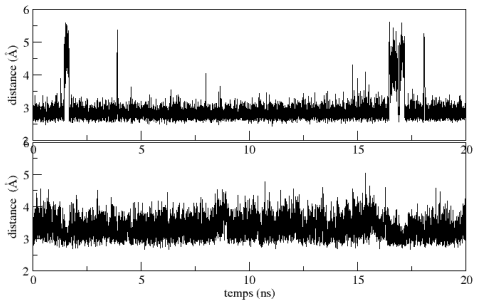


FIG. IV.28.: La longueur de liaison H entre l'oxygène de Thr203 et l'oxygène du phénolate du chromophore (en haut) et entre l'oxygène peptidique de Thr203 et l'azote peptidique de His148 (en bas), en fonction du temps.

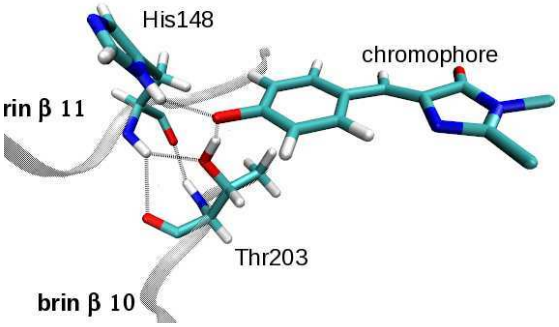


FIG. IV.29.: Réseau de liaison H entre le chromophore, Thr203 et His148, dans la structure cristalline de la GFP. L'oxygène du phénolate du chromophore est en liaison H avec l'oxygène du groupment hydroxyle de Thr203 et avec l'azote *ND1* (nomenclature dans Fig. C.1) de His148. Thr203 et His148 sont liés par deux liaisons H peptidiques (entre l'oxygène *O* peptidique et l'azote *N* peptidique). L'azote *N* de His148 est également lié à l'oxygène hydroxyle de Thr203 par une liaison H.

IV.1.3.2.2. Conclusion sur les torsions en τ .

Les torsions en τ sont très minoritaires. Elles sont toutes précédées par la rupture de la liaison H entre le phénolate et Thr203. Cette rupture est accompagnée par une rotation de la chaîne latérale de Thr203 qui vient bloquer fortement la rotation du cycle phénolate et donc la torsion en φ. La torsion en τ est aussi contrainte par la chaîne latérale de Thr203 dans un sens et par celle de Thr62 dans l'autre. Les mouvements de déformation du chromophore, décrits ci-dessus, permettent une torsion en τ qui permet d'éviter ces obstacles et aussi de garder les liaisons H du phénolate avec His148 et W1. La perte de la liaison H avec Thr203 enlève une des trois points d'attache de l'oxygène du phénolate ce qui peut rendre plus facile les mouvements qui conduisent le chromophore jusqu'à la sortie de la région FS.

IV.1.4. Étude de mouvements collectifs de la protéine

Nous avons effectué une analyse en modes normaux (NMA) de vibration de la GFP dans le but de déterminer si un ou plusieurs mouvements à basse fréquence, i.e. des mouvements collectifs de la protéine, peuvent être liés à un mécanisme de 'libération' de la torsion du chromophore. Cette approche s'inspire de ce qui a été fait par exemple dans le cas de l'hémoglobine [122, 123]. On note en particuliers que les FPs ont un seul domaine au lieu de plusieurs dans les globines et quelle sont donc *a priori* plus rigide, en particulier le tonneau entourant le chromophore. On peut donc s'attendre à ce qu'il n'y ait que des mouvements de petite amplitude ce qui justifie d'utiliser la notion de mode normal. Les calculs ont été effectués à l'aide de la suite CHARMM [124]. Les détails sont décrits dans l'annexe D.

Pour identifier les modes normaux significatifs, nous avons d'abord cherché les déplacements moyens des atomes de la protéine liés à la torsion de φ. Pour cela, un vecteur de déplacement δr de dimension $3N$ est défini, où N est le nombre d'atomes de la protéine hors le chromophore et les molécules d'eau. On identifie ensuite les 'épisodes de torsion' c'est-à-dire des paires de 'snapshots' de la simulation MD qui encadrent une torsion en φ (un premier snapshot ou le

chromophore est plan et un deuxième où il est perpendiculaire). Pour chaque snapshot, les coordonnées de tous les atomes sont exprimées dans les axes d'inertie de la protéine. On pose alors :

$$\delta r_{i,k} = \mathcal{N} \left[r_{i,k}^{(post)} - r_{i,k}^{(pre)} \right] \tag{IV.1}$$

où $\mathcal{N}$ est un facteur de normalisation, $r_{i,k}$ est une coordonnée pondérée en masse ($i = x$, y ou z et k = numéro d'un atome - hors chromophore et molécules d'eau), $r_{i,k}^{(pre)}$ est la valeur de la coordonnée $r_{i,k}$ dans le premier snapshot de l'épisode de torsion en φ et $r_{i,k}^{(post)}$ dans le second. On évalue alors la moyenne $\overline{\delta r}_{i,k}$ des composants $\delta r_{i,k}$ sur les épisodes de torsion (soit sur 220 paires de snapshots) : le vecteur $\overline{\delta r}$ décrit une déformation de la protéine caractéristique de ce qui se passe pendant la torsion φ.

Nous pouvons alors comparer cette déformation aux mouvement collectifs de la protéine en évaluant les produits scalaires de $\overline{\delta r}$ avec les modes normaux (également exprimées dans les axes d'inertie de la protéine). Les résultats obtenus pour les 100 modes normaux qui correspondent aux mouvements les plus lents de la protéine sont présentés sur la figure IV.30.

On voit qu'il n'y a pas de mode normal qui donne une contribution dominante à $\overline{\delta r}$. Certains modes normaux donnent des contributions qui paraissent plus significatives (14, 15, 28 ...) mais pas au point que l'on puisse négliger les autres.

On note en particulier qu'il ne semble pas y avoir de saturation de la norme de la projection de $\overline{\delta r}$ quand on augmente le nombre de modes normaux : même des modes des fréquence relativement élevée donnent une contribution non négligeable. Cela peut être dû au fait que le mouvement collectif cherché ici n'est pas de très grande amplitude, contrairement par exemple à ce qui se passe dans l'hémoglobine [122, 123]. Il est donc couplé avec beaucoup d'autres modes normaux et n'apparaît pas comme réduit à un petit nombre de modes.

On ne peut donc pas ici représenter le déplacement moyen lié à la torsion de φ par un ou quelques modes normaux.

Une autre façon d'utiliser les modes normaux consiste à remarquer que, dans le vide, le chromophore effectue une torsion en moyenne au bout de 2,6 ps. Ce délai s'explique par la forme particulière du paysage d'énergie autour de la géométrie plane (FS) plutôt que par la hauteur des barrières d'énergie très faibles le long des chemins de torsion : les sorties étroites de la FS forment un goulet d'étranglement et créent une barrière entropique (*entropic bottle neck*). Les modes normaux capables de libérer la torsion doivent avoir une période T supérieure à cette limite, T>2,6 ps, soit une fréquence limite $\omega \lesssim 11\,cm^{-1}$: certains modes plus rapides peuvent éventuellement ouvrir la voie à la torsion mais, la période étant trop courte, ils la referment avant que le chromophore ait eu le temps de quitter la région FS.

Cette fréquence limite obtenue est indiquée sur la figure IV.31. On voit qu'en prenant les modes normaux, de 7 ($\omega = 5,7\,cm^{-1}$) à 25 ($\omega = 11,6\,cm^{-1}$) (les modes normaux 1-6 représentent les translations et rotations d'ensemble) on couvre raisonnablement le domaine des fréquences appropriées. Mais on voit sur la figure IV.30 que ces modes ne représentent qu'environ 25% du déplacement moyen nécessaire $\overline{\delta r}$ (et deux entre eux, numéro 14 et 15, environ 12%).

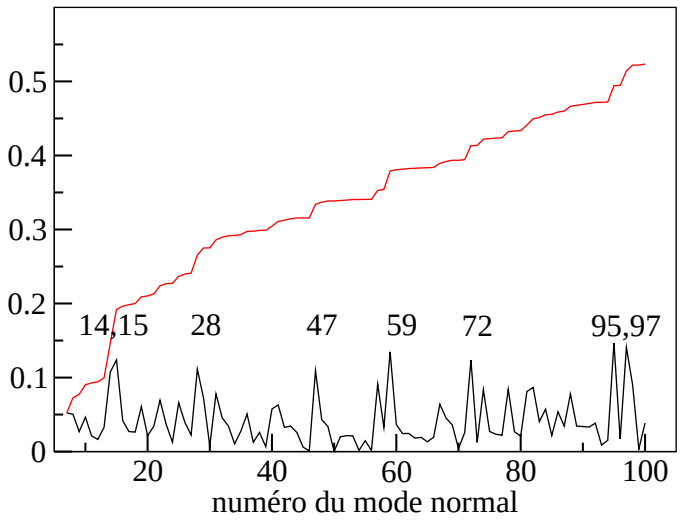


FIG. IV.30.: Décomposition en mode normaux du déplacement moyen $\overline{\delta r}$ de tous les atomes (hors le chromophore et les molécules d'eau) qui permettent la torsion de φ. Le produit scalaire de $\overline{\delta r}$ avec chaque mode normal est affiché en noir et la fonction $\mathcal{N}(k)$ (racine carrée de la norme somme des carrés des produits scalaires sur les modes normaux de 7 à k) est affichée en rouge. Les nombres au dessus des courbes noires donnent les numéros des modes ayant un grand produit scalaire.

On arrive ainsi à une description de la dynamique de torsion plus complète que ce qui apparaît simplement, par exemple, à partir des variations de l'angle φ présenté sur la figure IV.8 : il n'existe pas en réalité de mouvement de la protéine que l'on puisse décrire à l'aide d'un petit nombre (un, deux, trois ...) de modes normaux capables de libérer la torsion pendant un temps suffisant. La protéine peut libérer effectivement le chromophore mais seulement en partie : il y a des périodes de temps où la torsion est possible mais une partie seulement de la barrières entropique est ouverte (l'ouverture complète requiert que la projection de $\overline{\delta r}$ ait une norme de 1). Pour passer quand même, le chromophore doit se déformer comme vu pour la GFP et la YFP.

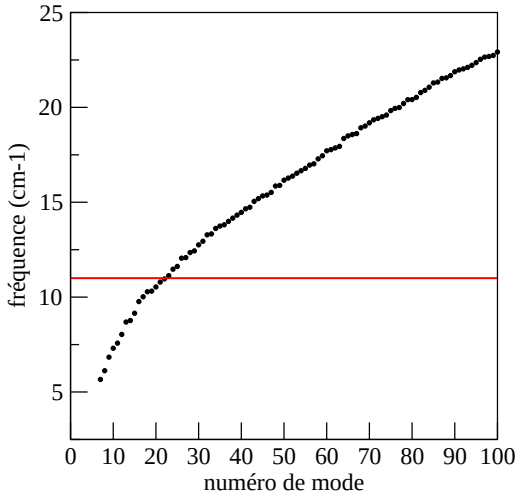


FIG. IV.31.: La fréquence ω des 100 modes normaux de vibration de la GFP.

On arrive finalement à une description où l'extinction de fluorescence résulte de deux facteurs en grande partie indépendantes.

- des mouvements de la protéine (y compris les molécules d'eau internes) qui libèrent en partie un chemin pour la torsion

- une déformation du chromophore qui lui permet de passer dans l'espace libéré

Le chemin étant libéré seulement en partie, le temps requis par le chromophore pour tourner est plus long que 2,6 ps. Comme les mouvements très lents de la protéine n'ont qu'une toute petite composante sur $\overline{\delta r}$ on doit plutôt utiliser la notion de coïncidence : le chemin de torsion se libère juste au moment où le chromophore est près d'une sortie de la région FS. Il n'est alors nécessaire que ce soit pendant un temps très long. Mais la coïncidence se produit rarement.

IV.1.5. Conclusion

Les mouvements de torsion du chromophore dans la GFP sont majoritairement des torsions en φ. Afin de permettre ces torsions, plusieurs mouvements clés, peu coûteux en énergie, doivent coïncider pendant un temps suffisant (quelques picosecondes). Le mécanisme qui retarde la torsion du chromophore (de l'échelle de la picoseconde dans le vide à l'échelle de la nanoseconde dans la GFP) ne vient pas d'une grande barrière énergétique mais d'une coïncidence entre des mouvements clés qui se produit rarement. Ces événements clés se produisent fréquemment dans la protéine et sont présentent dans la dynamique de la protéine à l'état fondamental du chromophore. Ils sont amplifiés quand le chromophore est à l'état excité, du fait des forces internes qui cherchent à effectuer des torsions du chromophore.

IV.2. Analyse de la YFP

La protéine YFP ne diffère de la GFP que par quatre mutations : T203Y, T65G, V68L et S72A. On peut s'attendre à ce que l'interaction de π-stacking entre le phénolate du chromophore et le cycle phénol de Tyr203 provoque un blocage important des torsions. Le remplacement de Thr65 par une glycine produit des modifications dans l'environnement de l'imidazolinone susceptibles d'apporter également des modifications aux mouvements de ce cycle.

IV.2.1. L'environnement du chromophore

Le chromophore interagit avec la protéine, d'une part au travers des deux liaisons covalentes faisant la jonction avec la protéine, et d'autre part, *via* plusieurs liaisons H (Fig. IV.32). L'oxygène *O2* de l'imidazolinone (nomenclature dans Fig. IV.33) est lié par deux liaisons H avec Arg96 et Gln69, l'azote *N2* est lié à Glu222. L'oxygène du phénolate forme deux liaisons H, l'une avec His148 et l'autre avec une molécule d'eau (W1), qui n'est pas en contact avec le solvant.

Les mouvements du chromophore sont également contraints par l'encombrement stérique des chaînes latérales de Tyr203 et Thr62, situées de part et d'autre du phénolate (Fig. IV.34). Par ailleurs l'interaction entre Tyr203 et le phénolate peut être stabilisante sous l'effet du π-stacking.

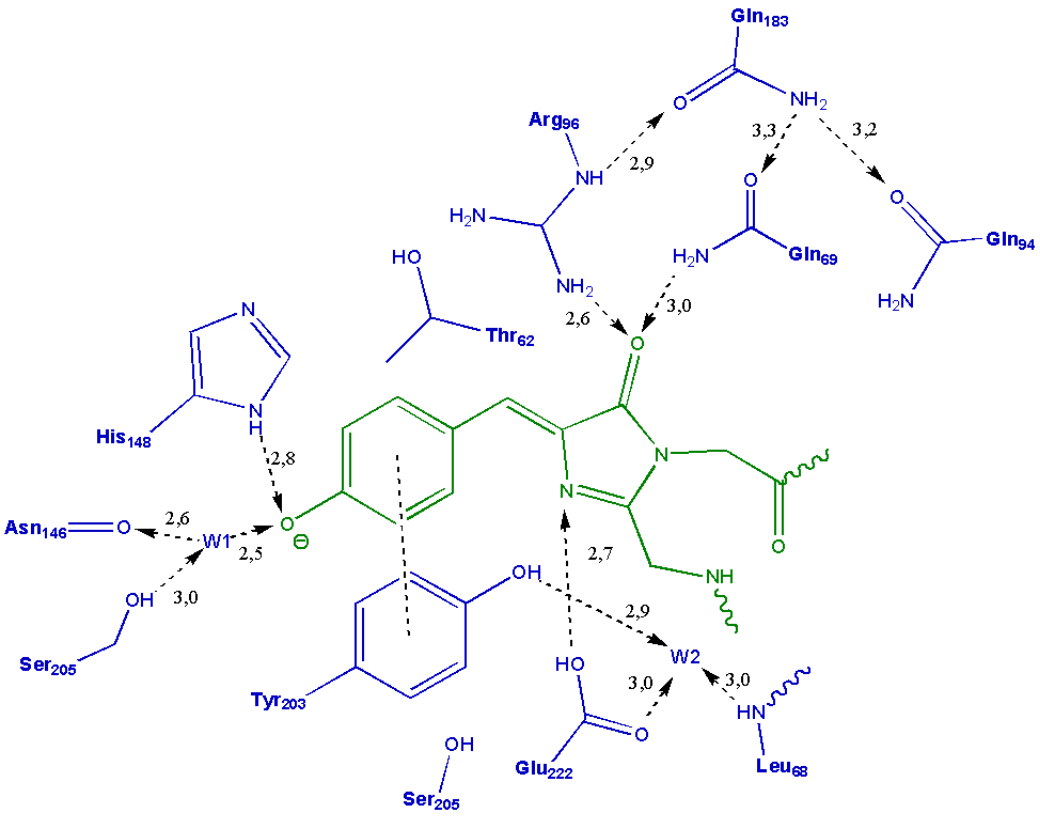


FIG. IV.32.: Le chromophore de la YFP (en vert) et une partie du réseau de liaisons hydrogène dans son environnement. Les distances inter-hétéroatomes sont tiré de la structure cristalline [58] et données en Ångström.

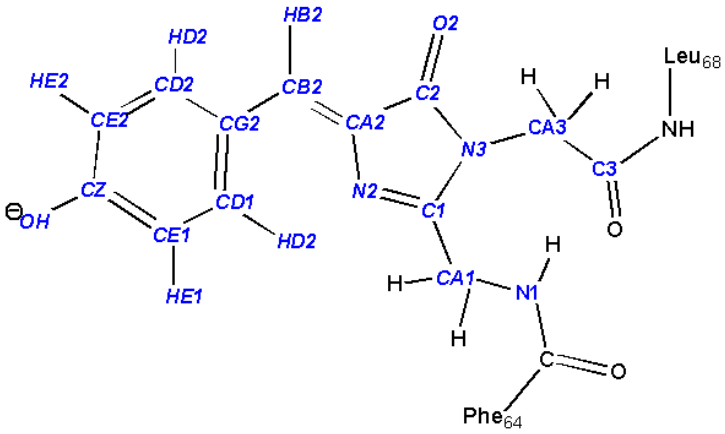


FIG. IV.33.: Le chromophore de la YFP (formé par le tripeptide Gly65-Tyr66-Gly67) avec les noms des atomes indiqués en bleu. Le groupement carboxyle de Phe64 et le groupement NH de Leu68 sont également montrés sur le schéma.

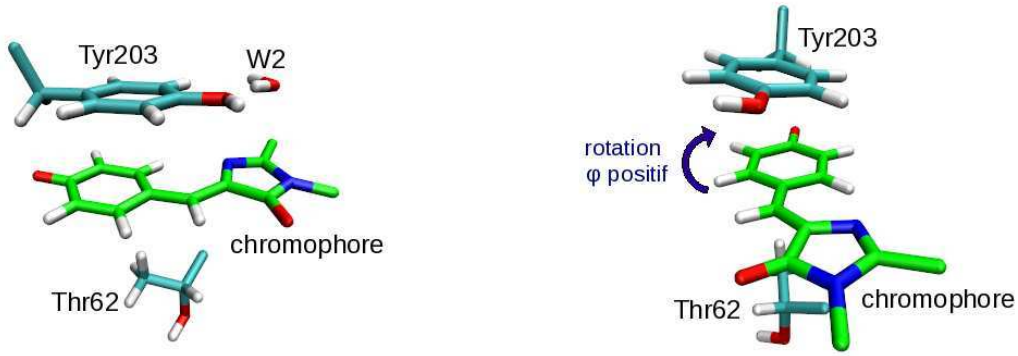


FIG. IV.34.: L'encombrement stérique dans le voisinage immédiat du chromophore. Le sens de la rotation en φ est indiqué sur la figure à droite.

IV.2.2. Dynamique moléculaire à l'état fondamental

Nous avons effectué une série de simulations de dynamiques moléculaires de la YFP avec le chromophore dans sa forme anionique. Les coordonnées de la structure cristalline [58] (nom PDB : 1YFP) ont été utilisées comme point de départ. Le pH de cristallisation (7,5) étant proche du pKa du chromophore au sein de la protéine (7,0) laisse penser que le cristal contient un mélange de la forme neutre et anionique du chromophore. Cependant, la bonne résolution de la structure suggère que l'environnement protéique est peu affecté par le changement d'état de protonation du chromophore. Cette structure donne donc un point de départ correct pour une dynamique de la forme anionique du chromophore.

Les acides aminés 'titrables' ont été choisis dans leur état standard de protonation à pH neutre, sauf pour Glu222 dont la structure cristalline suggère qu'il est donneur de proton dans une liaison H (Fig. IV.32).

Après la mise à l'équilibre du système (décrite dans le section C.6) une simulation de 20 ns avec le chromophore à l'état fondamental est effectuée. La structure globale de la protéine commence à dévier de la structure cristalline à partir de la moité de la simulation (Fig. IV.35). L'écart quadratique moyen (*root mean square deviation* - rmsd) des carbones α des brins β par rapport à la structure cristalline est stable et égal à $0,8 \pm 0,1$ Å pendant la première moitié de la simulation, puis il augmente régulièrement. Nous n'avons pas cherché à analyser l'origine de cette dérive, et n'avons retenu pour la suite que les 12 premières ns de la simulation.

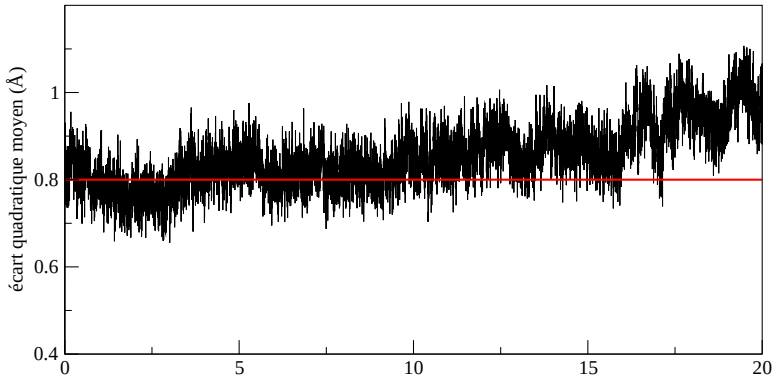


FIG. IV.35.: L'écart quadratique moyen entre les carbones α des brins β au cours de la mise en équilibre et de la simulation de 20 ns à l'état fondamental de la GFP, avec comme référence la structure cristalline.

Au cours de cette dynamique, le chromophore reste proche d'une géométrie plane. La valeur moyenne de l'angle dièdre τ, $7^\circ \pm 7^\circ$, est très proche de la valeur dans la structure cristalline (4,7°). En revanche, la valeur moyenne de φ, $7^\circ \pm 7^\circ$ est un peu plus grande et de signe opposé à la valeur dans la structure cristalline (-2,3°). Cette différence peut venir des effets de température lors de la dynamique ou être lié au π stacking avec Tyr203, dont on va voir par la suite qu'il n'est pas représenté de façon optimale.

La présence des liaisons H autour du chromophore au cours de la dynamique est montrée sur la figure IV.36. La liaison H entre l'oxygène de l'imidazolinone et Arg96 est très stable alors que celle avec Gln69 est de temps en temps rompue et remplacée par une liaison H avec Gln94. On observe une liaison H entre Gln69 et Tyr203, non permanente, non présente dans la structure cristalline, dont l'existence est fortement corrélée avec celle entre Gln69 et l'imidazolinone (Fig. IV.37). Il y a ainsi une boucle de liaisons H entre l'oxygène et l'azote N2 de l'imidazolinone, passant par Gln69, Thr203, W2 et Glu222.

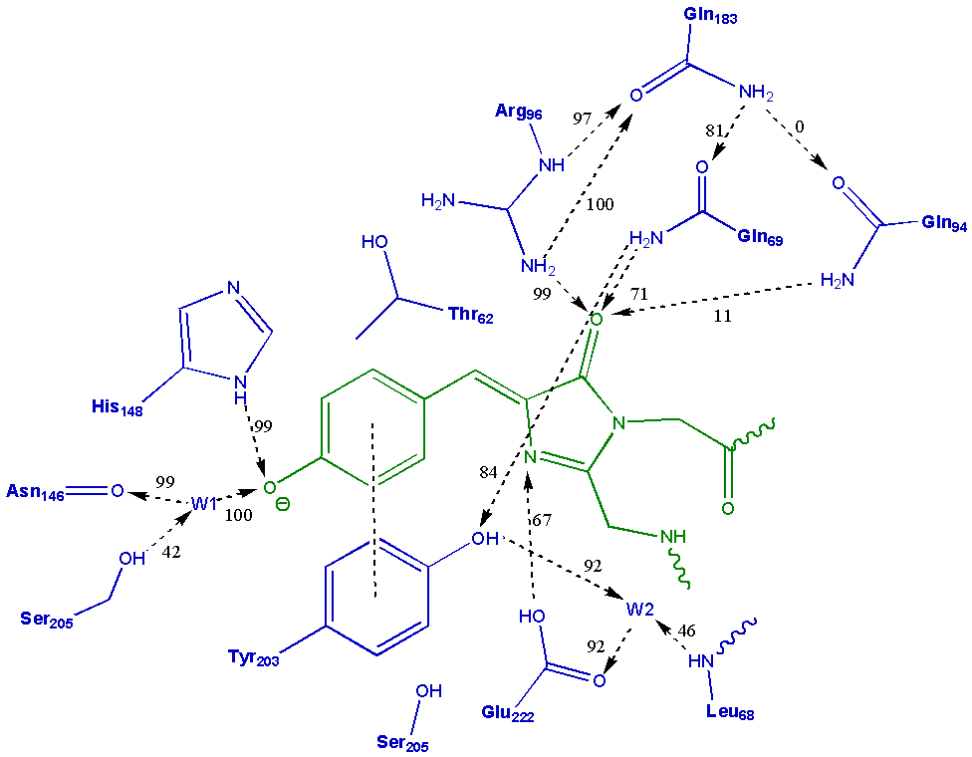


FIG. IV.36.: Le chromophore de la YFP (en vert) et une partie du réseau de liaison hydrogène qui l'entoure. La présence des liaisons hydrogène au cours de la dynamique à l'état fondamental est indiquée en % et est basée sur les critères suivants : la distance entre atomes lourds (X) doit être inférieure à 3,2 Å et l'angle X-H-X supérieur à 120°.

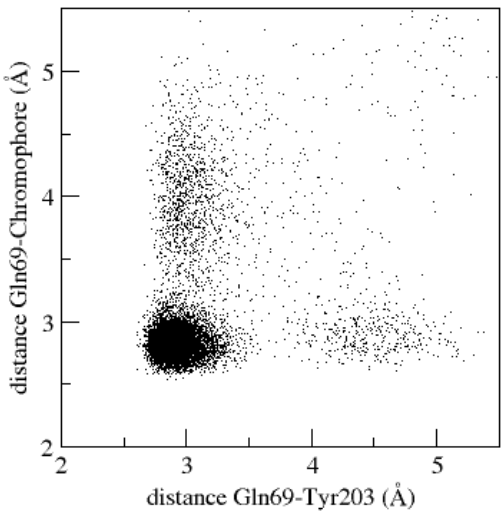


FIG. IV.37.: Corrélation entre la liaison H entre Gln69 et le chromophore et la liaison H entre Gln69 et Tyr203. La figure affiche la distance entre atomes lourds de chaque liaison H.

IV.2.2.1. Le π- stacking entre le chromophore et Tyr203

Dans la structure cristalline, le cycle phénolate du chromophore et le cycle phénol de Tyr203 se trouvent en position de π-stacking. Afin de caractériser la position relative de ces deux cycles, nous avons introduit plusieurs paramètres. On définit tout d'abord un repère XYZ pour chaque cycle. L'axe X est suivant *CD1-CD2*, l'axe Y suivant la normale au plan du cycle, l'axe Z suivant *CG2-CZ* (en supposant un cycle parfaitement plan et symétrique). Soient XX, YY et ZZ les angles entre les axes respectifs de chaque cycle. Ces trois angles caractérisent l'orientation relative des deux cycles. Nous avons également introduit deux distances. Soient C et T les centres de gravité respectifs des cycles du chromophore et de la tyrosine, et H la projection de T sur le plan du cycle du chromophore. Dans le cas où les deux cycles sont parallèles, l'angle YY est nul, les angles XX et ZZ caractérisent la rotation d'un des cycles autour de leur normale commune, TH est la distance entre les deux plans, et CH le décalage entre les deux cycles.

Les valeurs de ces paramètres dans la structure cristalline sont montrées dans le tableau IV.2, colonne 2 ("RX"). Les deux cycles ne sont pas tout à fait parallèles (angle YY = 12°) et il y a un décalage de 0,9 Å entre eux . Les résultats de la simulation sont présentés par tranche de 2 ns successives, afin de voir l'évolution dans le temps. On voit que l'angle entre les deux cycles oscillent autour de 0°, avec des fluctuations de l'ordre de 11°, ce qui est en accord correct avec la structure cristalline. Par contre le décalage entre les deux cycles (distance CH) est en moyenne deux fois plus grande que dans la structure cristalline (1,8 Å au lieu de 0,9 Å). On a donc dans la simulation une position de Tyr203 qui peut s'avérer moins efficace pour bloquer les torsions du chromophore que dans la structure cristalline.

Le décalage entre les résultats des simulations et les données RX peut provenir d'effets spécifiques dans la structure cristalline mais aussi d'une représentation imparfaite du π-stacking au travers des paramètres du champ de forces.

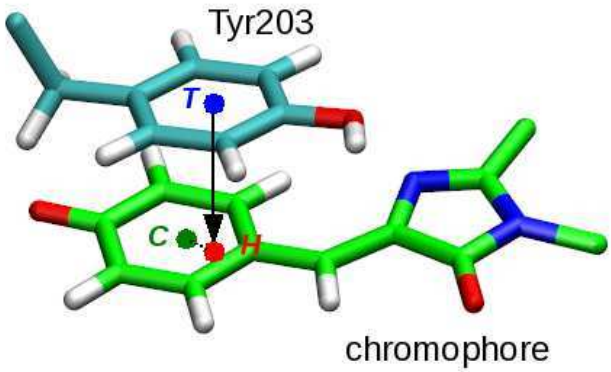


FIG. IV.38.: Le centre de gravité du cycle de phénolate du chromophore ($\boldsymbol{C}$) est indiqué par un point vert, le centre de gravité du cycle de phénol de Tyr203 ($\boldsymbol{T}$) est indiqué par un point bleu et la projection de $\boldsymbol{T}$ dans le plan du cycle phénol ($\boldsymbol{H}$) est indiqué par un point rouge.

TAB. IV.2.: Les paramètres décrivant la position relatives du cycle phénolate du chromophore et du cycle phénol de Tyr203 dans la structure cristalline (RX) et le long de la simulation MD à l'état fondamental (valeur moyenne et fluctuations entre parenthèses). Les angles sont donnés en degrés et les distances en Ångström. Les résultats de la simulation sont données par blocs de 2 ns successifs.

		MD					
	RX	0-2 ns	2-4 ns	4-6 ns	6-8 ns	8-10 ns	10-12 ns
angle XX	11,9	12(6)	12,6(5,7)	11(5)	12(6)	13(5)	16(6)
angle YY	12,0	0(11)	0(11)	0(10)	0(12)	0(10)	0(11)
angle ZZ	2,0	10(5)	10,8(5,1)	9(4)	9(4)	11(4)	13(5)
distance T-H	3,6	3,6	3,4(0,2)	3,4(0,2)	3,4(0,2)	3,5(0,2)	3,4(0,2)
distance C-H	0,9	1,8(0,7)	1,9(0,7)	1,5(0,5)	1,6(0,5)	1,9(0,5)	1,9(0,5)

IV.2.3. Dynamique moléculaire à l'état excité

Nous avons effectué 30 simulations, de durée maximale 8 ns, avec le chromophore à l'état excité. Les coordonnées et les vitesses initiales sont prises de façon régulière, tous les 0,4 ns, dans les premières 12 ns de la simulation de l'état fondamental. Les simulations sont faites par blocs de 2 ns et arrêtées après qu'une torsion 'prolongée' (supérieure à 12 ps, voir chapitre V) du chromophore a été observée. Par conséquent, les simulations ont des durées différentes, qui donnent un temps de simulation total de 120 ns.

Sur l'ensemble des 30 simulations, on observe différents aspects de la dynamique de torsion du chromophore semblables à ceux de la dynamique de la GFP (section IV.1). La figure Fig. IV.39 présente, superposées à la PES interne du chromophore, les géométries visitées au cours des 120 ns de simulation de la YFP, ainsi que celles visitées lors des simulations du chromophore dans le vide (section III.5). La comparaison de ces deux figures permet de visualiser plusieurs effets de la protéine. Alors que les 5 minima de la PES du chromophore (FS, TwP^+, TwP^-, TwI^+ et TwI^-) sont explorés de façon symétrique et uniforme dans la dynamique du chromophore isolé, ils sont visités de façon très inégale et dissymétrique en présence de la protéine. Les contraintes de la protéine limitent fortement les torsions vers la région TwP^+ et complètement les torsions vers TwP^- et TwI^-. Les torsions majoritaires sont celles vers la région TwI^+. On remarque qu'au cours de ces torsions le chromophore n'atteint pas le fond du puits ($\tau = 90^\circ$, $\varphi = 0^\circ$) mais reste bloqué à $\tau = 84^\circ \pm 7^\circ$ et $\varphi = -21^\circ \pm 10^\circ$. Aucune torsion de type hula-twist ($\tau = \pm 90^\circ$ et $\varphi = \pm 90^\circ$) n'est observée.

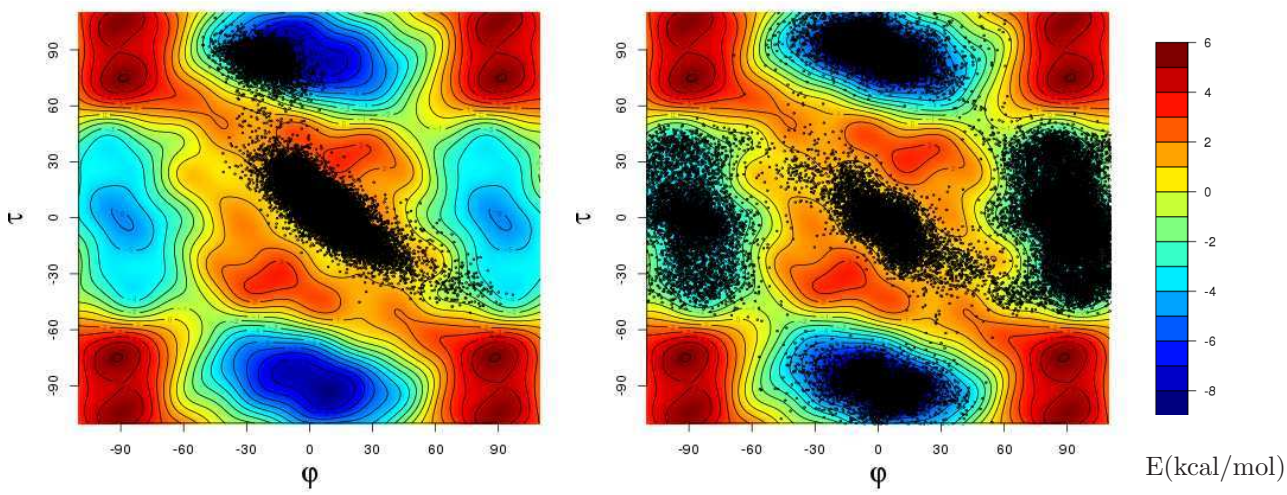


FIG. IV.39.: La surface d'énergie potentielle de torsion en fonction de τ (abscisse) et φ (ordonnée). Les configurations visitées au cours d'une trajectoire sont représentées par des points noirs. A gauche : les 30 trajectoires dans la YFP. A droite : plusieurs trajectoires dans le vide.

La figure IV.40 montre les trajectoires des angles τ et φ dans l'une des 30 simulations, où plusieurs torsions en τ sont observées. Trois aspects dynamiques des torsions apparaissent sur cette figure. Premièrement, la protéine peut maintenir le chromophore dans sa géométrie plane (FS) pendant plusieurs nanosecondes mais chaque passage entre FS et une géométrie perpendiculaire se fait en moins de 2 ps. Deuxièmement, le chromophore réside plus ou moins longtemps en géométrie perpendiculaire : dans certain cas on observe un aller-retour d'un 'snapshot' à l'autre, mais dans d'autres cas le temps de résidence en géométrie perpendiculaire peut être supérieur à plusieurs centaines de picosecondes. Troisièmement, le chromophore quitte la région de la géométrie plane (FS) par une torsion concertée entre τ et φ avec une corrélation négative ($\tau > 0^\circ$ et $\varphi < 0^\circ$).

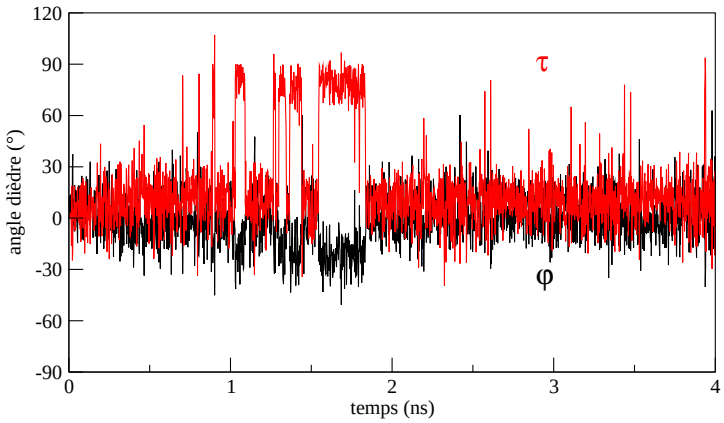


FIG. IV.40.: Simulation de 4 ns de la YFP. Le chromophore visite la région TwI^+ plusieurs fois. L'angle τ fluctue alors autour de la valeur 84° et l'angle φ autour de -21°.

La distribution des angles φ et τ au cours des 120 ns de simulation est également présentée sous

forme d'histogrammes sur la figure. IV.41. On y voit clairement que les torsions vers $\tau = 90^\circ$ sont beaucoup plus fréquentes que les torsions vers $\varphi = 90^\circ$ au sein de la YFP.

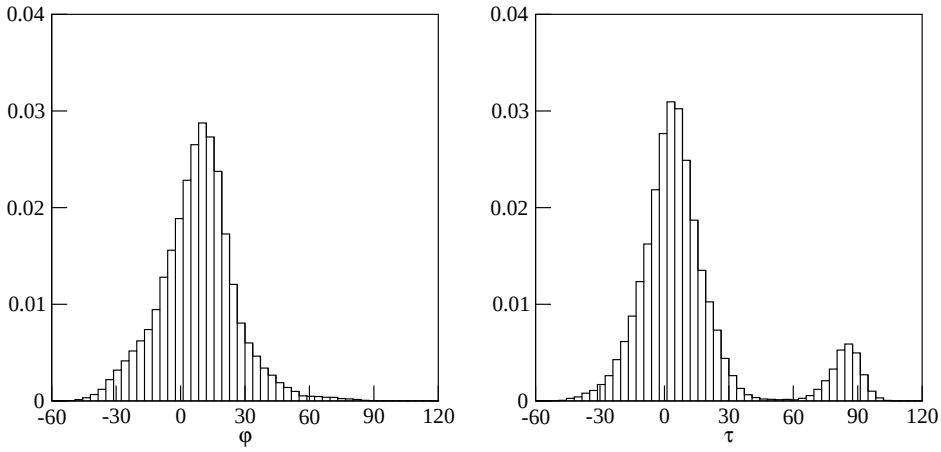


FIG. IV.41.: La distribution des valeurs de φ (à gauche) et τ (à droite) au cours des 120 ns de simulation.

IV.2.3.1. Analyse du mécanisme moléculaire des torsions τ

Deux facteurs doivent fortement contribuer à contraindre les torsions du chromophore. D'une part les deux résidus Tyr203 et Thr62 créent une répulsion stérique avec le cycle phénolate lors d'une torsion τ ou φ, et maintiennent le chromophore dans une géométrie plane ($\tau = 0^\circ$, $\varphi = 0^\circ$) (Fig. IV.34). Cette conformation du chromophore est également favorisée par des interactions van der Waals entre le phénolate et le cycle phénol de Tyr203 (type π-stacking).

D'autre part, l'oxygène du phénolate du chromophore est lié par des liaisons H à His148 et à une molécule d'eau (W1).

Afin d'éviter ces contraintes stériques et de garder ses points d'attaches à l'environnement, le chromophore s'arque et réoriente le cycle imidazolinone (Fig. IV.42) lors des torsions vers TwI^+ . De cette manière, le cycle phénolate se déplace très peu et son oxygène peut maintenir les liaisons H avec His148 et W1 (Fig. IV.43). Cette déformation spécifique du chromophore combine à la torsion τ une torsion φ de petite amplitude en sens inverse ($\varphi \approx -21^\circ$), et une pyramidalisation de l'atome *N3* (nomenclature voir Fig. IV.33) (Fig. IV.44).

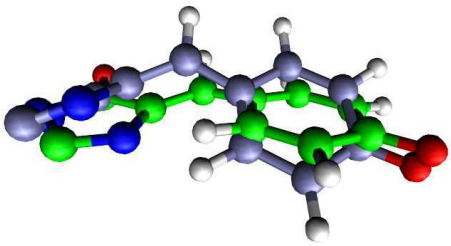


FIG. IV.42.: Superposition du chromophore dans deux snapshots successifs (plan en vert, perpendiculaire en bleu) dans la dynamique de la YFP. Grâce à une petite torsion en φ et une réorientation de l'imidazolinone, l'axe du phénolate se déplace peu et l'oxygène peut maintenir ses liaisons H.

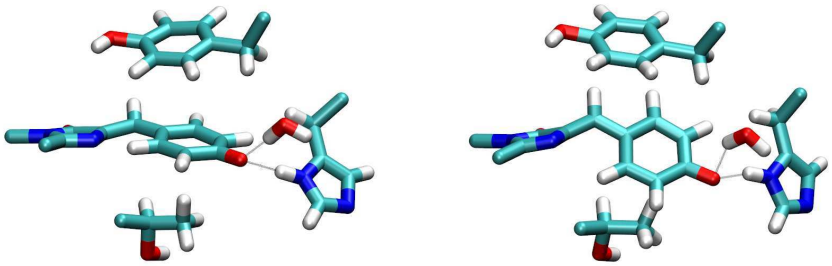


FIG. IV.43.: Le chromophore dans sa géométrie plane (à gauche) et perpendiculaire (à droite). Les liaison H avec His148 et la molécule d'eau sont maintenues. La gêne stérique des résidus Tyr203 et Thr62, placé au-dessus et en-dessous du chromophore, suscite une déformation de l'imidazolinone lors de la torsion τ .

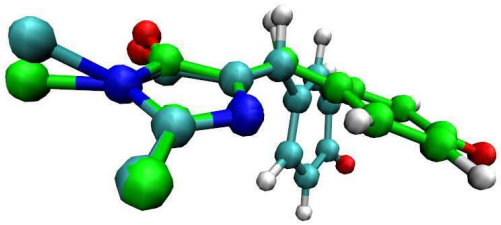


FIG. IV.44.: Superposition des atomes de l'imidazolinone de deux géométries du chromophore : une géométrie plane (en vert) et une géométrie perpendiculaire de la région TwI^+ (en cyan) dans la dynamique de la YFP. L'azote *N3* de l'imidazolinone se pyramidalise alors que le carbon *C1* reste plan.

L'angle moyen de pyramidalisation de N3 dans la région de TwI^+ ($-11 \pm 7^\circ$) est significativement différent de l'angle moyen dans la région de FS ($2 \pm 8^\circ$) (Fig. IV.45, à gauche). En revanche, la modification de l'angle de pyramidalisation de *C1* est négligeable : $4 \pm 6^\circ$ dans la région de TwI^+ et $3 \pm 6^\circ$ dans la région de FS (Fig. IV.45, à droite). De ce fait, l'angle dièdre autour de la liaison C1-N3 est également modifié, il passe de $1 \pm 10^\circ$ dans la région de FS à $22 \pm 10^\circ$ dans la région de TwI^+ (Fig. IV.46). Cette déformation s'étend dans les liaisons reliant l'imidazolinone au reste de la protéine, que nous appellerons les 'attaches peptidiques'. Il s'avère que l'impact est plus grand du côté de *N3*, i.e. le côté de la chaîne peptidique de Gly67 (le chromophore étant formé de Gly65-Tyr66-Gly67) (Fig. IV.46 et IV.47).

Sur l'ensemble des coordonnées visitées au cours des 120 ns de dynamique on observe donc une déformation significative du cycle imidazolinone et de ses attaches peptidiques lors du passage dans la région de TwI^+. Cette déformation est le résultat de deux contraintes, exercées sur le chromophore, qui s'opposent. La protéine, majoritairement les résidus Thr62, Tyr203 et His148, empêche des grands déplacements du cycle phénolate alors que le potentiel interne du chromophore favorise une torsion en τ de grande ampleur.

Afin de savoir si la déformation de l'imidazolinone est une conséquence de l'opposition entre les contraintes exercées par la protéine et les forces internes du chromophore qui le tirent vers la région TwI+, ou si elle se produit avant que le chromophore ait quitté la région de FS, nous avons suivi l'angle de pyramidalisation de l'atome N3 le long du chemin d'énergie minimum (MEP) entre les deux minima FS et TwI+ de la surface d'énergie potentielle NADiA. Nous avons ensuite sélectionné toutes les géométries de la dynamique de 120 ns de la YFP dans lesquelles les angles τ et φ sont proches du MEP (Fig. IV.48 à gauche). Afin d'augmenter la statistique dans la région où le chromophore passe très peu de temps, i.e. où $40^\circ < \tau < 60^\circ$, nous avons effectué une simulation supplémentaire de 2 ns où un biais favorisant l'angle $\tau = 50^\circ$ est imposé (en vert sur la figure IV.48). L'évolution de l'angle moyen de pyramidalisation de l'atome *N3* le long de du MEP est montrée sur la figure IV.48 à droite.

On observe une baisse régulière mais assez faible de l'angle de pyramidalisation dans la région de FS. L'angle de pyramidalisation de *N3* passe de $+3^\circ \pm 8^\circ$ dans la géométrie plane ($\tau = 0^\circ$) à $-3^\circ \pm 7^\circ$ à la sortie de la région de FS ($\tau = 38^\circ$). La pyramidalisation de *N3* n'augmente significativement qu'à $\tau > 60^\circ$, où les forces internes de torsion du chromophore sont fortes. Cette pyramidalisation accompagne de façon progressive et corrélée la torsion en τ et ne la précède pas.

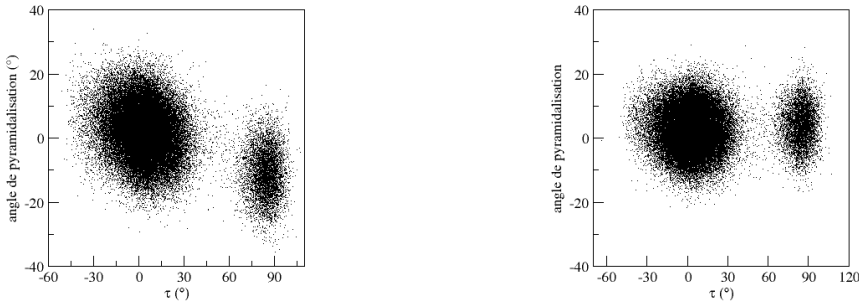


FIG. IV.45.: La pyramidalisation de l'atome *N3* (à gauche) et de l'atome $C1$ (à droite) de l'imidazolinone du chromophore en fonction de l'angle τ.

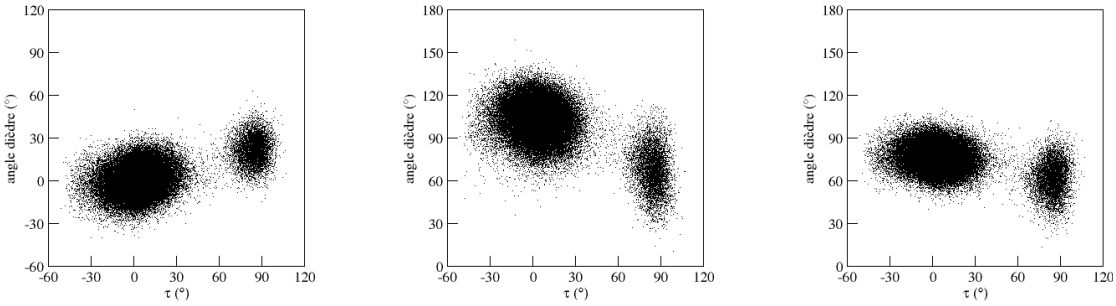


FIG. IV.46.: L'angle dièdre *CA3-N3-C1-CA1* (à gauche), l'angle dièdre *C3-CA3-N3-C1* (à centre) et *N1-CA1-C1-N3* (à droite) en fonction de l'angle τ (nomenclature voir Fig. IV.33).

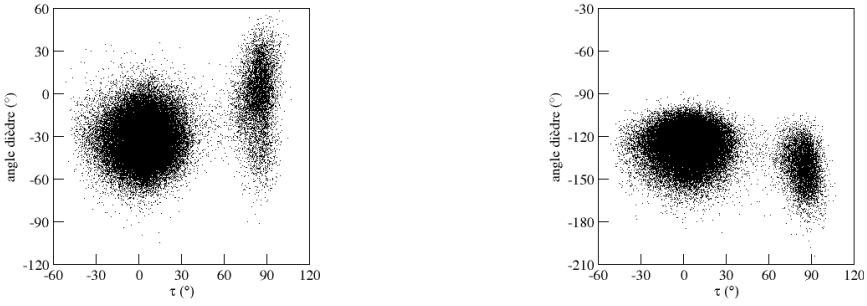


FIG. IV.47.: L'angle dièdre *N-C3-CA3-N3* (à gauche) et *C-N1-CA1-C1* (à droite) de l'imidazolinone du chromophore en fonction de l'angle τ (nomenclature voir Fig. IV.33) .

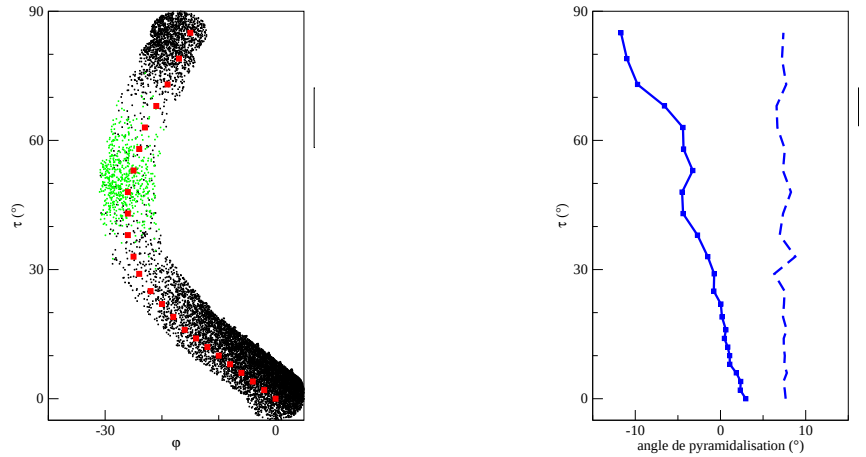


FIG. IV.48.: A gauche : le chemin d'énergie minimum (MEP) entre la région de FS ($\tau = 0°, \varphi = 0°$) et la région de TwI$^+$ ($\tau = 90°, \varphi = 0°$) est indiqué par des carrés rouges. Une sélection de géométries qui présentent des angles de τ et φ proches de ceux du MEP parmi les géométries visitées au cours des 120 ns de dynamique de la YFP (en noir) ou des 2 ns d'une dynamique biaisée (en vert).
A droite : l'angle moyen de pyramidalisation de N3 (traits pleins) et son écart type (tirets) dans chaque point du MEP, basé sur la sélection de géométries issues des simulations de la YFP (points noirs et verts sur la figure à gauche).

Nous avons analysé les mouvements spécifiques de la protéine au moment du passage dans la région de TwI$^+$. On peut remarquer deux variations significatives dans l'environnement du chromophore. Premièrement, la distance moyenne entre les chaînes latérales de Tyr203 et Thr62 augmente (Fig. IV.49) de 7,7 Å dans la géométrie plane à 8,6 Å dans la géométrie perpendiculaire. Deuxièmement, la liaison H entre l'azote *N2* de l'imidazolinone et Glu222 est plus stable dans la géométrie TwI$^+$que dans la géométrie plane (Fig. IV.50, à droite), tandis que les autres liaisons H impliquant le chromophore ne sont pas modifiées.

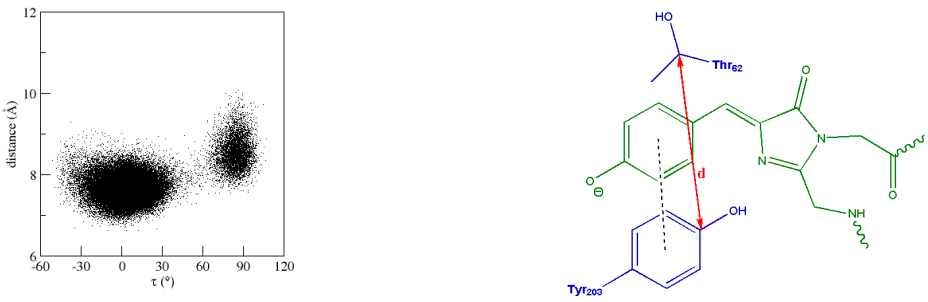


FIG. IV.49.: La distance entre le carbone *CB* de Thr62 et le carbone *CZ* de Tyr203 en fonction de τ.

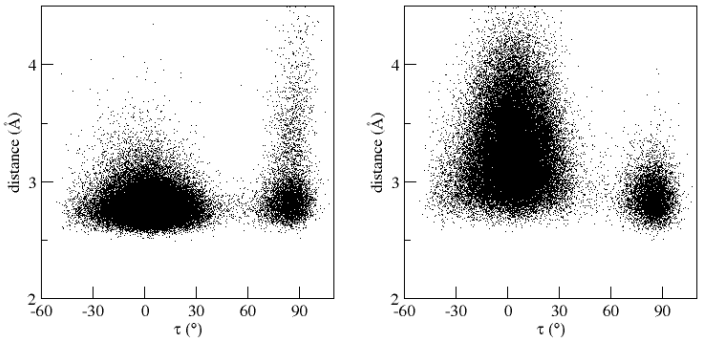


FIG. IV.50.: Liaison H entre *O2* de l'imidazolinone et Arg96 (à gauche), et entre *N2* de l'imidazolinone et Glu222 (à droite), en fonction de l'angle τ.

Nous allons examiner successivement ces deux caractéristiques.

Dans l'éloignement entre Thr62 et Thr203, le déplacement de Thr62 relatif au chromophore est limité par le fait que ces deux résidus sont tous les deux placés sur l'hélice α centrale. De plus, la rotation de la chaîne latérale de Thr62 est restreinte par une liaison H stable entre son groupement hydroxyle et le cycle imidazole de His181 (Fig. IV.51). Considérant que les positions relatives de Thr62 et du chromophore changent peu quand on passe des conformations planes aux conformations perpendiculaires, nous pouvons interpréter la variation de la distance Thr62-Tyr203 principalement comme un déplacement de Tyr203.

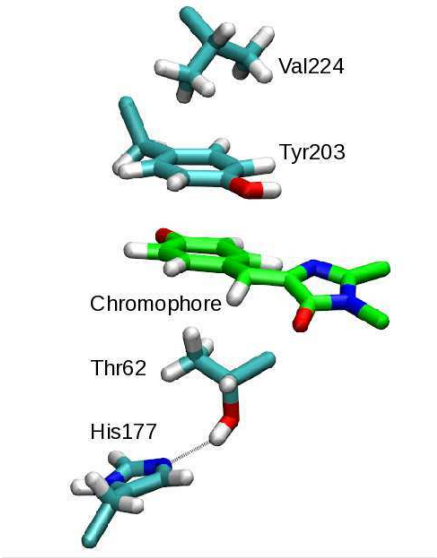


FIG. IV.51.: Le chromophore (en vert) et les deux résidus Tyr203 (au-dessus du chromophore) et Thr62 (en dessous du chromophore) qui bloquent les torsions τ et φ du chromophore. La rotation de la chaîne latérale de Tyr203 est empêchée par Val224 (en haut) (gène stérique) et celle de Thr62 par His177 (en bas) (liaison H).

En effet, la distance moyenne entre le brin β_{10} portant Tyr203 (en utilisant la même numérotation des brins que dans la GFP) et l'hélice α portant le chromophore, est plus grande quand le

chromophore se trouve dans la région de TwI^+ que quand il se trouve dans la région de FS (Fig. IV.52). Ce déplacement a été identifié au niveau de Ser202, Tyr203 et Gln204, placées sur le brin β_{10} en face du chromophore. Il s'agit donc d'un mouvement corrélé de plusieurs résidus voisins. En revanche, nous n'avons pas trouvé de corrélation significative entre la torsion en τ et la rotation du cycle phénol de Tyr203 ou de la torsion de sa chaîne latérale (autour de la liaison *CA-CB*).

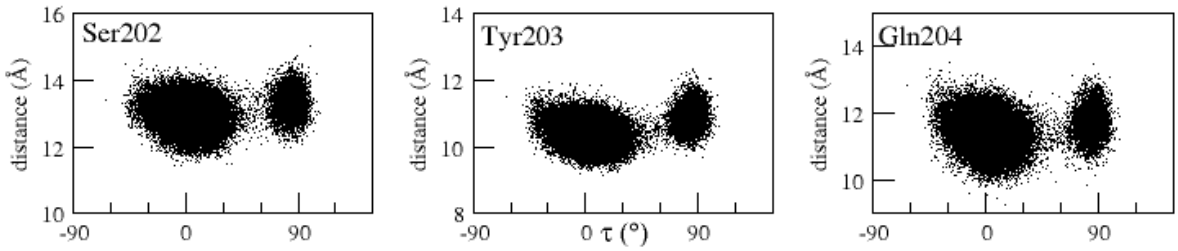


FIG. IV.52.: Déplacement du brin β_{10} portant de Tyr203 en fonction de τ. Distances entre le carbone α de Thr62 et les carbones α des résidus indiqués sur chaque graphe.

Nous examinons maintenant la deuxième caractéristique : le renforcement de la liaison H entre l'imidazolinone et Glu222.

Les résidus Arg96 et Glu222 sont placés de part et d'autre du tonneau β de la protéine (Fig. IV.53). La figure IV.50 montre que le cycle imidazolinone maintient ses liaisons H avec Arg96 et Glu222 dans les deux géométries, plane et perpendiculaire. S'ils s'écartent au moment où tous les deux sont en liaison H avec l'imidazolinone, ils soumettent le cycle à deux forces de sens opposé qui peuvent le faire légèrement basculer, ce qui augmente la valeur de τ et peut déclencher une torsion τ du chromophore (Fig. IV.54).

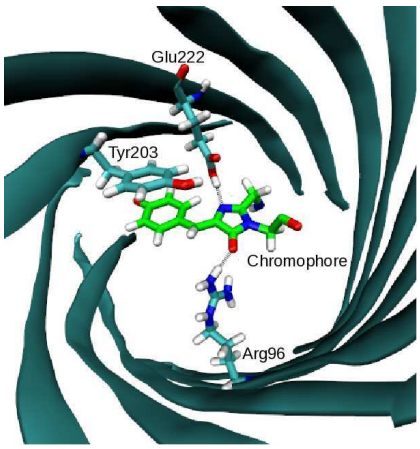


FIG. IV.53.: Le chromophore au centre en vert. Arg96 en bas (brin β_3), Tyr203 (brin β_{10}) en π-stacking avec le chromophore et Glu222 en haut (brin β_9).

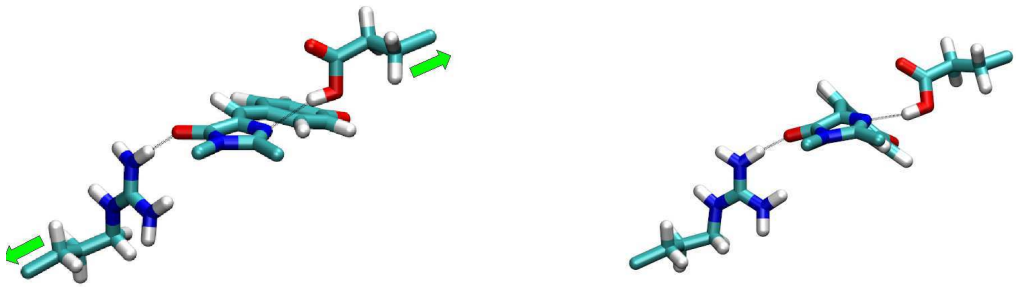


FIG. IV.54.: Le chromophore dans sa géométrie plane (à gauche) et perpendiculaire (à droite). La liaison H avec Arg96 est maintenue et la liaison H avec Glu222 est renforcée.

Nous avons donc étudié le déplacement des brins β_3 et β_9, portant Arg96 et Glu222 respectivement, vis-à-vis de l'hélice α portant le chromophore, en utilisant le carbone α de Thr62 comme point de référence. Aucune corrélation entre la torsion en τ et la distance entre les carbones α de Arg96 et Thr62 n'est observée. En revanche, la distance moyenne entre les carbones α de Glu222 et Thr62 augmente quand on passe de la région de FS ($\tau = 0^\circ$) à la région de TwI^+ ($\tau = 90^\circ$) (Fig. IV.55). Cette tendance est également observée pour les résidus voisins de Glu222, placés sur le brin β_9 : Phe223, Val224 et Thr225. Il s'agit donc d'un déplacement du brin β_9 et non du seul mouvement local de Glu222.

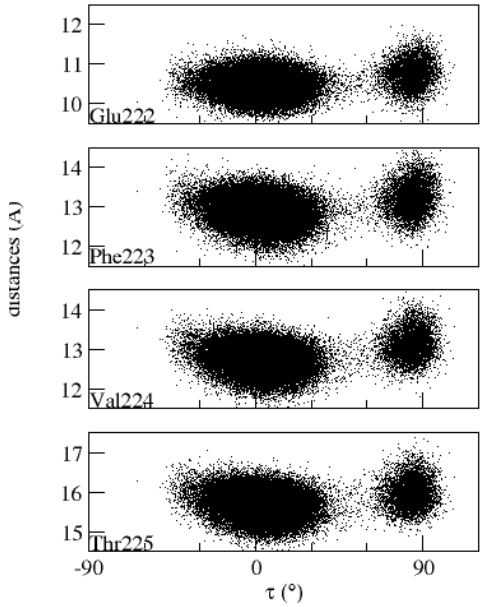


FIG. IV.55.: Déplacement du brin β_9 portant Glu222 en fonction de l'angle τ. Distance entre le carbone α de Thr62 et les carbones α des résidus indiqués sur chaque graphe. La distance est en moyenne plus grande quand le chromophore visite les régions de géométrie perpendiculaire : TwI^+ à $\tau > 60^\circ$ ou TwP^+ à $\tau < -40^\circ$

IV.2.3.1.1. Conclusions sur les torsions en τ .

Nous avons identifié plusieurs mouvements clés de la protéine liés à la torsion en τ. D'une part, le brin β_{10} doit s'éloigner de l'hélice α centrale afin d'écarter la chaîne latérale de Tyr203 et libérer de la place pour la torsion du chromophore. Ce déplacement affaiblit *a priori* l'interaction de π-stacking et lève ainsi une deuxième contrainte. D'autre part, le brin β_9 doit également s'écarter du centre du tonneau afin d'éloigner Glu222 du cycle imidazolinone. Le mouvement doit se produire au moment où Glu222 est en liaison H avec l'azote *N2* afin de faire basculer le cycle et amener le chromophore jusqu'à la sortie de la région FS. Pour y arriver le chromophore doit également se déformer au niveau des jonctions covalentes avec le reste de la protéine. De cette façon il peut maintenir les liaisons H du phénolate avec His148 et W1.

IV.2.3.2. Analyse des torsions en φ

Les torsions en φ amènent le chromophore dans la région de TwP^+, jamais dans TwP^-. Les passages dans TwP^+sont très brefs, le chromophore y reste en moyenne environ 2 ps.

On n'observe pas de changement de l'angle de pyramidalisation de l'atome *N3* pendant les torsions. En revanche, on observe une corrélation négative entre l'angle de pyramidalisation de l'atome *CA2* et la torsion en φ (Fig. IV.56). L'atome *CA2* se pyramidalise plus, et dans le sens contraire, lors des passages dans la région de TwP^+ ($\varphi > 60^\circ$, $\tau \approx -30^\circ$) que lors des passages dans la région de TwI^+ ($\varphi \approx -20^\circ$, $\tau > 60^\circ$). Cette pyramidalisation donne une forme courbée au chromophore qui lui permet d'éviter un répulsion stérique avec Thr62 et Tyr203 et qui permet de garder le π-stacking avec Tyr203. On observe également un écartement entre la chaîne latérale de Thr62 et Tyr203 (Fig. IV.56). Ce déplacement est moins important que celui observé lors des torsions en τ.

Nous n'avons pas cherché à analyser ces torsions plutôt instables de façon plus détaillée.

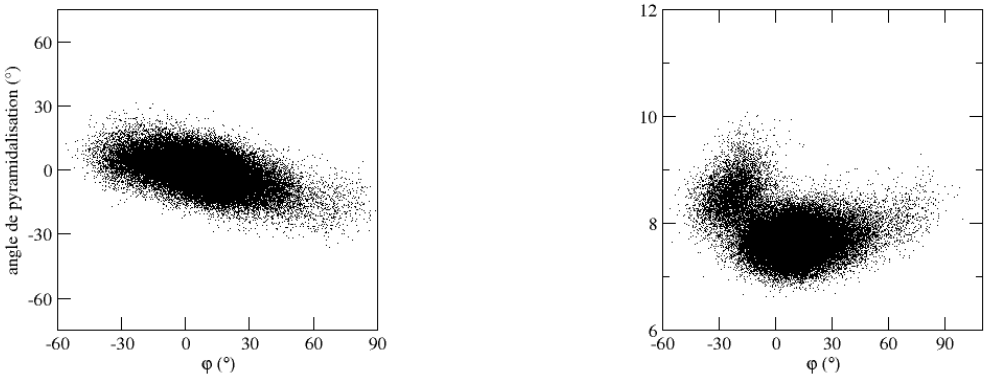


FIG. IV.56.: A gauche : l'angle de pyramidalisation de l'atome *CA2* en fonction de φ. A droite : la distance entre le carbon *CB* de Thr62 et le carbone *CZ* de Tyr203 (Fig. IV.49) en fonction de φ. Les passages dans la région TwP^+ correspondent à $\varphi > 60^\circ$ alors que les passages dans la région de TwI^+ correspondent à $\varphi \approx -20^\circ$.

IV.2.4. Conclusion

Dans la YFP, la désactivation non-radiative se produit principalement par la torsion en τ du chromophore, qui se fait dans le sens opposé à Tyr203. Par rapport à la GFP, la mutation Thr203 en Tyr203 a pour effet d'inhiber presque totalement les torsion en φ et de ne laisser la place qu'à des torsions en τ. Ces torsions ont lieu selon un mécanisme impliquant de petits déplacements ou déformations de brins β dont des chaînes latérales sont en interaction avec le chromophore. Peu coûteux en énergie, ces mouvements combinés de quelques résidus n'augmentent pas notablement le volume libre autour du chromophore mais ils lui permettent de se faufiler jusqu'à la sortie de la région FS sans perdre la plupart des interactions avec la protéine. Ces mouvements clés ont cependant une probabilité faible de se produire simultanément et suffisamment longtemps.

IV.3. Analyse de Padron

On rappelle le schéma d'activation de fluorescence de Padron où le chromophore *trans* non-fluorescent peut être transformé en la forme *cis* fluorescente :

$$\underset{\text{(désactivé)}}{trans} \rightleftarrows \underset{\text{(activé)}}{cis}$$

Il a été montré [82] que la fluorescence de la forme *trans* du chromophore au sein de Padron n'augmente pas quand la température est abaissée (170 K) contrairement à ce qui est observé en solvant [125]. Il en a été conclu que la désactivation non-radiative qui limite la fluorescence dépend d'autre chose que des torsions du chromophore. Cependant, Faro *et al.* [83] ont mis en évidence que l'isomérisation *trans*→*cis* du chromophore au sein de Padron est possible même à basse température (100 K) où l'environnement protéique est très rigide. Cependant, le rendement est très faible ($\sim 5.10^{-6}$) par rapport au rendement à température ambiante ($\sim 2.10^{-4}$). Cela indique, au contraire, que les contraintes de la protéine empêchent peu les torsions du chromophore. Par ailleurs, les résultats de Faro *et al.* montrent que l'isomérisation se produit sans changement d'état de protonation du chromophore (*trans* anionique→*cis* anionique), ce qui contredit l'hypothèse d'une désactivation par protonation du chromophore.

Nous avons effectué des simulations du chromophore *trans* anionique à l'état excité, au sein de Padron, à basse température (100 K), afin d'étudier sa flexibilité et d'analyser le mécanisme moléculaire qui permet une isomérisation.

IV.3.1. L'environnement du chromophore

Dans la structure cristalline, et aussi dans la dynamique, le chromophore est entouré par un réseau de liaisons qui reste stable au cours de la simulation : d'une part, les deux liaisons covalentes faisant la jonction avec la protéine, d'autre part, plusieurs liaisons hydrogène indiqué sur la figure IV.57 (l'oxygène O2 de l'imidazolinone (noms d'atome dans Fig. IV.58) avec Arg91 et Arg66, l'oxygène du phénolate avec Tyr159 et la molécule d'eau W1). Les mouvements du chromophore sont également contraints par des encombrements stériques : d'un côté par la chaîne latérale de Met59, de l'autre côté par le cycle imidazolinone de His193, liés par des liaisons H aux groupements carboxyles de Glu211 et Glu144.

Ces trois résidus Glu144, His193 et Glu211, appelés par la suite *la triade*, peuvent exister dans plusieurs états de protonation. Le cycle imidazolinone de His193 peut être protoné (neutre) sur l'azote δ ou sur l'azote ε ou bien doublement protoné (cationique). Le groupement carboxyle des acides glutamiques peut être déprotoné (anionique) ou protoné sur l'un des deux oxygènes (neutre). Nous avons envisagé trois schémas de protonation de la triade : 1) Glu144 déprotoné - His193 doublement protoné - Glu211 déprotoné ; 2) Glu144 protoné - His193 protoné sur l'azote δ - Glu211 déprotoné ; 3) Glu144 déprotoné - His193 protoné sur l'azote ε - Glu211 protoné

(Fig. IV.59). La stabilité des trois schémas a été étudiée par des simulations avec le chromophore à l'état fondamental à 100K.

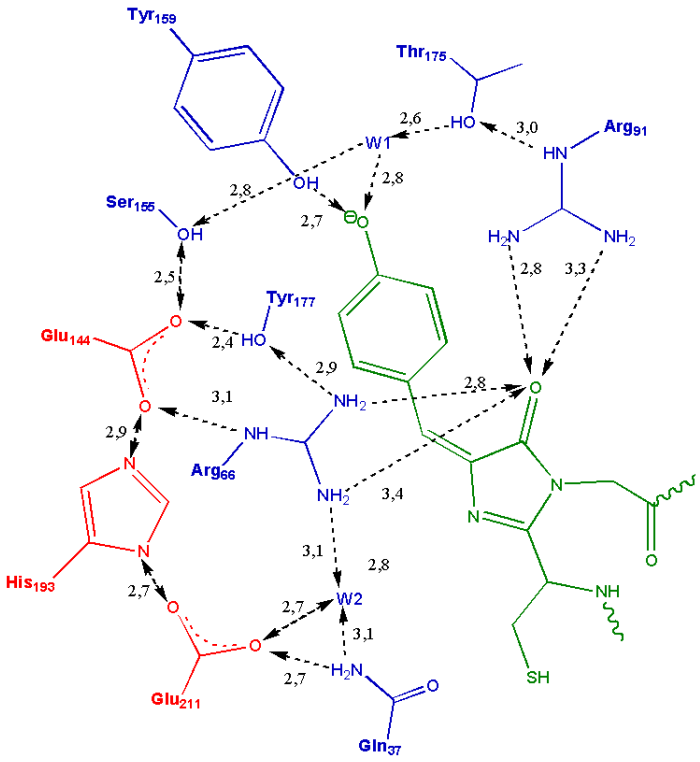


FIG. IV.57.: Le chromophore *trans* de Padron (en vert) et une partie du réseau de liaisons hydrogène dans son environnement. Les distances inter-hétéroatomes sont tirées de la structure cristalline [68] et données en Ångström. Le symbole W représente une molécule d'eau.

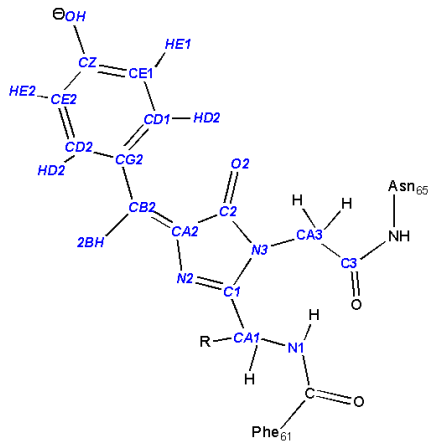


FIG. IV.58.: Le chromophore de Padron (formé par le tripeptide Cys62-Tyr63-Gly64) avec les noms des atomes indiqués en bleu. La chaîne latérale de Cys62 est indiqué par un R. Le groupement carboxyle de Phe61 et le groupement NH de Asn65 sont également montré sur le schéma.

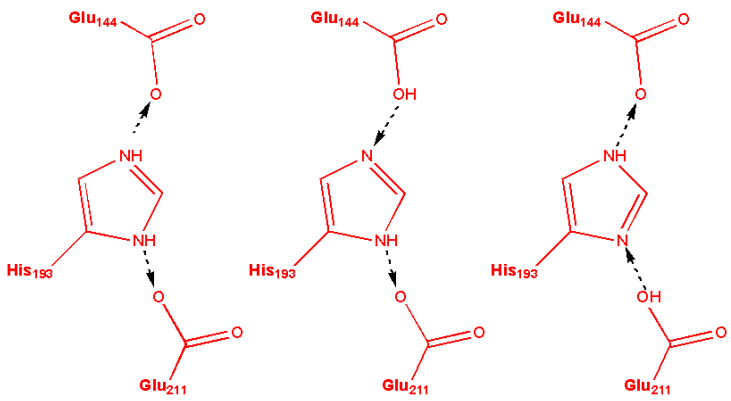


FIG. IV.59.: Les trois schémas de protonation de la triade Glu144 - His193 - Glu211.

IV.3.2. Dynamique moléculaire à l'état fondamental

Dans le cas des acides glutamiques protonés, il existe deux rotamères du groupement carboxyle : *syn* - l'hydrogène est dirigé vers l'intérieur du groupement et *anti* - l'hydrogène est dirigé vers l'extérieur du groupement (Fig. IV.60). Dans la structure cristalline, les groupements carboxyle de Glu211 et Glu144 se trouvent à peu près dans le même plan que l'imidazolinone de His193 avec des orientations qui imposent une position *anti* des hydrogènes de Glu211 ou Glu144 pour pouvoir former les liaisons H. Les hydrogènes des acides glutamiques protonés ont donc initialement été placés en situation *anti* dans les simulations.

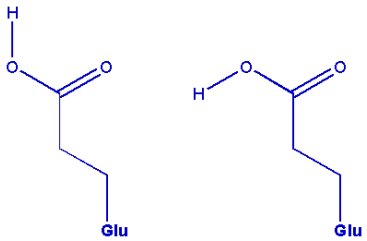


FIG. IV.60.: Les rotamères *syn* (à gauche) et *anti* (à droite) d'un acide glutamique protoné.

Il faut également prêter attention à l'orientation du groupement hydroxyle des tyrosines, qui présente deux rotamères. Dans la géométrie de départ, nous avons choisi l'orientation des groupements hydroxyle de Tyr159 est Tyr177 de façon à permettre l'établissement des liaisons H avec le chromophore et avec Glu144, respectivement.

Pour évaluer quel schéma de protonation de la triade reproduit au mieux les interactions présentes dans la structure cristalline, nous avons analysé les simulations à l'aide de plusieurs critères :

- la stabilité du réseau de liaisons H de la triade
- la stabilité du réseau de liaisons H du chromophore
- la présence ou non d'un π-stacking entre l'imidazolinone de His193 et le phénolate du chromophore

Les résultats des simulations sont résumés dans le tableau IV.3.

TAB. IV.3.: Résultat des simulations avec le chromophore à l'état fondamental pour les trois schémas de protonation de la triade :
Schéma 1 : Glu211 déprotoné - His193 doublement protoné - Glu144 déprotoné
Schéma 2 : Glu211 déprotoné - His193 protoné sur l'azote ε - Glu144 protoné
Schéma 3 : Glu211 protoné - His193 protoné sur l'azote δ - Glu144 déprotoné
Les simulations des schémas de protonation 1, 2 et 3 ont une durée de 6 ns, 6 ns et 2 ns, respectivement. Les liaisons H analysées sont affichées sur la figure IV.57. La distance entre atomes lourds de chaque liaison H dans la structure cristalline (RX) est indiquée en Ångström et leur existence dans les simulations est indiquée en pourcentage du temps avec les critères suivantes : la distance entre atomes lourds (X) doit être inférieure à 3,2 Å et l'angle X-H-X supérieur à 120°. Le nom du chromophore est GYC (glycine (G), tyrosine (Y) et cystéine (C) : les trois résidus à partir desquels ils est formé).

Liaisons H			Distances (Å)	% existence		
			RX	schéma 1	schéma 2	schéma 3
GYC(OH)	-	Tyr159(OH)	2,6	100	100	100
GYC(OH)	-	W1	2,8	100	100	100
GYC(O2)	-	Arg66(NH1/NH2)	2,7 / 3,3	62 / 40	34 / 35	97
GYC(O2)	-	Arg91(NH1/NH2)	2,8 / 3,3	100	100	100
Glu144(OE2)	-	His193(NE2)	2,8	100	85	100
Glu211(OE2)	-	His193(ND1)	2,7	100	100	***0***
Glu144(OE2)	-	Arg66(NE)	3	100	95	100
Glu144(OE1)	-	Ser155(OG)	2,5	100	***0,2***	100
Glu144(OE1)	-	Tyr177(OH)	2,4	100	74	100
Thr175(OG1)	-	W1	2,6	100	63	100
Ser155(OG)	-	W1	2,8	100	100	100
Thr175(OG1)	-	Arg91(NE)	3	100	100	100
Glu144(OE1/OE2)	-	Arg66(NH1/NH2)	>3,9	***86 / 35***	***73 / 36***	***100 / 100***

Le réseau de liaisons H de la structure cristalline est très bien reproduit dans le schéma 1 mais, comme dans les deux autres schémas, il y a une liaison H supplémentaire entre Glu144 et Arg66. Dans le schéma 2 certaines liaisons sont moins stables et une liaison manque, entre Glu144 et Ser155. Les liaisons H entre Ser155 et la molécule d'eau W1, et entre W1 et le phénolate du chromophore, sont maintenues. Le schéma 3 échoue à établir la liaison H entre His193 et Glu211. Pourtant, la structure cristalline suggère fortement l'existence d'une liaison H (distance 2,7 Å entre atomes lourds) entre ces deux résidus. Le schéma 3 doit donc être exclu.

Nous avons ensuite examiné les positions relatives de l'imidazolinone de His193 et du phénolate du chromophore. Dans la structure cristalline, les deux cycles ne sont pas en π-stacking ce qui peut être mis en relation avec le caractère non-fluorescent de la protéine. En effet, l'interaction de π-stacking limiterait les rotations du chromophore, notamment à l'état excité où des torsions τ et φ de grande ampleur sont responsables de l'extinction de fluorescence à l'échelle de la picoseconde. Dans les simulations à 100K, aucun π-stacking n'est observé à l'état fondamental dans les schéma 1 et 2 (observation des trajectoires par VMD). Cependant, dans le schéma 1, le cycle imidazolinone de His193 est chargé positivement. Il peut donc avoir une interaction électrostatique attractive entre ce cycle et le phénolate (chargé négativement) du chromophore.

Les schéma 1 et 2 ont également été examinés par des simulations avec le chromophore à l'état excité.

IV.3.3. Dynamique moléculaire à l'état excité

Les structures engendrées par les simulations avec le chromophore à l'état fondamental (schéma de protonation 1 et 2 de la triade) sont utilisées comme coordonnées initiales dans les simulations à 100K à l'état excité (S_1) du chromophore.

Dans le cas du schéma 1, nous avons effectué 30 simulations à l'état S_1 avec des coordonnées et des vitesses initiales extraites régulièrement de la simulation à l'état fondamental. Dans 28 de ces simulations, on observe une torsion en φ qui se produit au bout d'un temps situé dans l'intervalle [26 ps - 850 ps] avec une moyenne égale à 282 ps (Tab. IV.4). Dans les deux simulations restantes on n'observe aucune torsion pour l'une (simulation de 2 ns), et on observe une torsion en τ pour l'autre, se produisant après 200 ps. Les torsions en φ aboutissent à des géométries proches de la perpendiculaire : $\varphi = -69^\circ \pm 6^\circ$ et $\tau = 5^\circ \pm 4^\circ$. Le cycle phénolate du chromophore est alors proche d'une orientation en "T" par rapport au cycle imidazolinone de His193. La torsion en τ déplace le phénolate dans le sens opposé à la triade jusqu'à la géométrie $\varphi = -9^\circ \pm 8^\circ$ et $\tau = -76^\circ \pm 4^\circ$. Les torsions majoritairement observées (φ) ne conduisent pas à l'isomérisation et d'autre part elles se produisent à une échelle de quelques centaines de picosecondes, ce qui n'est pas en accord avec le caractère non-fluorescent de l'isomère *trans* dans Padron.

Dans le cas du schéma 2, nous avons également effectué 30 simulations à l'état excité. Les torsions du chromophore se produisent toutes avant 30 ps, avec une moyenne égale à 4 ps (Tab. IV.4), ce qui est en accord avec le fait que Padron *trans* est très peu fluorescent à 100K. Dans toutes ces simulations les torsions ont lieu suivant τ (déplaçant le phénolate dans le sens opposé à la triade) et peuvent donc conduire à une isomérisation *trans*→*cis*.

La différence entre les résultats de ces deux schémas provient entre autres, vraisemblablement,

de l'attraction électrostatique qui existe entre le phénolate du chromophore et His193 dans le schéma 1 (où His193 est cationique) et qui est absente dans le schéma 2.

Nous étudions maintenant plus précisément les simulations dans le schéma 2.

Dans chaque simulation le chromophore atteint la région de TwI$^-$($\tau = -90^\circ$) par une torsion qui se produit au bout de 4 ps en moyenne (0,2-28 ps). Afin de quitter la région de la géométrie plane, le chromophore effectue une torsion concerté de τ et φ, où $\tau < 0^\circ$ et $\varphi > 0^\circ$. La torsion τ amène ensuite le chromophore à une géométrie proche de la perpendiculaire où $\tau \approx -71 \pm 4^\circ$ et $\varphi \approx 0 \pm 7^\circ$. Cette torsion déplace le phénolate dans le sens opposé de la triade.

Toutes les liaisons H du chromophore sont maintenues pendant la torsion, en particulier, celles du phénolate (Fig. IV.61). Ces points d'attache retiennent le chromophore et l'empêchent de visiter toute la région de TwI$^-$.

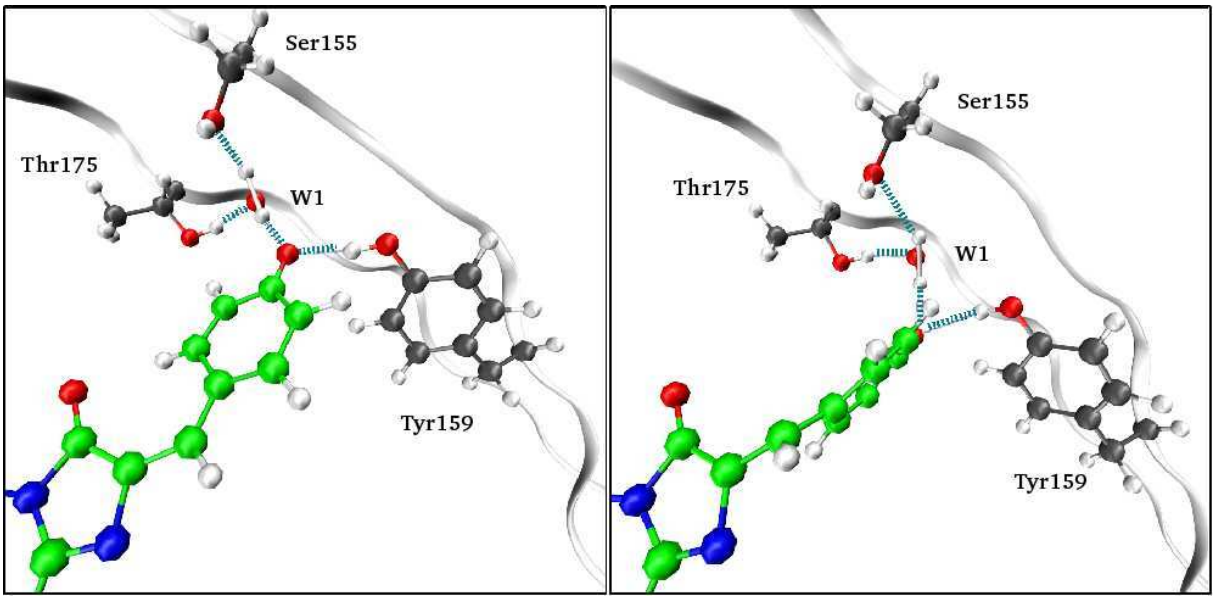


FIG. IV.61.: Le réseau de liaisons H autour de l'oxygène du phénolate du chromophore (en vert). Thr175 et Ser155 sont liés par deux liaisons H peptidiques. A gauche : le chromophore est plan. A droite : le chromophore a effectué une torsion jusqu'à $\tau \approx -70^\circ$.

TAB. IV.4.: Résultats des simulations à l'état excité, avec les schéma de protonation 1 et 2 de la triade. Le temps du premier passage (FPT - *first passage time*) dans une géométrie perpendiculaire est donné en picosecondes. Le FPT moyen est donné en bas du tableau pour chaque schéma.

Schéma 1	
sim	FPT(ps)
1	> 2000
2	692
3	80
4	850
5	640
6	120
7	200
8	434
9	100
10	434
11	720
12	122
13	122
14	74
15	74
16	26
17	68
18	86
19	64
20	66
21	74
22	76
23	74
24	120
25	500
26	768
27	72
28	78
29	710
30	642

Schéma 2	
sim	FPT(ps)
1	1,2
2	0,5
3	0,3
4	1,7
5	0,2
6	0,4
7	1,0
8	1,0
9	0,2
10	10,5
11	8,0
12	14,0
13	6,4
14	0,9
15	28
16	0,4
17	16,0
18	12,0
19	7,7
20	0,8
21	1,1
22	7,0
23	0,5
24	0,4
25	0,4
26	0,4
27	0,4
28	0,4
29	0,3
30	0,5

IV.3.4. Possibilité d'isomérisation

IV.3.4.1. Potentiel interne du chromophore

Le mécanisme de l'isomérisation *trans*→*cis* comporte plusieurs étapes. Après l'excitation du chromophore *trans* il doit d'abord effectuer une torsion en τ qui l'amène à la région de TwI^- ou TwI^+. Le chromophore doit ensuite atteindre la 'géométrie critique' d'isomérisation, c'est-à-dire la géométrie dans laquelle les forces dérivant du potentiel interne du chromophore à l'état fondamental (S_0) tendent à l'amener vers le minimum de la géométrie plane *cis*. L'isomérisation peut donc se produire si le chromophore se trouve dans cette géométrie critique au moment où la conversion interne (CI) $S_1 \rightarrow S_0$ a lieu. Dans le cas inverse, les forces internes vont chercher à le ramener vers le point de départ - le minimum de la géométrie plane *trans*.

Nous rappelons que la torsion le long de l'angle dièdre τ (défini par *C2-CA2-CB2-CG2*) a une période de 360°, ce qui donne lieu à cette relation :

$\tau =$	0°	90°	180°	270°	360°
$=$	-360°	-270°	-180°	-90°	0°
géométrie :	*trans*	TwI^+	*cis*	TwI^-	*trans*

La géométrie critique d'isomérisation dépend de la forme de la surface d'énergie potentielle (PES) de l'état S_0. Toniolo *et al. [73]* ont caractérisé le chemin d'énergie minimum (MEP) entre la géométrie plane *cis* et la géométrie perpendiculaire TwI^-, à l'état S_1 du chromophore anionique dans le vide. Ils ont optimisé la géométrie du chromophore à l'état S_1 au niveau SA-2-CAS(2,2) et ensuite calculé l'énergie de S_0 et S_1 au niveau SA-2-CAS(2,2)-PT2. Le résultat montre un minimum local à la géométrie perpendiculaire sur la PES de l'état S_0 : la géométrie $TwI^-(\tau = -90°)$ est stabilisée de 2,6 kcal/mol par rapport à la géométrie $\tau = -120°$. La description du paysage énergétique couvre uniquement le chemin *cis*→TwI^- mais le MEP du côté de l'isomère *trans* (TwI^-← *trans*) est *a priori* très similaire. Néanmoins, ce MEP S_0 est basé sur des géométries optimisées à l'état S_1 et ne décrit donc pas la PES relaxée de l'état S_0. Il faudrait une description plus précise de la PES de l'état S_0 dans la région des géométries perpendiculaires afin de pouvoir définir une géométrie critique d'isomérisation. Les résultats de Toniolo *et al.* nous indiquent simplement la possibilité d'un minimum local pour TwI^- à l'état S_0.

Nous envisageons donc deux modèles de la PES du chromophore à l'état S_0 : soit elle présente un maximum d'énergie centré sur $\tau \approx -90°$ (Fig. IV.62, à gauche) soit elle présente un minimum local, centré sur $\tau \approx -90°$, avec des barrières de 2-3 kcal/mol autour de $\tau \approx -60°$ et $\tau \approx -120°$(Fig. IV.62, à droite). Dans le premier modèle, le chromophore peut s'isomériser si $\tau < -90°$ au moment de la CI. Dans le deuxième modèle, le chromophore doit avoir atteint $\tau < -120°$ au moment de la CI pour que ses forces internes l'amènent directement vers la géométrie *cis*. Cependant, si la CI se produit quand $-120° < \tau < -60°$, le chromophore sera amené vers le minimum local de l'état S_0, à partir duquel il est possible d'atteindre la géométrie *cis*. Le chromophore doit dans ce cas franchir la barrière à $\tau \approx -120°$ (2-3 kcal/mol). Dans le vide, les deux modèles envisagées de la PES de l'état S_0 donnent une probabilité d'isomérisation de 50% à chaque passage dans une géométrie perpendiculaire ($\tau = \pm 90°$).

Dans nos 30 simulations à 100 K on n'observe aucune torsion du chromophore qui pourrait mener vers une isomérisation. Cependant, cela n'est pas contradictoire avec l'expérience puisque le rendement d'isomérisation mesuré expérimentalement est très faible ($\sim 10^{-6}$) .

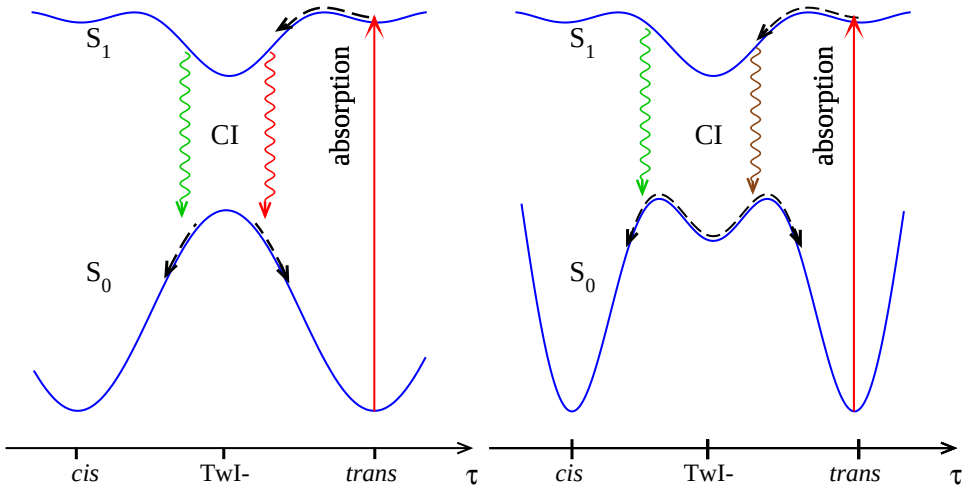


FIG. IV.62.: Le schéma des processus internes au chromophore, impliqués dans son isomérisation. La surface d'énergie potentielle (PES) de l'état excité (S_1) est basée sur des calculs quantiques (section III.4). Nous avons envisagé deux modèles de la PES de l'état fondamental (S_0) : l'un présente un maximum à TwI$^-$ ($\tau = -90^\circ$) (à gauche) et l'autre un minimum local (à droite). Les flèches ondulées vertes représentent une conversion interne (CI) qui amène le chromophore à un point sur la PES S_0 où les forces internes du chromophore vont chercher à l'amener vers l'isomère *cis*. La CI représentée par la flèche ondulée rouge va, au contraire, conduire le chromophore vers l'isomère *trans*. La CI représentée par la flèche ondulée marron conduit le chromophore à un minimum local sur la PES S_0 , à partir duquel il peut atteindre l'isomère *cis* aussi bien que l'isomère *trans*. Ce schéma ne prend en compte aucune influence de la protéine.

IV.3.4.2. Les contraintes de la protéine

Les contraintes imposées par la protéine limitent les torsions du chromophore. Dans les simulations à l'état S_1 de Padron *trans* (avec le schéma de protonation 2 de la triade), le chromophore est bloqué à $\tau \approx -71 \pm 4^\circ$ et $\varphi \approx 0 \pm 7^\circ$, et empêché d'atteindre le minimum TwI$^-$ $\tau = -90^\circ$. Pour réaliser une isomérisation, le chromophore doit atteindre une torsion plus importante, soit à l'état excité (modèle 1 de la PES S_0), soit à l'état fondamental (modèle 2 de la PES S_0).

Si les forces internes du chromophore sont suffisamment fortes pour pousser les résidus qui bloquent, il peut se frayer un chemin jusqu'à l'isomère *cis*. Cela est envisageable si le chromophore atteint la géométrie critique où la pente de la PES S_0 est raide et mène directement vers l'isomère *cis*, i.e. à $\tau \lesssim -90^\circ$ (modèle 1) ou à $\tau \lesssim -120^\circ$ (modèle 2). Cependant, pour atteindre une telle géométrie, les contraintes des résidus voisins doivent être levées.

Nous avons analysé les contraintes de la protéine tout du long du chemin d'isomérisation en utilisant une géométrie du chromophore issue de la simulation à l'état excité qui est proche de la géométrie perpendiculaire ($\tau = -75^\circ$) , et en effectuant une torsion en τ (l'environnement protéique et le reste du chromophore fixe étant fixe). Le calcul d'énergie d'interaction entre le chromophore et la protéine (van der Waals et électrostatique) à partir du point de départ ($\tau = -75^\circ$) jusqu'à l'isomère cis ($\tau = -180^\circ$) montre que les contraintes immédiates qui empêche le chromophore d'atteindre la géométrie critique sont les deux liaisons H de l'oxygène du phénolate avec Tyr159 et W1 et la répulsion stérique de Met59.

Afin d'effectuer une isomérisation au sein de la protéine, il faut donc une rupture de ces liaisons H et un déplacement de la chaîne latérale de Met59.

Dans les simulations à 100 K, les liaisons H entre de phénolate avec Tyr159 et W1 sont toujours présentes, ce qui signifie que leur rupture est rare. Afin d'étudier l'effet de leur rupture sur la torsion du chromophore, nous avons effectué des simulations à plus haute température (300 K). Dans ces simulations on observe effectivement des cas où les liaisons H sont rompues et que, dans ce cas, le chromophore effectue une torsion en τ qui dépasse -90°, avant d'être arrêté par Met59. Dans l'état fondamental à 300 K (où on échantillonne mieux la géométrie plane) les liaisons H sont rompues pendant 2% du temps.

On voit également dans les simulations à 300 K, un déplacement de la chaîne latérale de Met59 quand le chromophore arrive à $\tau < -90^\circ$. Cela peut être dû à un mouvement de la protéine ou aux forces internes du chromophore qui tendent à le ramener au minimum d'énergie ($\tau = -90^\circ$) à l'état excité. Il est probable que le même déplacement se produit à 100 K quand le chromophore effectue une torsion lorsque les liaisons H avec Tyr159 et W1 sont rompues.

Nous pouvons donc conclure que le chromophore peut atteindre une géométrie où $\tau < -90^\circ$ à 100 K mais que cela ne se produit que rarement. Si la conversion interne se produit dans cette géométrie, les forces internes du chromophore vont tendre à le ramener le chromophore à l'isomère *cis*. La torsion du chromophore à l'état fondamental ($\tau = -90^\circ \rightarrow \tau = -180^\circ$) doit provoquer un déplacement supplémentaire de Met59. En effet, les structures cristallines de la forme désactivé de la protéine (chromophore *trans*) et de la forme activé (chromophore cis) [83] montrent un que la chaîne latérale de Met59 s'est déplacé pendant l'isomérisation. Afin d'étudier ce mécanisme il faudrait développer un champ de forces de torsion pour le chromophore à l'état fondamental, optimisé notamment pour la région proche de $\tau = -90^\circ$.

IV.3.5. Conclusion

Nous avons trouvé que seul le schéma 2 de protonation de la triade (Glu211 déprotoné - His193 protoné sur l'azote ε - Glu144 protoné) est compatible avec le fait que Padron *trans* est non-fluorescent et susceptible de s'isomériser suite à une excitation.

Dans nos simulations le chromophore n'atteint jamais les géométries qui peuvent mener à une isomérisation (il ne va pas au delà de la région $\tau \approx -71 \pm 4^\circ$), ce qui n'est pas surprenant car le rendement d'isomérisation est très faible. Nous avons identifié les contraintes qui l'en empêchent. Des simulations biaisées seraient nécessaires pour étudier l'effet de la levée de ces contraintes.

V. La durée de vie non-radiative

Parmi les objectifs de cette thèse se trouve la modélisation de la cinétique de désactivation non-radiative de l'état excité du chromophore. L'ensemble des données expérimentales et théoriques converge pour associer la désactivation non-radiative à la torsion du chromophore autour des liaisons reliant les deux cycles. Il s'agit donc dans le cadre de l'équation cinétique standard (V.1) de déterminer la constante de vitesse non-radiative k_{nr} en considérant uniquement la désactivation non-radiative pouvant se produire dans les géométries perpendiculaires .

$$\frac{\mathrm{d}[S_1]}{\mathrm{d}t} = -k_{nr}[S_1] - k_r[S_1] \tag{V.1}$$

On définit le temps de vie non-radiatif τ_{nr} comme l'inverse de k_{nr}.

Pour cela, nous devons établir un modèle cinétique de la désactivation non-radiative permettant d'extraire τ_{nr} des dynamiques moléculaires classiques de la protéine à l'état excité. Ce modèle s'appuie sur l'utilisation de données de la littérature sur le processus non Born-Oppenheimer et sur le traitement statistique adapté des événements de torsions au cours des dynamiques pour extraire une valeur de τ_{nr} et évaluer l'incertitude associée.

V.1. Caractéristiques de la transition $S_1 \rightarrow S_0$ dans les géométries perpendiculaires

Nous considérons ici l'extinction de fluorescence produite par une torsion du chromophore à l'état excité. Plusieurs phénomènes peuvent conduire à l'extinction de fluorescence quand le chromophore atteint la région des géométries perpendiculaires ($\varphi = 90^\circ$ ou $\tau = 90^\circ$) :

- la probabilité d'une émission de fluorescence est faible car la force d'oscillateur de la transition $S_1 \rightarrow S_0$ est abaissée.

- l'émission de fluorescence, quand elle a lieu, est déplacée dans l'infra rouge (IR), c'est-à-dire hors de la fenêtre de détection, à cause du petit écart énergétique entre S_0 et S_1.

- une conversion interne est susceptible d'avoir lieu car l'écart énergétique entre les deux états électroniques S_0 et S_1 est petit (ca. 1 eV).

V.1.1. La force d'oscillateur $S_1 \rightarrow S_0$

La figure V.1 ci-dessous reporte la valeur du moment de transition $S_1 \rightarrow S_0$ calculé au niveau CASPT2 dans les régions des minima de la surface d'énergie potentielle du chromophore dans l'état S_1. On voit que le moment de transition diminue fortement si au moins l'un des angles de torsion est supérieur à 60° en valeur absolue, et il en est de même pour la force d'oscillateur, proportionnelle au carré du moment de transition,

La différence d'énergie entre les deux états électroniques est de 29 kcal/mol lorsque φ vaut 90° (Fig. V.2) et de 17 kcal/mol lorsque τ vaut 90°, ce qui laisse supposer une transition non-radiative de type conversion interne plutôt que de type intersection conique.

On peut donc considérer que la transition radiative disparaît lorsque la torsion φ ou τ dépasse 60° et que la disparition de l'état excité dans les géométries perpendiculaires est due à un processus de conversion interne.

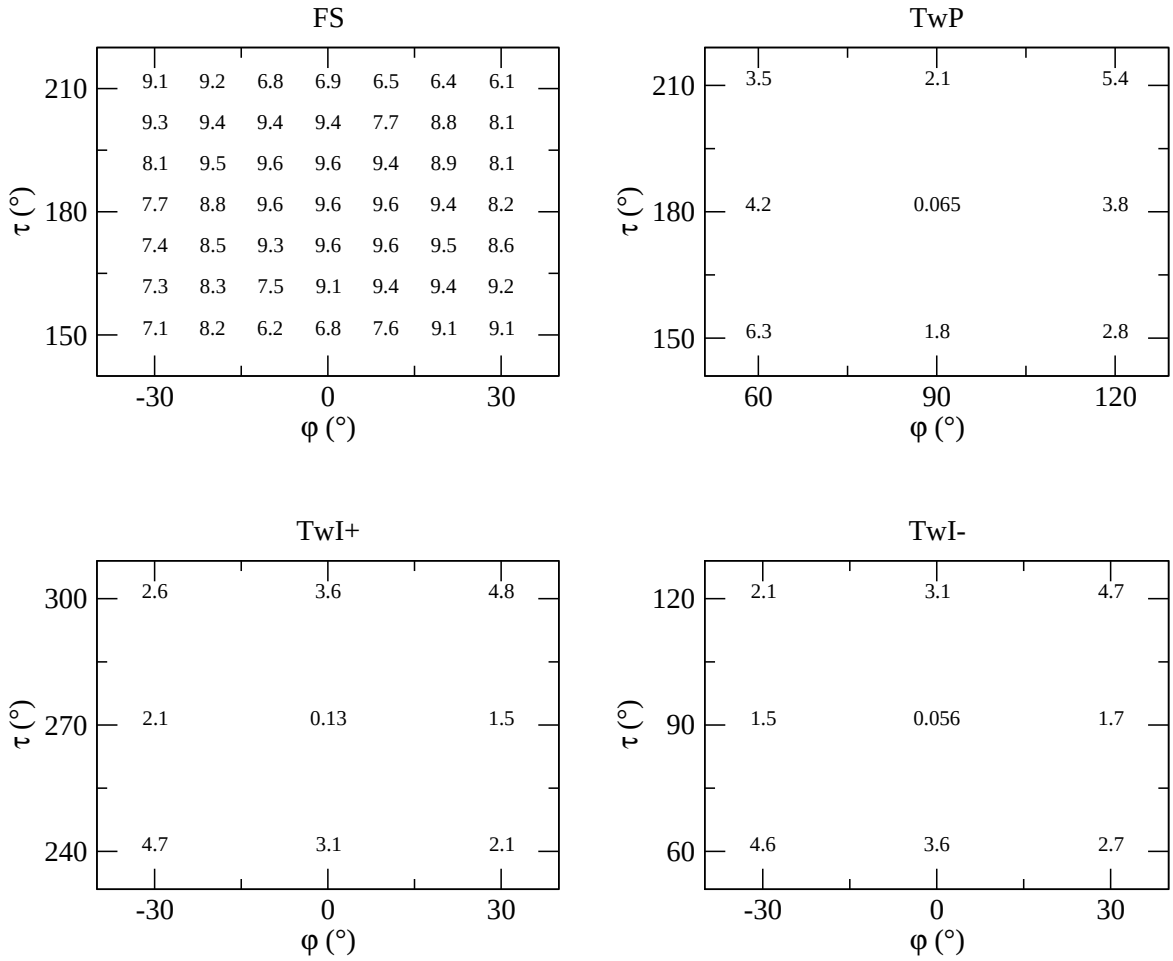


FIG. V.1.: Le moment dipolaire de la transition $S_1 \rightarrow S_0$ (en Debye) dans les régions des minima de la surface d'énergie potentielle du chromophore. La région de TwP représente les régions TwP^+ et TwP^-qui sont identiques.

V.1.2. Données sur la conversion interne $S_1 \rightarrow S_0$ du chromophore

Il n'existe pas de données expérimentales de la désactivation radiative ou non-radiative du chromophore dans le vide. Des études de dynamique *ab initio* (multiple spawning QM/MM) du chromophore anionique (Olsen *et al* [84]) montrent que le déclin de la population de l'état S_1 est d'environ 2% après 0,25 ps de dynamique dans la région des géométries perpendiculaires (cette région est atteinte, en moyenne, en 0,20 ps à partir de la géométrie plane S_1). Aucun retour à la géométrie plane S_1 n'est observé après la torsion du chromophore et aucune géométrie de

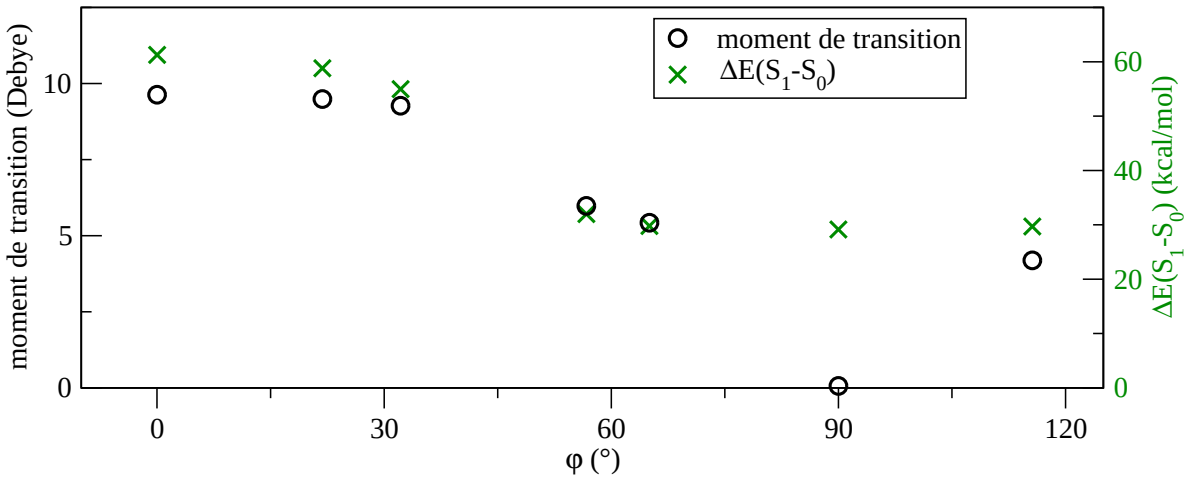


FIG. V.2.: Le moment dipolaire de la transition $S_1 \rightarrow S_0$ (en Debye) et l'écart énergétique entre les deux états (en kcal/mol) le long le coordonnée φ. Le tout est calculé au niveau CAS(4,3)-PT2 sur des géométries optimisées pour l'état S_1 au niveau CASSCF(4,3)/6-31G*.

type hula-twist ($\tau = 90^\circ$ et $\varphi = 90^\circ$) n'est atteinte. Le déclin de la population de l'état S_1 est donc entièrement dû à la conversion interne dans les géométries perpendiculaires ($\tau = 90^\circ$ ou $\varphi = 90^\circ$). Si on décrit ce déclin par l'exponentielle suivante :

$$\frac{A(t)}{A_0} = exp(-\frac{t}{\tau}) \tag{V.2}$$

où $A(t)$ représente la population S_1 à l'instant t et A_0 à l'instant $t = 0$, on en déduit une durée de vie τ non-radiative égale à 12 ps. Les résultats de Olsen *et al.* sont basés sur un petit nombre de simulations (15 trajectoires). Ce calcul de la durée de vie est très sensible, e.g. un déclin légèrement différent, 4% au lieu de 2%, correspond à une durée de vie τ de 6 ps - la moitié de la valeur obtenue auparavant. On doit donc considérer le résultat comme un ordre de grandeur plutôt que comme une valeur exacte.

Par ailleurs, l'environnement est susceptible d'avoir une influence significative sur le processus de conversion interne. Des études similaires sur le chromophore neutre (protonation du phénolate) montre que l'effet d'un solvant fortement polaire comme l'eau peut accélérer la conversion interne : un déclin de 50% de la population de l'état S_1 est observé après environ 20 fs dans l'eau et après environ 600 fs dans le vide [85]. L'accélération de l'extinction de la fluorescence en présence de l'eau est interprétée par la stabilisation des géométries correspondant à une éventuelle intersection conique, qui sont dans le vide situées 5 kcal/mol plus haut que la géométrie plane ([26]). La protéine dans son ensemble et l'environnement proche du chromophore constituent un milieu assez rigide, dont on attend des effets beaucoup plus faibles sur la PES de S_1 que ceux de l'eau) et nous avons considéré que la valeur de 12 ps pour le temps de vie non-radiatif du chromophore dans une géométrie perpendiculaire était une approximation acceptable.

V.2. Principes de l'analyse des dynamiques moléculaires de la protéine à l'état excité

Les dynamiques moléculaires que nous avons menées donnent accès aux temps de résidence du chromophore dans les géométries planes et perpendiculaires. Le chromophore contraint par la protéine reste au voisinage de la géométrie plane pendant des temps variant de plusieurs picosecondes à plusieurs nanosecondes. Lorsqu'il atteint des géométries perpendiculaires, il peut y rester jusqu'à la fin de la simulation ou y résider un temps variant de moins de 2 picosecondes à plusieurs centaines de picosecondes avant de revenir à la géométrie plane. Ce retour est une manifestation des forces exercées par la protéine sur le chromophore pour le maintenir dans une géométrie plane et de la nécessaire coïncidence entre plusieurs réarrangements de résidus proches du chromophore pour autoriser le maintien d'une géométrie perpendiculaire.

Nos simulations classiques ne décrivent pas le processus de changement d'état électronique et de retour à l'état fondamental. Compte tenu des résultats sur le chromophore isolé exposés précédemment, nous considérerons qu'une torsion est efficace (c'est-à-dire qu'elle conduit à une désexcitation non-radiative) si elle conduit à un angle τ ou φ supérieur à 60° et si elle dure au moins 12 ps. Nous appellerons ces situations des 'géométries critiques' (cette expression compacte englobant la notion de temps de résidence au moins égal à 12 ps). Dans, l'étude développée ci-dessous, nous faisons l'approximation que la probabilité de la transition non-radiative $S_1 \rightarrow S_0$ vaut 1 dans des géométries critiques et 0 en dehors.

Pour aller au delà, une analyse approfondie du processus non Born-Oppenheimer est nécessaire, notamment l'évolution dans le temps de la superposition des états électroniques S_1 et S_0 dans les géométries perpendiculaires. Les travaux de Olsen *et al.* [84] doivent permettre d'obtenir les informations nécessaires, cependant nous n'avons pas cherché à le faire dans le cadre de ce travail de thèse.

Atteindre une géométrie critique se produit un nombre limité de fois par simulation pour une protéine donnée, ce qui signifie que la cinétique de la désactivation est lente à l'échelle des simulations, ou encore que l'événement correspondant à atteindre une géométrie critique est rare. Une méthode habituelle qui permet de répondre à cette question consiste à biaiser la simulation, par exemple à l'aide du 'umbrella sampling' [111] ou de la 'meta dynamics' [112] ou à l'aide des divers techniques utilisant 'the reactive flux auto correlation function' (une utilisation pour déterminer la vitesse d'isomérisation du n-butane dans la référence [113]) ou la théorie de la 'blue moon' [114, 115]. Cependant ces méthodes impliquent d'avoir, à l'avance, une idée du chemin de la réaction et/ou de la géométrie de l'état de transition ce qui n'est pas le cas ici. Le 'parallel tempering' [116] permet, lui, une bonne évaluation de probabilités très faibles, mais ne donne pas d'information directe sur l'aspect temporel.

On pourrait envisager de combiner les deux types d'approches : le 'parallel tempering' pour essayer de caractériser l'état de transition puis une méthode de biais. Mais il se trouve qu'on peut obtenir des informations fiables de façon beaucoup plus simple, même en n'ayant observé que peu d'évènements au cours de la simulation et même en ignorant tout du chemin de réaction. Pour cela on va utiliser la notion de 'temps de premier passage' (FPT *first passage time*) [117] et, en particulier, le fait que l'on connaît à l'avance la forme de la distribution des FPT. Au

cours d'un trajectoire, nous définissons le temps de premier passage comme le temps qui s'écoule entre le début de la simulation et le moment où le chromophore atteint une 'géométrie critique', i.e. la situation où il reste pendant au moins 12 ps dans une géométrie perpendiculaire.

V.3. Temps moyen de premier passage et temps de vie non-radiatif

On considère un processus assimilable à une réaction chimique et on rappel la relation entre distribution des temps de premier passage (FPT) et profil du déclin du réactif quand on peut se placer dans le cadre des hypothèses suivantes :

- le processus peut être décrit par le passage à une géométrie critique G_c (pas nécessairement connue à l'avance)

- la probabilité P de passage à cette géométrie est indépendante du temps

- il y a réaction à chaque passage (comme on suppose dans la théorie de l'état de transition, TST) à la géométrie critique.

La probabilité que la réaction se produise dans un intervalle de temps Δ est donnée par $\Gamma\triangle$ où Γ désigne la densité de probabilité par unité de temps. On suppose que Γ est indépendant du temps. La probabilité d'observer k passages dans un intervalle de temps Δ est alors donnée par la distribution de Poisson :

$$P(k) = \frac{(\Gamma\Delta)^k e^{-\Gamma\Delta}}{k!} \tag{V.3}$$

Dans notre cas, on s'intéresse à la probabilité P_{FP} d'observer un premier passage (FP *first passage*) dans l'intervalle de temps $[t, t + \mathrm{d}t]$ où t et $\mathrm{d}t$ sont donnés, soit zéro passages (k=0) dans $[0, t]$ et un seul passage (k=1) dans $[t, t + \mathrm{d}t]$. On suppose que ces deux probabilités sont indépendantes, la probabilité cherchée est donc le produit de deux facteurs :

$$P_{FP} = P(0) \times P(1)$$

En utilisant l'équation V.3 on obtient :

$$\begin{aligned} P(0) &= e^{-\Gamma t} \\ P(1) &= \Gamma e^{-\Gamma \mathrm{d}t}\mathrm{d}t \end{aligned} \tag{V.4}$$

On en déduit la densité de probabilité d'un premier passage dans l'intervalle $[t, t + \mathrm{d}t]$en divisant par $\mathrm{d}t$ puis en faisant tendre $\mathrm{d}t$ vers zéro :

$$\rho(t) = \lim_{\mathrm{d}t\to 0} \frac{P_{FP}}{\mathrm{d}t} = \Gamma e^{-\Gamma t}$$

La temps moyen du premier passage (MFPT *mean first passage time*), est ainsi donné par

$$\tau = \int_0^\infty t\Gamma e^{-\Gamma t}\mathrm{d}t = \frac{1}{\Gamma} \tag{V.5}$$

d'où

$$\rho(t) = \frac{1}{\tau}\, e^{-t/\tau} \tag{V.6}$$

On considère alors le déclin de population du réactif. On regarde d'abord la simulation d'un réactif que l'on répète plusieurs fois. Si dn est le nombre moyen de réactions qui se produisent pendant l'intervalle de temps dt, alors, dn est proportionnel à dt : dn = λdt. Si l'intervalle dt considéré est égal au MFPT, ou le temps de vie, τ, alors, par définition dn=1. On trouve donc la relation dn =dt/τ. On considère maintenant toute la population des réactifs et définit $N(t)$ comme le nombre de systèmes n'ayant pas encore réagi au temps t. Le nombre moyen de réactions qui se produisent pendant l'intervalle de temps dt devient donc

$$\mathrm{d}N(t) = \frac{N(t)}{\tau}\mathrm{d}t \tag{V.7}$$

Le déclin de population des réactifs peut alors être écrite par une exponentielle :

$$N(t) = N(0)\exp\left(-\frac{t}{\tau}\right) \tag{V.8}$$

La distribution des FPT et le déclin de la population de réactifs sont décrits par *la même* fonction exponentielle, d'où

$$k_{reac} = \frac{1}{\tau} \tag{V.9}$$

Le fait que la distribution des FPT est nécessairement exponentielle, eq.(V.6), et les trois hypothèses générales sur le processus impliquent une relation simple entre la constante de vitesse de la réaction et le MFPT, eq.(V.9).

Ces résultats permettent d'illustrer la difficulté à prédire la valeur de k_{reac} par simulation dans le cas d'un évènement rare. Cela provient simplement de prédire la valeur du MFPT. On montre sur la figure V.3 les deux situations correspondant à un temps de simulation T très supérieur ou du même ordre de grandeur que le MFPT τ : à gauche on a $T \gg \tau$ (T =6, τ = 1 dans les unités de la figure) - ce qu'on peut réaliser en pratique seulement si τ est assez petit - et on voit que la distribution des FPT est presque entièrement contenue dans le temps T de simulation ; au contraire, à droite, on a $T \lesssim \tau$ ($T = \tau$=6) - ce à quoi on est contraint si τ est plus grand - et la distribution des FPT est contenue seulement en partie dans le temps de simulation.

Donc si on observe n passages aux temps à la géométrie critique au cours de la simulation et que l'on évalue la moyenne

$$\bar{t}_n = \frac{1}{n}\sum_{i=1,n} t_i \tag{V.10}$$

où les $t_i (1 \leq i \leq n)$ sont FPT observés, on doit s'attendre à avoir

$\bar{t}_n = \tau$ si $\tau \ll T$ (figure V.3, à gauche) ;

$\bar{t}_n < \tau$ sinon (figure V.3, à droite).

La question posée ici est de déduire comment déduire une valeur fiable de τ à partir des FPT observés dans le second cas, c'est-à-dire pour un évènement rare.

On note par ailleurs que chaque passage à la géométrie critique ne conduit pas nécessairement à une réaction (il arrive que la porte de l'enzyme se referme avant la diffusion du substrat, que la différence entre les potentiels d'ionisation cesse d'être inférieure au couplage avant que le

transfert d'électron puisse se faire, *etc.* ...) : c'est un limite de la TST ; dans une simulation, on peut en tenir compte en ne considérant que les cas où le système reste assez longtemps dans la situation *ad hoc* (≈ 12 ps avec $\varphi \approx 90^o$ dans le cas de la fluorescence de la GFP).

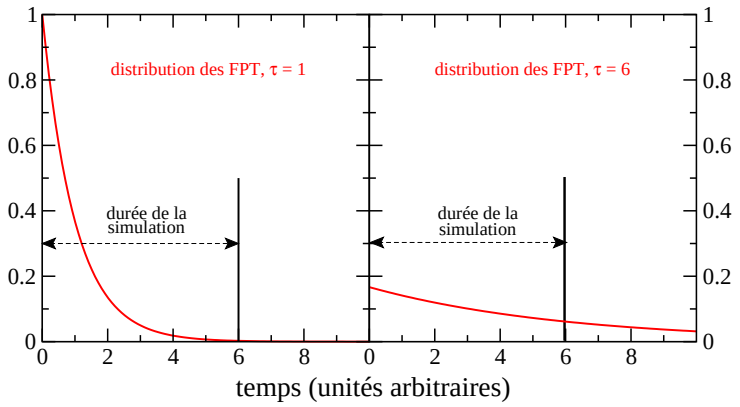


FIG. V.3.: La distribution de FPT dans deux situations : une durée de la simulation très supérieure (à gauche) ou du même ordre de grandeur que le MFPT (à droite).

V.3.1. Évaluation directe du MFPT par simulation

Une première approche pour évaluer un MFPT par simulation consiste simplement à faire la moyenne $\bar{t}_n$ des FPT t_i observés au cours de N simulations, *cf.* l'eq.(V.10). Mais on doit alors noter, voir la figure V.3 à droite, que $\bar{t}_n$ peut être considérée comme une approximation du MFPT de la distribution $\rho(t)$ seulement si le temps de simulation T est très supérieur au MFPT. Sinon, $\bar{t}_n$ est plutôt une approximation du MFPT d'une distribution partielle $\rho(t \in [0;T])$, limitée par la durée T des simulations (en supposant qu'elles ont toutes la même durée). En utilisant l'eq.(V.6) on obtient une expression de cette MFPT partiel :

$$\tau_{part}(\tau, T) = \frac{\int\limits_{0,T} \mathrm{d}t\, t\, \rho(t)}{\int\limits_{0,T} \mathrm{d}t\, \rho(t)} = \frac{\tau - (T+\tau)\exp\left(-\dfrac{T}{\tau}\right)}{1 - \exp\left(-\dfrac{T}{\tau}\right)} \tag{V.11}$$

où τ désigne le MFPT de la distribution complète de FPT $\rho(t)$.

En posant $\bar{t}_n \approx \tau_{part}$ on peut alors obtenir une prédiction τ_{pred} de τ en résolvant l'équation V.11(où $\bar{t}_n$ et T sont connus) :

$$\bar{t}_n = \frac{\tau_{pred} - (T+\tau_{pred})\exp\left(-\dfrac{T}{\tau_{pred}}\right)}{1 - \exp\left(-\dfrac{T}{\tau_{pred}}\right)} \tag{V.12}$$

Remarques :

- on peut vérifier que cette racine τ_{pred} n'existe pas nécessairement : il faut pour cela qu'on ait $0 \leq \bar{t}_n \leq T/2$ (de fait $\bar{t}_n$ varie seulement de 0 à $T/2$ quand τ_{pred} prend toutes les valeurs possibles de 0 à l'∞) ; on doit donc sélectionner parmi les simulations disponibles un sous ensemble tel que $\bar{t}_n \leq T/2$;

- quand elle existe et sauf cas exceptionnel, la racine τ_{pred} est différente du τ cherché.

Afin de se faire une idée de la différence entre τ_{pred} et τ on cherche la distribution des τ_{pred} obtenus en considérant les t_i comme les tirages d'une variable aléatoire t de la distribution partielle

$$\begin{aligned} \rho(t \in [0;T]) &= \frac{1}{\tau} \frac{\exp\left(-\frac{t}{\tau}\right)}{1-\exp\left(-\frac{T}{\tau}\right)} \\ 0 \leq \quad t \quad &\leq T \end{aligned} \tag{V.13}$$

et $\bar{t}_n$ comme la moyenne de n tirages t_i de t avec la contrainte que $\bar{t}_n$ doit être dans $[0,\, T/2]$. On trouve en fait que la distribution de la moyenne $\bar{t}_n$ contrainte à l'intervalle $[0,\, T/2]$ tend vers le même distribution normale p_n que la moyenne non contrainte sauf que cette distribution doit être tronquée et donc renormée.

On peut ainsi exprimer la distribution g_n des prédictions τ_{pred} à l'aide des fonctions p_n, et $\bar{t}_n(\tau_{pred})$, eq.(V.12) :

$$g_n(\tau_{pred}) \quad \propto \quad p_n(\bar{t}_n) \left| \frac{\mathrm{d}\bar{t}_n(\tau_{pred})}{\mathrm{d}\tau_{pred}} \right| \tag{V.14}$$

Le calcul de g_n consiste donc dans les étapes suivantes
- on parcourt les valeurs de τ_{pred} ;
- pour chaque valeur de τ_{pred} on détermine la fonction $\bar{t}_n(\tau_{pred})$, eq.(V.12), et de sa dérivée ;
- puis on détermine la distribution $p_n(\bar{t}_n)$;
- enfin on obtient $g_n(\tau_{pred})$ à partir de l'eq.(V.14)

Remarque : la distribution g_n obtenue ainsi dépend du MFPT τ (supposé exact) utilisé pour évaluer la distribution p_n. Elle dépend de la même manière de la durée totale T de chaque simulation. Par contre elle ne dépend pas du nombre N de simulations effectivement réalisables (*cf.* le § suivant).

Des exemples de la valeur moyenne et de l'écart type de g_n obtenus à partir de l'eq.(V.14) sont montrés sur la figure V.4 à gauche.

On voit que la moyenne des prédictions est toujours raisonnablement précise, l'écart relatif $(\tau_{pred} - \tau)/T$ étant toujours inférieur à 1/2 (en valeur absolue) sauf pour les toutes petites valeurs de n ($n < 9$ dans le cas de la figure). Mais pour savoir si une simulation donnée peut conduire à un résultat proche de cette moyenne il est plus significatif de considérer l'éc

art type de la distribution des prédictions. On trouve alors que, pour un évènement fréquent ($\tau = T/2$), on obtient effectivement un écart type satisfaisant, de l'ordre de $T/2$, avec un nombre

limité de FPT observés ($n \simeq 20$). Par contre, pour un évènement plus rare ($\tau \simeq T$ ou $\simeq 2T$) on trouve qu'on ne peut obtenir une précision du même ordre que si le nombre de FPT observés est beaucoup plus grand ($n \gg 200$) ce qui semble difficile à réaliser.

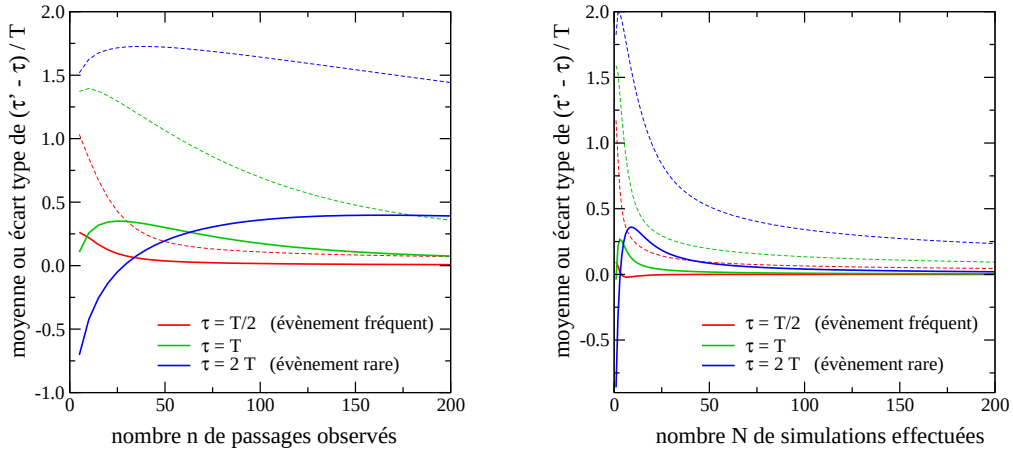


FIG. V.4.: Incertitude sur la prédiction τ_{pred} du MFPT en fonction de la durée T des simulations et du MFPT τ (supposé exact) utilisé pour engendrer les FPT.
A gauche : valeurs du MFPT évaluées à l'aide de la moyenne des n FPT observés au cours des simulations (§ V.3.1) ;
A droite : valeurs du MFPT évaluées à l'aide du nombre de premiers passages observés au cours de N simulations (§ V.3.2) ;
Traits pleins : valeurs moyennes de $(\tau_{pred} - \tau)/T$;
Tirets : écarts types de $(\tau_{pred} - \tau)/T$.

V.3.2. Évaluation du MFPT à partir du nombre de FPT obtenus par simulation

On peut obtenir une évaluation du MFPT sans passer par le calcul de $\bar{t}_n$, eq.(V.10), et sans la résolution de l'eq.(V.12) : si on a fait N simulations de longueur T et qu'il y a un passage à la géométrie critique seulement dans n cas (aucun passage dans $N - n$ cas) cela signifie que l'intégrale de 0 à T de la distribution des FPT est égale à n/N. En utilisant l'eq.(V.6) on obtient

$$\begin{aligned} \frac{1}{\tau} \int_{0,T} \exp\left(-\frac{t}{\tau}\right) \mathrm{d}t &= \frac{n}{N} \\ 1 - \exp\left(-\frac{T}{\tau}\right) &= \frac{n}{N} \end{aligned} \tag{V.15}$$

d'où la prédiction τ_{pred} du MFPT

$$\tau_{pred} = \frac{T}{\ln \dfrac{N}{N-n}} \tag{V.16}$$

Cette relation ne peut être utilisée que si $0 < n < N$.

Pour évaluer l'incertitude statistique de τ_{pred}, on considère, comme précédemment, les FPT

comme des variables aléatoires et on en déduit la distribution q des n obtenus au cours des M séries de N simulations ; en fait q est la 'distribution binomiale'

$$\begin{aligned} q(n) &= \binom{n}{N} z^n (1-z)^{N-n} \\ z &= 1 - \exp\left(-\frac{T}{\tau}\right) \end{aligned} \tag{V.17}$$

La distribution h des valeurs de τ_{pred} est ensuite donnée par

$$\begin{aligned} h\left(\frac{\tau_{pred}}{T}\right) &= q(n) \left| \frac{\partial n}{\partial \left(\tau_{pred}/T\right)} \right| \\ &= q(n) \left| -N \left(\frac{T}{\tau_{pred}}\right)^2 \exp\left(-\frac{\mathrm{T}}{\tau_{\mathrm{pred}}}\right) \right| \end{aligned}$$

où $q(n)$ est évaluée avec la valeur de τ supposée exacte et la dérivé de n est déduit de l'eq. V.15 avec $\tau = \tau_{pred}$:

$$n = N \left[1 - \exp\left(-\frac{T}{\tau_{pred}}\right)\right]$$

Le calcul numérique de la distribution h implique une difficulté venant du fait que $q(n)$ est une distribution discontinue. On doit donc interpoler la distribution $q(n)$ en remplaçant les coefficients du binôme par une combinaison de fonctions gamma. Des exemples de distributions obtenues de cette façon sont donnés sur la figure V.4 à droite. On constate que l'écart relatif $(\tau_{pred} - \tau)/T$ est plutôt plus petit que celui qu'on obtient à l'aide de la méthode du § V.3.1 et, surtout, que l'écart type (relatif à T), σ/T, est beaucoup plus petit ; ainsi, dans le cas $\tau = 2T$ (évènement rare), on obtient une valeur de σ/T de l'ordre de 1/2 avec environ 50 simulations alors qu'avec la méthode du § V.3.1 il faut plus de 200 premiers passages observés ce qui implique un nombre de simulations $\gg 200$ donc bien supérieur à 50.

A l'aide des 50 et 30 simulations réalisées pour les protéines GFP et YFP, nous utilisons l'évaluation du MFPT à partir du nombre de FPT à la géométrie critique.

V.4. Détermination de τ_{nr} de la GFP

V.4.1. Torsion du chromophore

Dans la dynamique de la GFP avec le chromophore à l'état excité (300 ns), décrite dans la section IV.1, on observe des torsions vers les régions TwP^-, TwI^+ et TwP^+. On compte 1036 passages à la géométrie critique ($\varphi < -60^\circ$) dans la région TwP^-, 68 passages dans la région de TwI^+ et 31 dans la région TwP^+. Le temps pendant lequel le chromophore reste dans ces zones varie fortement d'un passage à l'autre et d'une zone à l'autre (Fig. V.5 et V.6).

Nous avons utilisé les critères suivants pour définir le temps du premier passage (FPT) dans chaque simulation : le chromophore doit rester dans une région où $|\varphi|>60^\circ$ ou $|\tau|>60^\circ$ (géométrie critique) au moins 12 ps. Nous avons observé un tel passage dans 25 des 50 simulations de 6 ns

que nous avons effectuées. Dans six cas le premier passage se produit dans la région de TwI^+ et dans les autres il se produit dans la région TwP^-.

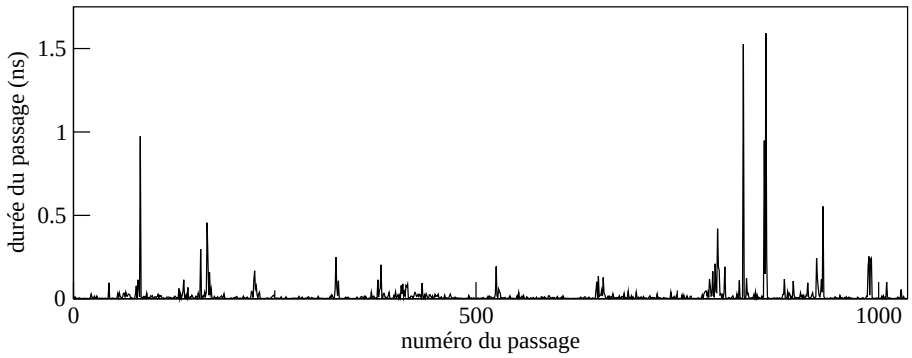


FIG. V.5.: La durée des 1036 passages dans la région TwP^-au cours du temps de simulation (300 ns).

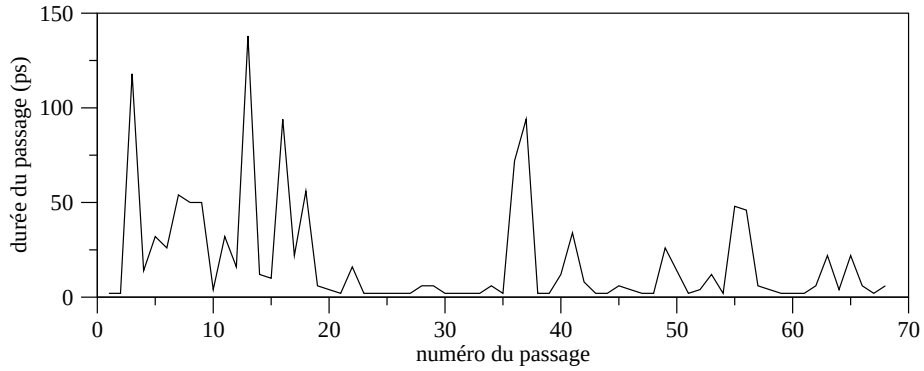


FIG. V.6.: La durée des 68 passages dans la région TwI^+ au cours du temps de simulation (300 ns).

V.4.2. Calcul de τ_{nr}

Les FPT sont présentés dans le tableau V.1.

Si on applique la méthode décrite dans la section V.3.2 (eq. V.16) on obtient l'évaluation $\tau_{nr} = 8,6 \pm 2,6\,ns$ (écart type déduit du graphique dans la figure V.4, c'est à dire, l'intervalle de confiance à 63%), ce qui est en très bon accord avec la valeur de $8,3\,ns$ déduite des données expérimentales [23, 75] et présente une incertitude raisonnable.

Si on applique la méthode décrite dans la section V.3.1 (eq. V.12) on obtient une évaluation de $\tau_{nr} = 7,4 \pm 8,4\,ns$ ($\overline{t_n} = 2,6\,ns$). L'incertitude (déduite du graphique dans la figure V.4) est du même ordre de grandeur que la valeur de τ_{nr} , ce qui fait que cette méthode n'est pas ici utilisable.

TAB. V.1.: Les temps de premier passage (FPT) à une géométrie perpendiculaire dans les simulations de la GFP à l'état excité. Un temps >6 ns signifie que le chromophore est resté plan pendant les 6 ns de simulation.

no	FPT (ns)	no	FPT (ns)
1	>6	26	1,6
2	>6	27	>6
3	>6	28	0,3
4	4,1	29	0,5
5	>6	30	3,2
6	>6	31	0,2
7	4,9	32	3,7
8	>6	33	>6
9	>6	34	0,4
10	>6	35	2,2
11	0,3	36	2,2
12	>6	37	2,3
13	>6	38	2,9
14	>6	39	3,7
15	0,0	40	5,2
16	>6	41	>6
17	>6	42	>6
18	1,1	43	5,6
19	4,9	44	4,5
20	>6	45	>6
21	>6	46	4,9
22	>6	47	>6
23	>6	48	>6
24	>6	49	4,0
25	0,0	50	2,3

V.5. Détermination de τ_{nr} de la YFP

V.5.1. Torsion du chromophore

Dans la dynamique de la YFP avec le chromophore à l'état excité (120 ns), décrite dans la section IV.2, on observe des torsions vers les régions TwI^+ et TwP^+. On compte 301 passages à la géométrie critique ($\varphi < -60^\circ$) dans la région TwI^+ et 413 passages dans la région TwP^+. Le temps pendant lequel le chromophore reste dans ces zones varie, comme dans la GFP, fortement d'un passage à l'autre et d'une zone à l'autre (Fig. V.7 et V.8).

Nous avons utilisé les même critères que pour la GFP pour définir le temps du premier passage (FPT) dans chaque simulation. On observe un tel passage dans 27 des 30 simulations de 8 ns

maximale que nous avons effectuées. A chaque fois le premier passage se produit dans la région de TwI^+.

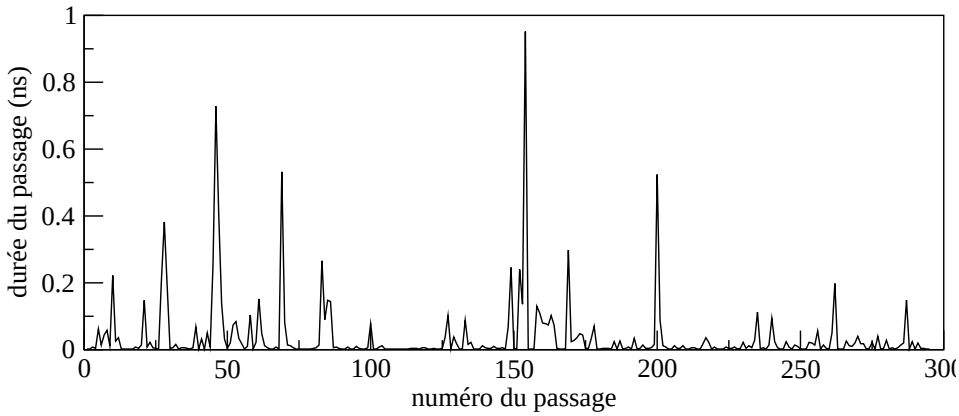


FIG. V.7.: La durée des 301 passages dans la région TwI^+au cours du temps de simulation (120 ns). Le chromophore y reste en moyenne 40 ps avec des durées du passage comprises entre 2 ps et 952 ps.

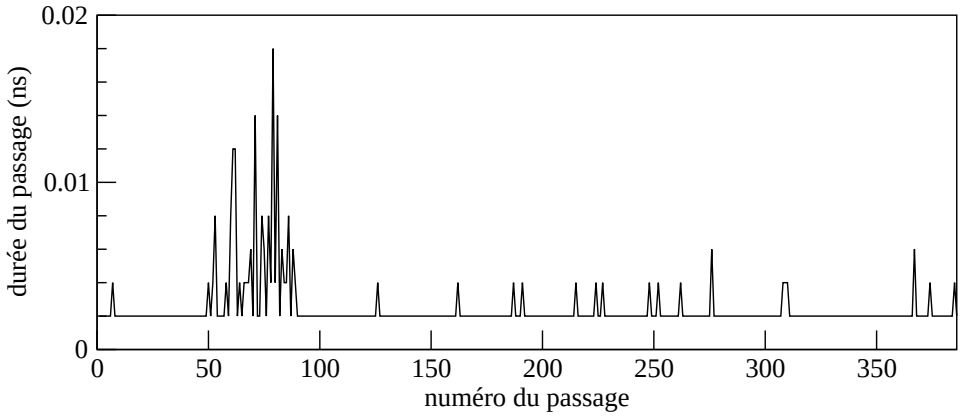


FIG. V.8.: La durée des 413 passages dans la région TwP^+ au cours du temps de simulation (120 ns). Les passages sont très brefs : une durée maximale de 18 ps est observée mais la durée moyenne est proche de 2 ps (la précision étant limité par le pas d'écriture de 2 ps)

V.5.2. Calcul de τ_{nr}

Les FPT sont présentés, dans le tableau V.2.

Si on applique la méthode décrite dans la section V.3.2 (eq. V.16) on obtient $\tau_{nr} = 3,5 \pm 0,2\, ns$ (écart type déduit du graphique dans la figure V.4, c'est à dire, l'intervalle de confiance à 63%).

L'incertitude sur cette prédiction est très faible mais l'intervalle de confiance ne recouvre pas la valeur 9,5 ns, déduite des données expérimentales [78, 1].

La méthode décrite dans la section V.3.1 (eq. V.12) donne, comme pour la GFP, une prédiction très imprécise : $\tau_{nr} = 1,54 \pm 9,6\,ns$ ($\overline{t_n} = 1,50\,ns$). L'intervalle de confiance englobe la valeur expérimentale mais il est trop grand pour que l'on puisse en tirer une autre conclusion que simplement l'accord avec l'ordre de grandeur expérimental.

Par contre, la méthode d'évaluation décrite dans la section V.3.2 qui conduit à une prédiction de $\tau_{nr} = 3,5 \pm 0,2\,ns$ permet d'aboutir à une conclusion plus précise : d'une part le décalage entre la prédiction et la valeur expérimentale est dû à une mauvaise description du processus de désactivation non-radiative, plutôt qu'au nombre de simulations effectuées.

Nous avons envisagé deux pistes pour cela.

Premièrement, dans la YFP, les contraintes de la protéine empêchent le chromophore d'atteindre le minimum énergétique à $\tau = 90°$, $\varphi = 0°$ et il oscille autour de $\tau = 84° \pm 7°$, $\varphi = -21° \pm 10°$. On peut supposer que la conversion interne est moins efficace dans cette géométrie qu'au minimum du puits. Il faudra donc prendre en compte ce facteur dans la description de la désactivation non-radiative.

Deuxièmement, le champ de forces de Amber ne décrit pas parfaitement le π-stacking entre le phénolate du chromophore et le cycle phénol de Tyr203 au cours des simulations effectuées (voir section IV.2.2.1). Une interaction plus attractive entre le deux cycles peut favoriser la géométrie plane du chromophore. Il faudra donc développer un champ de force spécifique à cette interaction avec le chromophore à l'état fondamental ainsi qu'avec le chromophore à l'état excité .

TAB. V.2.: Les temps de premier passage (FPT) à une géométrie perpendiculaire des simulations de la YFP à l'état excité. Un temps >8 ns signifie que le chromophore est resté plan pendant les 8 ns de simulation.

no	**FPT (ns)**	**no**	**FPT (ns)**	**no**	**FPT (ns)**
1	1,0	11	5,4	21	>8
2	2,6	12	2,7	22	0,6
3	>8	13	1,7	23	0,01
4	0,004	14	0,004	24	1,0
5	0,2	15	3,5	25	1,9
6	0,1	16	>8	26	2,2
7	0,1	17	1,3	27	0,5
8	2,2	18	2,4	28	2,8
9	2,0	19	1,3	29	0,9
10	1,5	20	0,9	30	0,2

V.6. Conclusion du chapitre

Nous avons établi deux modèles qui permettent d'extraire une évaluation de τ_{nr} des simulations effectuées de la GFP et la YFP, et d'estimer l'incertitude sur ces prédictions. Ces modèles s'appuient sur une description simple du processus de désactivation non-radiative : le chromophore se désactive s'il reste dans une géométrie perpendiculaire pendant une durée 'critique'. La définition de cette durée est basée sur des dynamiques *ab initio* du chromophore dans le vide [84]. Une étude plus approfondie sur le processus non Born-Oppenheimer dans les géométries perpendiculaires permettrait d'améliorer la description utilisée dans nos deux méthodes.

VI. Conclusion générale et perspectives

De nombreuses études expérimentales et théoriques ont été menées ces dix dernières années sur les protéines fluorescentes. De là, se forge une compréhension essentiellement qualitative du rôle de la protéine vis-à-vis de l'obtention ou non d'une émission radiative : il apparaît que la protéine permet la fluorescence en bloquant les processus qui la désactivent ; ces processus de désactivation sont très rapides et efficaces (à l'échelle de la picoseconde) dans le cas du chromophore seul, et ils sont bien identifiés comme étant des torsions autour des liaisons intercycles. Dans la protéine, la sensibilité des temps de vie de fluorescence à des mutations proches ou non du chromophore, à des modifications de pH ou de température laisse supposer un contrôle de la dynamique du chromophore par différents paramètres, sans qu'ils soient pour autant identifiés et mis en relation.

Deux difficultés théoriques ont principalement été affrontées dans ce travail. Il s'agit tout d'abord de la modélisation de la surface d'énergie potentielle de torsion du chromophore à l'état excité basée sur des calculs quantiques de haute précision, sur une interpolation des valeurs obtenues par une expression analytique appropriée en fonction des angles de torsion et avec une précision suffisante pour reproduire des barrières de l'ordre de la kcal/mole, et enfin, sur l'implémentation de cette expression analytique dans le programme parallèle AMBER. Le deuxième aspect concerne la simulation et l'analyse statistique d'événements peu fréquents à l'échelle de la nanoseconde, et dont on ne connaît pas le chemin de réaction, ici les déformations de la protéine et du chromophore conduisant aux géométries favorables à la conversion interne.

Grâce à ces développements et aux simulations qu'ils ont permises, nous avons réalisé la première modélisation de la désactivation non-radiative par conversion interne à l'échelle de la nanoseconde dans trois protéines fluorescentes différentes. L'analyse des dynamiques moléculaires classiques nous donne une évaluation quantitative des temps de vie de l'extinction de fluorescence, en accord avec les données expérimentales. Par ailleurs elle nous a permis d'identifier les mouvements moléculaires concertés de la protéine et du chromophore conduisant à cette extinction.

De ces résultats, émerge une représentation plus complète du mécanisme qui libère la torsion du chromophore ou qui la déclenche : il peut venir d'un mouvement spécifique de la protéine, qui se produit à l'échelle de la nanoseconde, ou bien de plusieurs mouvements spécifiques, plus fréquents (rupture de liaisons hydrogène, rotation de chaînes latérales, dynamique d'agrégats d'eau), mais qui coïncident seulement à l'échelle de la nanoseconde. Ces mouvements spécifiques n'ont pas un coût énergétique important mais la nécessité de leur coïncidence crée un délai de l'ordre de quelques nanosecondes alors que dans le vide la torsion se produit en quelques picosecondes. Dans le cas des protéines étudiées, on a identifié en grande partie les mécanismes et les acides aminés qui sont impliqués. Ces travaux seront poursuivis et enrichis selon trois axes : (i) la mise au point d'une meilleure description des interactions de type π-stacking dont le rôle est

central dans la YFP et les protéines photoactivables comme Padron, (ii) l'étude du mécanisme moléculaire de l'isomérisation du chromophore dans Padron, ce qui nécessite la mise au point d'un potentiel couplant les torsions autour des deux liaisons intercycles dans l'état fondamental, (iii) l'élaboration d'un modèle cinétique en trois étapes (torsion dans S_1, transition non Born-Oppenheimer S_1-> S_0, retour en S_0 plan) indispensable pour prévoir de façon plus efficace les temps de désactivation. Nos approches pourront également contribuer à l'étude d'autres processus biologiques impliquant plusieurs états électroniques, et/ou des coïncidences comme les transferts d'électrons dans les protéines.

Bibliographie

[1] Tsien R. Y., The green fluorescent protein, *Annual Review of Biochemistry* **1998** *67*, 509-544.

[2] Wang L., Jackson W. C., Steinbach P. A., Tsien R. Y., Evolution of new nonantibody proteins via iterative somatic hypermutation, *Proceedings of the National Academy of Sciences of the United States of America* **2004** *101*, 16745–16749.

[3] Shaner N. C., Campbell R. E., Steinbach P. A., Giepmans B. N., Palmer A. E., Tsien R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein, *Nature Biotechnology* **2004** *22*, 1567–1572.

[4] Shagin D. A., Barsova E. V. , Yanushevich Y. G., Fradkov A. F., Lukyanov K. A., Labas Y. A., Semenova T. N., Ugalde J. A., Meyers A., Nunez J. M., Widder E. A., Lukyanov S. A., Matz M. V., GFP-like Proteins as Ubiquitous Metazoan Superfamily : Evolution of Functional Features and Structural Complexity, *Molecular Biology and Evolution* **2004** *21*, 841-850.

[5] Deheyn D. D., Kubokawa K., McCarthy J. K., Murakami A., Porrachia M., Rouse G. W., Holland N. D., Endogenous Green Fluorescent Protein (GFP) in Amphioxus, *The Biological bulletin* **2007** *213*, 95–100.

[6] Salih A., Larkum A., Cox G., Kuhl M., Hoegh-Guldberg O., Fluorescent pigments in corals are photoprotective, *Nature* **2000** *408*, 850–853.

[7] Bou-Abdallah F., Chasteen N. D., Lesser M. P., Quenching of superoxide radicals by green fluorescent protein, *Biochimica et Biophysica Acta* **2006** *1760*, 1690–1695.

[8] Zimmer M., Green Fluorescent Protein (GFP) : Applications, Structure, and Related Photophysical Behavior, *Chemical Reviews* **2002** *102*, 759–781.

[9] Chudakov D. M., Lukyanov S., Lukyanov K. A., Fluorescent proteins as a toolkit for in vivo imaging, *Trends Biotechnology* **2005** *23*, 605–613.

[10] Day R. N., Davidson M. W., The fluorescent protein palette : tools for cellular imaging, *Chemical Society Reviews* **2009** *38*, 2887–2921.

[11] Wachter R. M., The Family of GFP-Like Proteins : Structure, Function, Photophysics and Biosensor Applications. Introduction and Perspective, *Photochemistry and Photobiology* **2006** *82*, 339–344.

[12] Chalfie M., Tu Y., Euskirchen G., Ward W. W., Prasher D. C., Green Fluorescent Protein as a Marker for Gene Expression, *Science* **1994** *263*, 802–805.

[13] Lippincott-Schwartz J., Patterson G. H., Development and Use of Fluorescent Protein Markers in Living Cells, *Science* **2003** *300*, 87-91.

[14] Egner A., Geisler C., von Middendorff C., Bock H., Wenzel D., Medda R., Andresen M., Stiel A. C., Jakobs S., Eggeling C., Schönle S., Hell S. W., Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters, *Biophysical Journal* **2007** *93*, 3285–3290.

[15] Llopis J., McCaffery J. M., Miyawaki A., Farquhar M. G., Tsien R. Y., Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins, *Proceedings of the National Academy of Sciences of the United States of America* **1998** *95*, 6803–6808.

[16] Kneen M, Farinas J, Li Y, Verkman AS, Green fluorescent protein as a noninvasive intracellular pH indicator, *Biophysical Journal* **1998** *74*, 1591–1599.

[17] Miesenböck G., De Angelis D. A., Rothman J. E., Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, *Nature* **1998** *394*, 192–195.

[18] Miyawaki A., Llopis J., Heim R., McCaffery J. M., Adams J. A., Ikura M., Tsien R. Y., Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin, *Nature* **1997** *388*, 882–887.

[19] Romoser V. A., Hinkle P. M., Persechini A., 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, *The Journal of Biological Chemistry* **1997** *272*, 13270–13274.

[20] Hanson G. T., Aggeler R., Oglesbee D., Cannon M., Capaldi R. A., Tsien R. Y., Remington S. J., Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators, *The Journal of Biological Chemistry* **2004** *279*, 13044–13053.

[21] Fredj A., Pasquier H., Demachy I., Jonasson G., Lévy B., Derrien V., Bousmah Y., Manoussaris G., Wien F., Ridard J., Erard M., Mérola F., The Single T65S Mutation Generates Brighter Cyan Fluorescent Proteins with Increased Photostability and pH Insensitivity, *PlosONE* soumis **2012**

[22] Megley C. M., Dickson L. A., Maddalo S. L., Chandler G. J., Zimmer M., Photophysics and dihedral freedom of the chromophore in yellow, blue, and green fluorescent protein, *The Journal of Physical Chemistry B* **2009** *113*, 302–308.

[23] Patterson G. H., Knobel S. M., Sharif W. D., Kain S. R., Piston, D. W., Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy, *Biophysical Journal* **1997** *73*, 2782–2790.

[24] Niwa H., Inouye S., Hirano T., Matsuno T., Kojima S., Kubota S., Ohashi M., Tsuji F. I., Chemical nature of the light emitter of the Aequorea green fluorescent protein, *Proceedings of the National Academy of Sciences of the United States of America* **1996** *93*, 13617.

[25] de la Lande A., Rezac J., Lévy B., Sanders B. C., Salahub D. R., Transmission Coefficients for Chemical Reactions with Multiple States : Role of Quantum Decoherence, *Journal of the American Chemical Society* **2011** *133*, 3883-3894.

[26] Martin M. E., Negri F., Olivucci, M., Origin, Nature, and Fate of the Fluorescent State of the Green Fluorescent Protein Chromophore at the CASPT2//CASSCF Resolution, *Journal of the American Chemical Society* **2004** *126*, 5452–5464.

[27] Shimomura O., Johnson F. H., Saiga Y., Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea, *Journal of Cellular and Comparative Physiology* **1962** 59, 223–239.

[28] Morin J. G., Hastings J. W., Energy transfer in a bioluminescent system, *Journal of Cellular Physiology* **1971** *77*, 313–318.

[29] Inouye S., Tsuji F. I., Aequorea green fluorescent protein : Expression of the gene and fluorescence characteristics of the recombinant protein, *FEBS Letters* **1994** *341*, 277–280.

[30] Morise H., Shimomura O., Johnson F. H., Winant J., Intermolecular Energy Transfer in the Bioluminescent System of Aequorea, *Biochemistry* **1974** *13*, 2656–2662.

[31] Prasher D. C., Eckenrode V. K., Ward W. W., Prendergast F. G., Cormier M. J., Primary structure of the Aequorea victoria green-fluorescent protein, *Gene* **1992** *111*, 229–233.

[32] Plautz J. D., Day R. N., Dailey G. M., Welsh S. B., Hall J. C., Halpain S., Kay S. A., Green fluorescent protein and its derivatives as versatile markers for gene expression in living Drosophila melanogaster, plant and mammalian cells, *Gene* **1996** *173*, 83–87.

[33] van Roessel P., Brand A. H., Imaging into the future : visualizing gene expression and protein interactions with fluorescent proteins, *Nature Cell Biology* **2002** *4*, E15–E20.

[34] Hadjantonakis A. K., Dickinson M. E., Fraser S. E., Papaioannou V. E., Technicolour transgenics : imaging tools for functional genomics in the mouse, *Nature Reviews Genetics* **2003** *4*, 613–625.

[35] Stewart Jr. C. N., Go with the glow : fluorescent proteins to light transgenic organisms, *Trends in Biotechnology* **2006** *24*, 155–162.

[36] Heim R., Prasher D. C., Tsien R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **1994** *91*, 12501–12504.

[37] Davis D. F., Ward W. W., Cutler M. W., Post-translational chromophore formation in recombinant GFP from E. coli requires oxygen, *Bioluminescence and Chemiluminescence, Fundamentals and Applied Aspects*. Campbell A. K., Kricka L. J., Stanley P. E. (eds.) John Wiley & Sons **1994** 596-599.

[38] Shimomura O., Structure of the chromophore of aquorea green fluorescent protein, *FEBS Letters* **1979** *104*, 220-222.

[39] Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., Ward, W. W., Chemical Structure of the Hexapeptide Chromophore of the Aequorea Green-Fluorescent Protein, *Biochemistry* **1993** *32*, 1212-1218.

[40] Cubitt A. B., Heim R., Adams S. R., Boyd A. E., Gross L. A., Tsien R. Y., Understanding, improving and using green fluorescent proteins, *Trends in Biochemical Sciences* **1995** *20*, 448–455.

[41] Reid B. G., Flynn G. C., Chromophore Formation in Green Fluorescent Protein, *Biochemistry* **1997** *36*, 6786–91.

[42] Barondeau D. P., Putnam C. D., Kassmann C. J., Tainer J. A., Getzoff E. D., Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped

intermediate structures, *Proceedings of the National Academy of Sciences of the United States of America* **2003** *100*, 12111–12116.

[43] Ormö M., Cubitt A. B., Kallio K., Gross L. A., Tsien R. Y., Remington S. J., Crystal Structure of the Aequorea victoria Green Fluorescent Protein, *Science* **1996** *273* 1392–1395.

[44] Yang F., Moss L. G., Phillips Jr. G. N., The molecular structure of green fluorescent protein, *Nature Biotechnology* **1996** *14*, 1246–51.

[45] Sniegowski J. A., Lappe J. W., Patel H. N., Huffman H. A., Wachter R. M., Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein, *The Journal of Biological Chemistry* **2005** *280*, 26248–26255.

[46] Wood T. I., Barondeau D. P., Hitomi C., Kassmann C. J., Tainer J. A., Getzoff E. D., Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis, *Biochemistry* **2005** *44*, 16211–16220.

[47] Ward W. W., Cody C. W., Hart R. C., Cormier M. J., *Photochemistry and Photobiology* **1980** *31*, 61 1-615.

[48] Brejc K., Sixma T. K., Kitts P. A., Kain S. R., Tsien R. Y., Ormö M., Remington S. J., Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein, *Proceedings of the National Academy of Sciences of the United States of America* **1997** *94*, 2306–2311.

[49] Chattoraj M., King B. A., Bublitz G. U., Boxer S. G., Ultra-fast excited state dynamics in green fluorescent protein : Multiple states and proton transfer, *Proceedings of the National Academy of Sciences of the United States of America* **1996** *93*, 8362–8367.

[50] Palm G. J., Zdanov A., Gaitanaris G. A., Stauber R., Pavlakis G. N., Wlodawer A., The structural basis for spectral variations in green fluorescent, *Nature Structural Biology* **1997** *4*, 361–65.

[51] Mérola F., Lévy B., Demachy I., Pasquier H., Photophysics and Spectroscopy of Fluorophores in the Green Fluorescent Protein Family. In Advanced Fluorescence Reporters in Chemistry and Biology I : Fundamentals and Molecular Design, Springer Series on Fluorescence, 1st ed., Demchenko, A. P., Ed. ; Springer : Berlin, **2010** *8*, 347383.

[52] Perozzo M. A., Ward K. B., Thompson R. B., Ward W. W., X-ray diffraction and timeresolved fluorescence analyses of Aequorea green fluorescent protein crystals, *The Journal of Biological Chemistry* **1988** *263*, 7713–7716.

[53] Villoing A., Ridhoir M., Cinquin B., Erard M., Alvarez L., Vallverdu G., Pernot P., Grailhe R., Mérola F., Pasquier H., Complex fluorescence of the cyan fluorescent protein : comparisons with the H148D variant and consequences for quantitative cell imaging, *Biochemistry* **2008** *47*, 12483–12492.

[54] Ai H. W., Shaner N. C., Cheng Z., Tsien R. Y., Campbell R. E., Exploration of New Chromophore Structures Leads to the Identification of Improved Blue Fluorescent Proteins, *Biochemistry* **2007** *46*, 5904–5910.

[55] Tomosugi W., Matsuda T., Tani T., Nemoto T., Kotera I., Saito K., Horikawa K., Nagai T., An ultramarine fluorescent protein with increased photostability and pH insensitivity, *Nature Methods* **2009** *6*, 351–353.

[56] Rizzo M. A., Springer G. H., Granada B., Piston D. W., An improved cyan fluorescent protein variant useful for FRET, *Nature Biotechnology* **2004** *22*, 445–449.

[57] Goedhart J., van Weeren L., Hink M. A., Vischer N. O., Jalink K., Gadella T.W.J., Bright cyan fluorescent protein variants identified by fluorescence lifetime screening. *Nature Methods* **2010** *7*, 137–139.

[58] Wachter R. M., Elsliger M. A., Kallio K., Hanson G. T., Remington S. J., *Structure* **1998** *6*, 1267-1277.

[59] Baird G. S., Zacharias D. A., Tsien R. Y., Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral, *Proceedings of the National Academy of Sciences of the United States of America* **2000** *97*, 11984–11989.

[60] Matz M. V., Fradkov A. F., Labas Y. A., Savitsky A. P., Zaraisky A. G., Markelov M. L., Lukyanov S. A., Fluorescent proteins from nonbioluminescent Anthozoa species, *Nature Biotechnology* **1999** *17*, 969–973.

[61] Verkhusha V. V., Lukyanov K. A., The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins, *Nature Biotechnology* **2004** *22*, 289–296.

[62] Masuda H., Takenaka Y., Yamaguchi A., Nishikawa S., Mizuno H., A novel yellowish-green fluorescent protein from the marine copepod, Chiridius poppei, and its use as a reporter protein in HeLa cells, *Gene* **2006** *372*, 18–25.

[63] Patterson G. H., Lippincott-Schwartz J., A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **2002** *297*, 1873–1877.

[64] Ando R., Mizuno H., Miyawaki A., Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting, *Science* **2004** *306*, 1370–1373.

[65] Wilmann P. G., Turcic K., Battad J. M., Wilce M. C., Devenish R. J., Prescott M., Rossjohn J., The 1.7 A crystal structure of Dronpa : a photoswitchable green fluorescent protein, *Journal of Molecular Biology* **2006** *364*, 213–224.

[66] Andresen M., Stiel A. C., Trowitzsch S., Weber G., Eggeling C., Wahl M. C., Hell S. W., Jakobs S., Structural basis for reversible photoswitching in Dronpa. *Proceedings of the National Academy of Sciences of the United States of America* **2007** *104*, 13005–13009.

[67] Stiel A. C., Trowitzsch S., Weber G., Andresen M., Eggeling C., Hell S. W., Jakobs S., Wahl M. C., 1.8 Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants, *Biochemical Journal* **2007** *402*, 35–42.

[68] Andresen M., Stiel A. C., Folling. J., Wenzel D., Schonle A., Egner A., Eggeling C., Hell S. W., Jakobs S., Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy, *Nature Biotechnology* **2008** *26*, 1035–1040.

[69] Gurskaya N. G., Fradkov A. F., Pounkova N. I., Staroverov D. B., Bulina M. E., Yanushevich Y. G., Labas Y. A., Lukyanov S., Lukyanov K. A., A colourless green fluorescent protein homologue from the non-fluorescent hydromedusa Aequorea coerulescens and its fluorescent mutants, *Biochemical Journal* **2003** *373*, 403–408.

[70] Ganesan S., Ameer-Beg S. M., Ng T. T., Vojnovic B., Wouters F. S., A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACh)

for Förster resonance energy transfer with GFP, *Proceedings of the National Academy of Sciences of the United States of America* **2006** *103*, 4089–4094.

[71] Murakoshi H., Lee S. J., Yasuda R., Highly sensitive and quantitative FRET–FLIM imaging in single dendritic spines using improved non-radiative YFP, *Brain Cell Biology* **2008** *36*, 31–42.

[72] Valeur B., Molecular Fluorescence : Principles and Applications, Wiley-VCH Verlag GmbH, Weinheim, **2001**

[73] Toniolo A., Olsen S., Manohar L., Martínez T. J., Conical intersection dynamics in solution : The chromophore of Green Fluorescent Protein, *Faraday Discuss* **2004** *127*, 149–163.

[74] Mandal D., Tahara T., Meech S. R, Excited-State Dynamics in the Green Fluorescent Protein Chromophore, *The Journal of Physical Chemistry B* **2004** *108*, 1102–1108.

[75] Volkmer A., Subramaniam V., Birch D. J. S., Jovin T. M., One- and Two-Photon Excited Fluorescence Lifetimes and Anisotropy Decays of Green Fluorescent Proteins, *Biophysical Journal* **2000** *78*, 1589–1598.

[76] Ward W. W., Bokman, S. H., Reversible denaturation of Aequorea green-fluorescent protein : physical separation and characterization of the renatured protein, *Biochemistry* **1982** *21*, 4535– 4540.

[77] Strickler S. J., Berg R. A., Relationship between Absorption Intensity and Fluorescence Lifetime of Molecules, *Journal of Chemical Physics* **1962** *37*, 814–822.

[78] Schwille P., Kummer S., Heikal A. A., Moerner W. E., Webb W. W., Fluorescence correlation spectroscopy reveals fast optical excitation- driven intramolecular dynamics of yellow fluorescent proteins, *Proceedings of the National Academy of Sciences of the United States of America* **2000** *97*, 151- 156.

[79] Heim R., Cubift A. B., Tsien R. Y., Improved green fluorescence, *Nature* **1995** *373*, 663–664.

[80] www.pdb.org

[81] Elsliger M. A., Wachter R. M., Hanson G. T., Kallio K., Remington S. J., Structural and Spectral Response of Green Fluorescent Protein Variants to Changes in pH, *Biochemistry* **1999** *38*, 5296-5301.

[82] Brakemann T., Weber G., Andresen M., Groenhof G., Stiel A. C., Trowitzsch S., Eggeling C., Grubmuller H., Hell S. W., Wahl M. C., Jakobs S., Molecular Basis of the Light-driven Switching of the Photochromic Fluorescent Protein Padron, *The Journal of Biological Chemistry* **2010** *285*, 14603-14609.

[83] Regis Faro A., Carpentier P., Jonasson G., Pompidor G., Arcizet D., Demachy I., Bourgeois D., Low-Temperature Chromophore Isomerization Reveals the Photoswitching Mechanism of the Fluorescent Protein Padron, *Journal of the American Chemical Society* **2011** *41*, 16362-16365.

[84] Olsen S., Lamothe K., Martínez T. J., Protonic Gating of Excited-State Twisting and Charge Localization in GFP Chromophores : A Mechanistic Hypothesis for Reversible Photoswitching, *Journal of the American Chemical Society* **2010** *132*, 1192–1193.

[85] Virshup A. M., Punwong C., Pogorelov T. V., Lindquist B. A., Ko C., Martínez T. J., Photodynamics in Complex Environments : Ab Initio Multiple Spawning Quantum Mechanical/Molecular Mechanical Dynamics, *Journal of Physical Chemistry B* **2009** *113*, 3280–3291.

[86] Maddalo S. L., Zimmer M., The Role of the Protein Matrix in Green Fluorescent Protein Fluorescence, *Photochemistry and Photobiology* **2006** *82*, 367–372.

[87] Moors S. L. C., Michielssens S., Flors C., Dedecker P., Hofkens J., Ceulemans A., How Is cis-trans Isomerization Controlled in Dronpa Mutants ? A Replica Exchange Molecular Dynamics Study, *Journal of Chemical Theory and Computation* **2008** *6*, 1012–1020.

[88] Vallverdu G., Demachy I., Ridard J., Levy B., Using biased molecular dynamics and Brownian dynamics in the study of fluorescent proteins, *Journal of Molecular Structure : THEOCHEM* **2009** *898*, 73–83.

[89] Schäfer L. V., Groenhof G., Boggio-Pasqua M., Robb M. A., Grubmüller H., Chromophore Protonation State Controls Photoswitching of the Fluoroprotein asFP595, *PLoS Computational Biology* **2008** *4*, e1000034.

[90] Li X., Chung L. W., Mizuno H., Miyawaki A., Morokuma K., Primary Events of Photodynamics in Reversible Photoswitching Fluorescent Protein Dronpa, *The Journal of Physical Chemistry Letters* **2010** *1*, 3328–3333.

[91] Vendrell O., Gelabert R., Moreno M., Lluch J. M., Potential energy landscape of the photoinduced multiple proton-transfer process in the green fluorescent protein : Classical molecular dynamics and multiconfigurational electronic structure calculations, *Journal of the American Chemical Society* **2006** *128*, 3564-3574.

[92] Jonasson G., Teuler J.-M., Vallverdu G., Mérola F., Ridard J., Lévy B., Demachy I., Excited State Dynamics of the Green Fluorescent Protein on the Nanosecond Time Scale, *Journal of Chemical Theory and Computation* **2011** *7*, 1990–1997.

[93] Olsen S., Smith S. C., Bond Selection in the Photoisomerization Reaction of Anionic Green Fluorescent Protein and Kindling Fluorescent Protein Chromophore Models, *Journal of the American Chemical Society* **2008** *130*, 8677-8689.

[94] Polyakov I. V., Grigorenko B. L., Epifanovsky E. M., Krylov A. I., Nemukhin A. V., *Journal of Chemical Theory and Computation* **2010** *6*, 2377-2387.

[95] Olsen S., McKenzie R., A diabatic three-state representation of photoisomerization in the green fluorescent protein chromophore, *Journal of Chemical Physics* **2009** *130*, 184302.

[96] Case D. A., Darden T. A., Cheatham T. E., III, Simmerling C. L., Wang J., Duke R. E., Luo R., Crowley M., Walker R. C., Zhang W., Merz K. M., Wang B., Hayik S., Roitberg A., Seabra G., Kolossvary I., Wong K. F., Paesani F., Vanicek J., Wu X., Brozell S. R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D. H., Seetin M. G., Sagui C., Babin V., Kollman P. A., AMBER 10, University of California : San Francisco, CA, **2008**.

[97] Roos B. O., Taylor P. R., Siegbahn, P. E. M., A complete active space scf method (cassf) using a density matrix formulated super-ci approach, *Chemical Physics* **1980** *48*, 157-173.

[98] Bravaya K. B., Khrenova M., Grigorenko B., Nemukhin A. V., Krylov A. I., The effect of protein environment on electronically excited and ionized states of the green fluorescent protein chromophore, *Journal of Physical Chemistry B* **2011** *8*, 8296 – 8303.

[99] Olsen S., A Modified Resonance-Theoretic Framework for Structure-Property Relationships in a Halochromic Oxonol Dye, *Journal of Chemical Theory and Computation* **2010** *6*, 1089–1103.

[100] Braida B., Walter C., Engels B., Hiberty P. C., A Clear Correlation Between the Diradical Character of 1,3-Dipoles and their Reactivity Towards Ethylene and Acetylene, *Journal of the American Chemical Society* **2010** *132*, 7631–7637.

[101] Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Montgomery J. A., Jr., Vreven T., Kudin K. N., Burant J. C., Millam J. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q., Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C., Pople J. A., Gaussian 03, Revision C.02 ; Gaussian, Inc. : Wallingford, CT, **2004**.

[102] Werner H.-J., Knowles P. J., Manby F. R., Sch€utz M., Celani P., Knizia G., Korona T., Lindh R., Mitrushenkov A., Rauhut G., Adler T. B., Amos R. D., Bernhardsson A., Berning A., Cooper D. L., Deegan M. J. O., Dobbyn A. J., Eckert F., Goll E., Hampel C., Hesselmann A., Hetzer G., Hrenar T., Jansen G., Koppl C., Liu Y., Lloyd A. W., Mata R. A., May A. J., McNicholas S. J., Meyer W., Mura M. E., Nicklass A., Palmieri P., Pfl€uger K., Pitzer R., Reiher M., Shiozaki T., Stoll H., Stone A. J., Tarroni R., Thorsteinsson T., Wang M., Wolf, A. MOLPRO, version 2006.1, University College Cardiff Consultants Limited : Cardiff, Wales, **2010**.

[103] Liu, R. S. H., Asato A. E., The primary process of vision and the structure of bathorhodopsin : A mechanism for photoisomerization of polyenes, *Proceedings of the National Academy of Sciences of the United States of America* **1985** *82*, 259–263.

[104] Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., Modeling of Data. In Numerical Recipes in FORTRAN : The Art of Scientific Computing, 2nd ed., Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., Eds., Cambridge University Press : New York, **1992** 690691.

[105] Koster A. M., Calaminici P., Casida M. E., Dominguez V. D., Flores-Moreno R., Geudtner G., Goursot A., Heine T., Ipatov A., Janetzko F., del Campo J. M., Reveles J. U., Vela A., Zuniga B., Salahub D. R., deMon2k, Version 2, The deMon developers, Cinvestav, Mexico City **2006**.

[106] Bayly C., Cieplak P., Cornell W., Kollman P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges : The RESP Model, *The Journal of Physical Chemistry* **1993** *97*, 10269-10280.

[107] Ridard J., Lévy B., Effective atomic charges in alanine dipeptide, *Journal of Computational Chemistry* **1999** *20*, 473–482.

[108] Lévy B., Enescu M., Theoretical study of Methylene Blue ; a new method to determine partial atomic charges ; investigation of interaction with Guanine, *Journal of Molecular Structure : THEOCHEM* **1998** *432*, 235.

[109] Vengris M., van Stokkum I. H. M., He X., Bell A. F., Tonge P. J., van Grondelle R., Larsen D. S., Ultrafast excited and ground-state dynamics of the green fluorescent protein chromophore in solution, *Journal of Physical Chemistry A* **2004** *108*, 4587-4598.

[110] Martínez T. J., Insights for Light-Driven Molecular Devices from Ab Initio Multiple Spawning Excited-State Dynamics of Organic and Biological Chromophores, *Accounts of Chemical Research* **2006** *39* 119–126.

[111] Torrie G. M., Valleau J.P., Nonphysical sampling distributions in Monte Carlo free-energy estimation - Umbrella sampling, *Journal of Computational Physics* **1977** *23*, 187-199.

[112] Laio A., Parinello M., Escaping free-energy minima, *Proceedings of the National Academy of Sciences of the United States of America* **2002** *99*, 12562-12566.

[113] Rosenberg R. O., Berne B. J., Chandler D., Isomerization dynamics in liquids by molecular dynamics, *Chemical Physics Letters* **1980** *75*, 162-168.

[114] Ciccotti G., Ferrario M., Blue Moon approach to rare events, *Molecular simulation* **2004** *30*, 787-793.

[115] Ciccotti G., Kapral R., Vanden Eijnden E., Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics, *ChemPhysChem* **2005** *6*, 1809-1814.

[116] Swendsen R. H., Yang J. S., Replica Monte Carlo simulation of spin-glasses, *Physical Review Letters* **1986** *57*, 2607-2609.

[117] Hänggi P.,Talkner P., Borkovec M., Reaction Rate Theory : Fifty Years After Kramers, *Reviews of Modern Physics* **1990** *62*, 251–342.

[118] Marcus R.A., On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I, *Journal of Chemical Physics* **1956** *24*, 966-978.

[119] Litvinenko K. L., Webber N. M., Meech S. R., An ultrafast polarisation spectroscopy study of internal conversion and orientational relaxation of the chromophore of the green fluorescent protein, *Chemical Physics Letters* **2001** *346*, 47-53.

[120] Hendlich M., Rippmann F., Barnickel G., Ligsite : automatic and efficient detection of potential small molecule-binding sites in proteins, *Journal of Molecular Graphics and Modelling* **1997** *15*, 359–363. http ://www.modelling.leeds.ac.uk/pocketfinder/

[121] Humphrey W., Dalke A., Schulten K., VMD - Visual Molecular Dynamics, *Journal of Molecular Graphics* **1996** *14.1*, 33-38.

[122] Mouawad L., Perahia D., Motions in Hemoglobin Studied by Normal Mode Analysis and Energy Minimization : Evidence for the Existence of Tertiary T-like, Quaternary R-like Intermediate Structures, *Journal of Molecular Biology* **1996** *258*, 393-410.

[123] Mouawad L., Perahia D., Robert C.H., Guilbert C., New Insights into the Allosteric Mechanism of Human Hemoglobin from Molecular Dynamics Simulations, *Biophysical Journal* **2002** *82*, 3224-3245.

[124] Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M., CHARMM : A program for macromolecular energy, minimization, and dynamics calculations, *Journal of Computational Chemistry* **1983** *4*, 187-217.

[125] Webber N. M., Litvinenko K. L., Meech S. R., Radiationless Relaxation in a Synthetic Analogue of the Green fluorescent Protein Chromophore, *The Journal of Physical Chemistry B* **2001** *105*, 8036-8039.

[126] Ryckaert J.-P., Ciccotti G., Berendsen H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints : Molecular Dynamics of n-Alkanes, *Journal of Computational Physics* **1977** *23*, 327–341.

[127] Miyamoto S., Kollman P. A., SETTLE : An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *Journal of Computational Chemistry* **1992** *13*, 952–962.

[128] Loncharich R. J., Brooks B. R., Pastor R. W., Langevin dynamics of peptides : The frictional dependence of isomerization rates of N-actylananyl-N'-methylamide. *Biopolymers* **1992** *32*, 523-535.

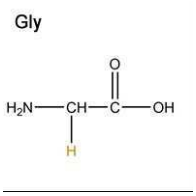
[129] Darden T., York D., Pederson L., PARTICLE MESH EWALD - AN N.LOG(N) METHOD FOR EWALD SUMS IN LARGE SYSTEMS, *Journal of Chemical Physics* **1993** 98, 10089-10092.

[130] Pastor R. W., Brooks B. R., Szabo A., An analysis of the accuracy of Langevin and molecular dynamics algorithms, *Molecular Physics* **1988** *65*, 1409-1419.

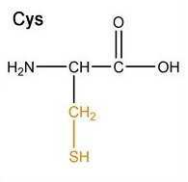
[131] Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J.R. Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics* **1984** *81*, 3684–3690.

[132] Reuter N., Lin H., Thiel W., 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, *The Journal of Physical Chemistry B* **2002** *106*, 6310-6321.

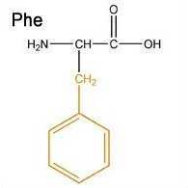
A. Acides Aminés



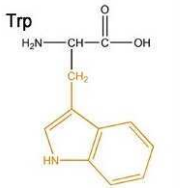
Glycine (G)



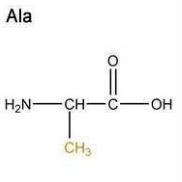
Cysteine (C)



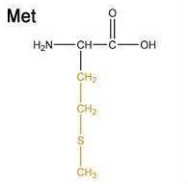
Phenylalanine (F)



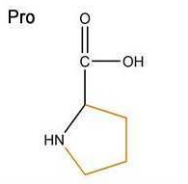
Tryptophane (W)



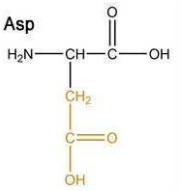
Alanine (A)



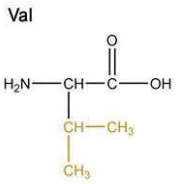
Méthionine (M)



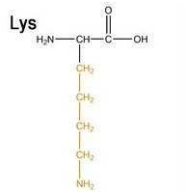
Proline (P)



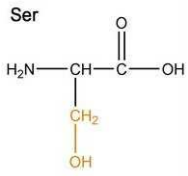
Acide aspartique (D)



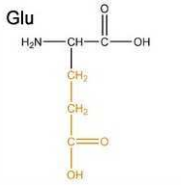
Valine (V)



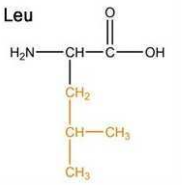
Lysine (K)



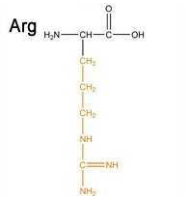
Serine (S)



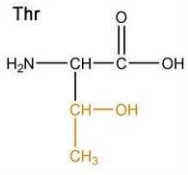
Acide glutamique (E)



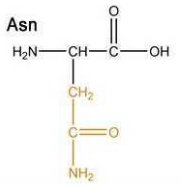
Leucine (L)



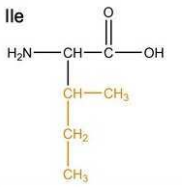
Arginine (R)



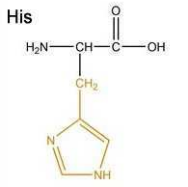
Thréonine (T)



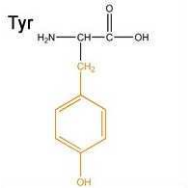
Asparagine (N)



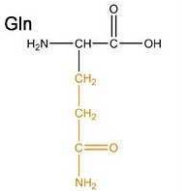
Isoleucine (I)



Histidine (H)



Tyrosine (Y)



Glutamine (Q)

FIG. A.1.: Les 20 acides aminés naturels.

B. Interpolation

On cherche à interpoler un grand nombre de valeur (n=118) d'énergie par une expression linéaire sous la forme

$$V_{NADiA}(\tau, \varphi) = \sum k_i g_i(\tau, \varphi)$$

où $g_i(\tau, \varphi)$ est une fonction où un produit de fonctions sinus et cosinus dépendant des angles τ et φ et les k_i sont les paramètres on cherche à ajuster. On pose le système d'équations linéaires suivant :

$$Gk = q$$

où G est la matrice des valeurs des fonctions $g(\tau, \varphi)$ dans chaque point de calcul CASPT2 , k est le vecteur des paramètres k_i (les inconnues) et q est le vecteur des valeurs d'énergie CASPT2. Les paramètres sont ajustés par la méthode des moindres carrés où on cherche à minimiser les résidus r de l'interpolation :

$$\|r\| = \|Gk - q\|$$

La solution qui donne le plus petit norme $\|k\|$ est donnée par

$$k = (G^T G)^{-1} G^T q$$

Afin de calculer le vecteur k on définit la matrice de covariance $A = G^T G$ et on réexprime A par la méthode de décomposition en valeurs singulières :

$$A = U \omega V^T$$

où U est une matrice avec des colonnes orthonormées, ω est une matrice diagonale dont les éléments diagonaux sont des valeurs singulières et V^T est la transposée d'une matrice orthogonale V. La matrice A est carrée, et par conséquent, U, ω et V aussi. L'inversion de la matrice est donc simplement donné par

$$A^{-1} = V[\mathrm{diag}(\frac{1}{\omega_i})]U^T$$

où l'inverse de U et V^T est donné par leur transposée et l'inverse de ω est donné par la matrice diagonale dont les éléments diagonaux sont les inverses des valeurs singulières. Le vecteur k peut donc être calculé par

$$k = A^{-1}G^T q$$

Afin de trouver la fonction $g(\tau, \varphi)$ qui peut apporter la plus grande amélioration à l'interpolation $V_{\mathrm{NADiA}}(\tau, \varphi)$, nous déterminons la projection orthogonale $\wp^{\perp}$de la nouvelle fonction $g(\tau, \varphi)$ sur les fonctions de l'interpolation.

La fonction $f(\tau, \varphi)$ ayant le plus grand recouvrement entre $\wp^{\perp}$ et les résidus de l'interpolation est sélectionné et rajouté à $V_{\mathrm{NADiA}}(\tau, \varphi)$. La projection $\wp$ et la projection orthogonale $\wp^{\perp}$ sont données par

$$\wp(f) = GA^{-1}G^T f$$

$$\wp(f)^{\perp} = f - \wp(f)$$

où f est le vecteur des valeurs de la nouvelle fonction candidate dans chaque point de calcul CASPT2.

C. Simulation de dynamique moléculaire

C.1. Les lois du mouvement de Newton

Dans une simulation de dynamique moléculaire, les lois du mouvement de Newton sont utilisés pour calculer le mouvement des atomes. Chaque atome est traité comme un point, avec une masse et une charge, et ressent une force et une accélération. L'intégration des équations de mouvement donne une trajectoire qui décrit les positions, les vitesses et les accélérations des atomes en fonction du temps. La deuxième loi de Newton peut être récapitulée dans l'équation suivante :

$$\overrightarrow{F_i} = m_i \overrightarrow{a_i} \tag{C.1}$$

où $\overrightarrow{F_i}$ désigne la force exercée sur l'atome i, m_i la masse et $\overrightarrow{a_i}$ l'accélération de l'atome i. La force peut également être exprimée comme le gradient de l'énergie potentielle du système V :

$$F = -\nabla V \tag{C.2}$$

La combination de ces deux équations donne une relation entre la dérivée de l'énergie potentielle et le changement en position en fonction du temps :

$$-\frac{\partial V}{\partial r_i} = m_i \frac{\partial^2 r_i}{\partial t^2} \tag{C.3}$$

Dans une simulation de dynamique moléculaire l'énergie potentielle est définie par le champ de forces (voir chapitre A et section C.7). La définition de positions et de vitesses initiales, permet de résoudre l'équation C.3 numériquement par un algorithme d'intégration où les positions, les vitesses et les accélérations sont exprimées par une série Taylor. Le programme de dynamique moléculaire Amber utilise l'algorithme *leap frog* qui est un schéma d'intégration où les vitesses v sont calculé à un temps $t + \frac{1}{2}dt$:

$$v(t + \frac{1}{2}dt) = v(t - \frac{1}{2}dt) + a(t)dt \tag{C.4}$$

pour déduire les positions r à un temps $t + dt$:

$$r(t + dt) = r(t) + v(t + \frac{1}{2}dt)dt \tag{C.5}$$

C.2. Mécanique Statistique

La dynamique moléculaire a pour but de simuler le comportement d'un système macroscopique, e.g. une protéine, à partir des propriétés des constituants microscopiques, e.g. les atomes. La mécanique statistique introduit un lien entre les données microscopiques, e.g. les positions et les

moments, et les observables macroscopiques, e.g. la pression, l'énergie, ou la capacité calorifique. Le passage entre les deux échelles peut se faire par des moyennes d'ensemble. Un ensemble est une collection de tous les systèmes possibles, de différents états microscopiques, qui donne un même état macroscopique ou thermodynamique. Un état thermodynamique d'un système de n composants peut être caractérisé par n+2 variables macroscopiques, ou n+1 in vacuo. Ces variables peuvent être extensifs, e.g. le nombre de particules (N), le volume (V), l'énergie (E), ou intensifs, e.g. le potentiel chimique (μ), la pression (P) ou la température (T). L'ensemble thermodynamique sera déterminé par la sélection de n+2 variables et par la définition de valeurs de référence macroscopique. On dit qu'un système est à l'état d'équilibre thermodynamique si ces variables d'état ont des valeurs bien définies et constantes.

En mécanique statistique, les valeurs moyennes des grandeurs macroscopiques sont définies par des moyennes d'ensemble. La moyenne de l'observable A est ainsi donnée par :

$$\langle A \rangle_{ensemble} = \iint dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) \tag{C.6}$$

où p désigne les moments et r les positions du système. La densité de probabilité ρ de l'ensemble est donné par

$$\rho(p^N, r^N) = \frac{1}{Q} exp(-\frac{H(p^N, r^N)}{k_B T}) \tag{C.7}$$

où H désigne l'Hamiltonian, T la température, k_B la constante de Boltzmann et Q la fonction de partition.

En simulation de dynamique moléculaire, le calcul inaccessible de la fonction de partition Q, est évité par le calcul de la moyenne de l'observable A par une moyenne du temps :

$$\langle A \rangle_{temps} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \approx \frac{1}{M} \int_{t=1}^{M} A(p^N, r^N) \tag{C.8}$$

où t désigne le temps de simulation, M le nombre de pas de temps de la simulation et $A(p^N, r^N)$ la valeur instantanée de A. Ces relations conduisent à l'hypothèse d'ergodicité qui affirme qu'à l'équilibre, la valeur moyenne d'une grandeur calculée de manière statistique est égale à la moyenne d'un très grand nombre de mesures prises dans le temps :

$$\langle A \rangle_{ensemble} = \langle A \rangle_{temps} \tag{C.9}$$

C.3. Des conditions aux limites

Une simulation est contrôlée par des conditions aux limites. Dans cette étude nous avons appliqué des conditions aux limites spatiales ainsi que thermodynamiques.

Les contraintes spatiales imposent la forme et la taille ainsi que la périodicité du système. Dans cette étude, nous imposons une forme rectangulaire du système. Des molécules d'eau sont ajoutées pour former une boite rectangulaire avec une dimension adaptée à la distance limite des interactions à longue distance. Une périodicité infinie du système est également imposée dans les simulations NPT pour éviter les effets de taille finie et permettre l'utilisation de la méthode Particle Mesh Ewald (PME) décrite en chapitre C.5.

Les contraintes thermodynamiques définissent l'état macroscopique et permettent d'imposer des valeurs des observables thermodynamiques. Le calcul de pression peut être problématique à basse température donc la première partie de l'équilibration qui consiste à chauffer le système est effectuée dans l'ensemble canonique (NVT). Le volume reste alors constant et la température est réglée par un thermostat. La deuxième partie de l'équilibration ainsi que les simulations longues sont effectuées dans l'ensemble isotherme-isobare (NPT). La valeur moyenne de la pression ainsi que la température sont imposées par un barostat et un thermostat, respectivement, pour reproduire les conditions de référence, e.g. des conditions expérimentales.

C.4. Contrôle de la température et de la pression (thermostat et barostat)

La régulation de la température permet de comparer à des conditions expérimentales, d'étudier la dépendance en température, d'évacuer ou de rajouter de l'énergie lors des simulations hors équilibre, d'augmenter l'efficacité d'une recherche conformationelle, e.g. recuit simulé, ou d'éviter des dérives de l'énergie au cours des simulations longues, e.g. à cause d'une accumulation d'erreurs numériques. Dans les simulations de cette étude la température est contrôlée par le thermostat de Langevin implémenté dans Amber [128, eq. 11]. La dynamique de Langevin est une approche qui étend la dynamique moléculaire en incluant des effets de viscosité du solvant. L'équation de Newton (eq. C.1) est alors remplacée par l'équation de Langevin :

$$F_i = m_i a_i = -\nabla V_i - \gamma_i + R(t) \tag{C.10}$$

où γ_i désigne une force de frottement visqueux et $R(t)$ une force fluctuante qui représente les impacts aléatoires des molécules du solvant sur l'atome.

Dans nos simulations, la pression est réglée par un ajustement du volume du système. La dimension de la maille rectangulaire est modifiée par un petit ajustement à chaque pas de simulation pour faire converger la pression vers la valeur de référence. Avec ce traitement, la pression instantanée fluctue beaucoup mais la pression moyenne de la simulation est proche de la valeur de référence.

C.5. La méthode Particle Mesh Ewald

Afin de gagner en efficacité de calcul, l'énergie d'interaction électrostatique entre atomes est calculée par la méthode Particle Mesh Ewald (PME) [129]. La méthode s'applique à des systèmes périodiques où les particules interagissent par des lois de forces en carré inverse. Le principe consiste à remplacer la sommation classique des interactions électrostatiques entre charges ponctuelles (somme directe), par deux sommations : une pour les interactions à courte distance et une pour les interactions à longue distance. Les nouvelles sommations convergent efficacement dans l'espace réel et l'espace de Fourier, respectivement. La convergence rapide fait gagner du temps de calcul et permet de tronquer le calcul sans une perte significative de précision.

La sommation à courte distance dans l'espace réel (somme directe) se présente comme la sommation classique des interactions électrostatiques entre charges ponctuelles (eq. C.11). La somme intègre toutes les interactions en-dessous une distance limite définie par le rayon de coupure (ici : 10 Å).

$$E_{cd} = \sum_{i,j} \varphi_{cd}(\mathbf{r}_j - \mathbf{r}_i) \tag{C.11}$$

La sommation à longue distance dans l'espace de Fourier (somme réciproque) se présente sous la forme suivante :

$$E_{ld} = \sum_{k} \tilde{\Phi}_{ld}(k) \mid \tilde{\rho}(k) \mid^2 \tag{C.12}$$

où $\tilde{\Phi}_{ld}$ désigne la transformée de Fourier du potentiel électrostatique et où $\tilde{\rho}(k)$ désigne la transformée de Fourier de la densité de charge, qui est calculée par la transformée de Fourier rapide. Dans cet algorithme, le potentiel électrostatique créé par les charges du système est évalué sur une maille discrète de l'espace.

L'implémentation de NADiA (voir section III.4) comprend la suppression de certaines interactions électrostatiques intra-chromophore. Ce traitement ne modifie pas les charges des atomes impliqués et l'énergie électrostatique des interactions restant, notamment les interactions chromophore-protéine, n'est donc pas modifiée.

C.6. Préparation de la structure initiale

C.6.1. La structure cristalline

La simulation de dynamique moléculaire d'une protéine nécessite des cordonnées initiales. Nous avons utilisé des structures déterminées par diffraction de rayons X. Dans cette technique, la protéine est cristallisée et soumise à un rayonnement électo-magnétique. L'information sur les positions, les intensités et les phases relatives des réflexions de la diffraction est enregistrée. Ce résultat permet d'établir une carte de la densité électronique du cristal. Un modèle de la structure de la protéine est ensuite construit pour reproduire au mieux la densité électronique. La diffraction apparaît quand les rayons X interagissent avec le nuage d'électron des atomes. La diffraction de rayons X permet de déceler les positions de tous les atomes lourds (mais non les hydrogènes).

La précision d'une structure cristallographique est exprimée en Å et représente une distance. Les structures déposées dans la PDB (*Protein Data Bank*) ont une précision de 2,2 Å en moyenne (écart type 1,2 Å). Une telle résolution implique que deux atomes séparés par au moins cette distance peuvent être distingués sur la carte de densité électronique et leur positions peuvent être obtenues directement en haute précision. Les atomes qui sont séparés par moins de 2,2 Å ont des densités électroniques confondues sur la carte. Les positions exactes de ces atomes doivent être déduites de la forme de la densité électronique et de la connaissance de la structure chimique des acides aminés. Le niveau de précision de la carte de densité électronique dépend des paramètres expérimentaux, notamment la longueur d'onde du rayonnement et l'angle maximal de détection.

L'accord entre le modèle cristallographique et la densité électronique est caractérisé par le degré de fiabilité R (anglais : *residual factor* ou *reliability factor*). Une petite valeur de R indique que le modèle décrit bien les données observées.

La structure cristalline de la S65T-GFP, de la YFP et de Padron trans (B_{trans}) sont prises dans la la *protein data bank* [80] sous les noms de 1EMA[43], 1YFP[58] et 3ZUF [83], respectivement. Ces structures sont déterminée avec une résolution de 1,90 Å, 2,5 Å et 2,20 Å, respectivement, et une degré de fiabilité R inférieure à 20 %.

C.6.2. Compléter la structure

La structure cristalline contient uniquement les cordonnées relatives des atomes lourds. Les positions des atomes d'hydrogène n'apparaissent pas dans la diffraction de rayons X. Certains atomes lourds peuvent également manquer dans la structure, soit à cause d'une manque de données, soit à cause d'une grande flexibilité des atomes ce qui a pour effet de brouiller la carte de densité électronique. Afin d'avoir une structure complète et chimiquement correcte de la protéine, il faut donc compléter la structure cristalline avec les atomes manquants.

Le programme *leap* dans Amber permet d'automatiser un tel traitement. Il reconstitue les atomes manquants dans chaque résidu à partir d'un patron prédéfini. Il faut donc s'assurer qu'il existe un patron dans Amber pour chaque résidu de la protéine. Le chromophore a nécessité la création d'un tout nouveau patron qui est décrit dans la section III.8. Dans certain cas, il faut également faire un choix entre plusieurs patrons pour définir un résidu. Le programme *leap* est adapté au format du fichier de coordonnées téléchargables à la PDB. Lors de la lecture du fichier le choix de patron est fait à partir d'un code à trois lettres qui désigne l'acide aminé, e.g. ALA pour Alanine. Le plupart des codes utilisés par la PDB sont en accord avec *leap* mais il existe quelques exceptions. Les acides aminés titrables changent d'état de protonation en fonction du pH et peuvent donc être décrits par différents patrons. L'histidine a trois états de protonation possibles dans Amber. Ils sont désignés par les codes HID, HIE et HIP qui correspondent à *protoné sur azote* N_δ, *protoné sur azote* N_ϵ et *doublement protoné*, respectivement (Fig. C.1). L'acide glutamique a deux états de protonation possibles ; GLU pour *non protoné* et GLH pour *protoné*.

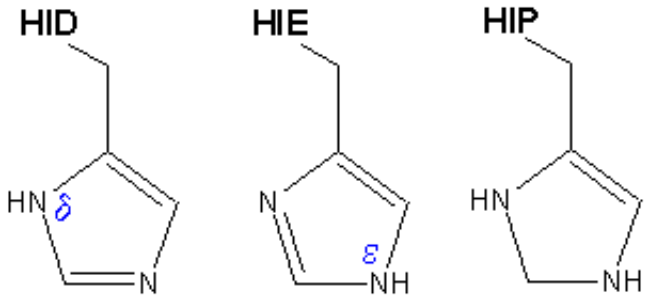


FIG. C.1.: Schéma des trois états de protonations de l'histidine utilisés dans Amber : protoné sur l'azote en position δ (HID), ε (HIE) ou les deux (HIP). L'azote en position δ est appelé *ND1* et celui en position ε est appelé *NE2.*

Nous avons également créé des nouveaux patrons pour au mieux préparer la structure cristalline. Amber a un seul patron pour la tyrosine (TYR) alors que deux rotamères existent. Le proton du groupement OH peut être placé dans deux directions différentes tout en restant dans le plan du

phénolate. Certes, il s'agit d'une simple rotation pour passer entre les deux conformations lors d'une dynamique mais une mauvaise définition initiale peut conduire à un réarrangement local de la protéine qui défavorise ce passage et/ou qui éloigne la géométrie de la structure cristalline. Nous avons donc créé un nouveau patron pour la deuxième rotamère.

Les acides aminés 'titrables' ont été choisis dans leur état standard de protonation à pH neutre. Pour les histidines, dont le pKa est proche de 7, la protonation a été choisie en fonction des possibilités de liaison H avec l'environnement proche [92]. Dans certains cas différents réseaux de liaisons H sont possibles et une série de tests sont nécessaires pour faire un choix (voir section IV.3).

Afin de simuler la dynamique de la protéine en solution nous avons rajouté une boite d'eau explicite autour de la protéine. Un petit nombre d'ions, Na^+ ou Cl^-, sont également rajoutés pour neutraliser la charge totale du système.

A la fin de la préparation de la protéine le programme *leap* crée un fichier de données d'entrée pour la dynamique qui contient les paramètres du champ de force de Amber 1999 (ff99) et la topologie du système.

C.6.3. Minimisation et mise à l'équilibre du système

Les cordonnées du fichier PDB correspondent à une structure cristallisée. Il faut donc chauffer le système pour générer une structure correspondant à la température ambiante. Pour contrôler ce passage, une procédure de minimisation et équilibration avec plusieurs étapes est utilisée. Les résidus reconstitués lors du traitement de *leap* peuvent être mal orientés dans l'espace et entraîner des clashs stériques. Lors d'une minimisation la libération de cette position va générer de l'énergie, on parle de points chauds, qui peut conduire à de grands changements structuraux. Pour éviter ce phénomène, les cordonnées sont d'abord fixées par des contraintes fortes (5000 kcal/mol) qui sont enlevées petit à petit au cours de la procédure de minimisation et d'équilibration.

La minimisation se déroule en trois étapes. D'abord, uniquement l'orientation des atomes d'hydrogène est optimisée. Cela permet d'optimiser les angles des liaisons H. Ensuite, les molécules d'eau sont également relâchées. Cela permet d'affiner le réseau de liaison H mais également d'optimiser les interactions entre la protéine et les molécules d'eau du solvant. Finalement, les contraintes sur la protéine sont atténuées (5 kcal/mol) pour optimiser l'ensemble des cordonnées sans pour autant permettre des trop grands changements structuraux.

L'équilibration du système se déroule en deux temps. Le système est d'abord chauffé progressivement de 100 K à 300 K à volume constant pendant 200 ps. Il s'agit d'une dynamique dans l'ensemble canonique (NVT), i.e. le nombre d'atome, le volume et la température sont conservés. Un thermostat est utilisé pour régler la température. Les contraintes imposées dans la minimisation sont toujours maintenues sur la protéine. Dans un deuxième temps, le système est équilibré à pression constante. On parle alors d'une dynamique dans l'ensemble isotherme-isobare (NPT), i.e. le nombre d'atomes, la pression et la température sont conservés. Au cours de cette équilibration, les contraintes sur la protéine sont diminuées progressivement dans l'ordre qui suit ; 5.0 kcal/mol, 2.0 kcal/mol, 1.0 kcal/mol, 0.5 kcal/mol, 0.1 kcal/mol jusqu-à que le

système soit complètement libre. Chaque contrainte est maintenu pendant 50 ps. Finalement une équilibration de 500 ps sans contrainte est effectuée.

C.6.4. Dynamique moléculaire : phase de production

Après la mise en équilibre du système, l'exploitation de la dynamique de la protéine peut commencer. Afin d'explorer la dynamique à l'état S_0 nous avons effectué une première simulation longue (>10 ns) de la protéine. Cela permet d'engendrer des structures, tout le long de la trajectoire, qui peuvent servir comme points de départ de différentes simulations à l'état excité (S_1). Différentes coordonnées et leurs vitesses correspondantes prises de l'état S_0 sont donc utilisées comme données initiales dans chaque simulation à l'état S_1. Elles sont ensuite directement soumises au potentiel du chromophore de l'état S_1. Ce mode opératoire et en analogie avec une transition électronique Franck-Condon où le chromophore arrive à l'état S_1 dans la géométrie d'équilibre de l'état S_0. Le chromophore se relaxe ensuite sur la surface d'énergie potentielle (PES) de l'état S_1, i.e. de la géométrie Franck-Condon à la géométrie fluorescente. La dynamique de la protéine à l'état S_1 est ensuite explorée pendant plusieurs nanosecondes (>6 ns).

Le pas de temps de la dynamique est limité par les mouvements les plus rapides du système : les vibrations des hydrogènes, qui implique un pas de 0,5-1 fs. Afin d'augmenter le pas de temps et ainsi accéléré le temps de calcul, nous avons utilisé l'algorithme SHAKE [126] qui contraint la longueur des liaisons contenant un hydrogène à sa valeur d'équilibre. Cela nous permet d'utiliser un pas de temps de 2 fs, plus économe en temps de calcul. La protéine est solvatée par des molécules d'eau du modèle TIP3P, ajustées pour l'algorithme SHAKE [127].

Le pas d'écriture des coordonnées est, en revanche, limité par l'espace disque nécessaire. Nous avons besoin d'effectuer des dizaines de simulations pendant plusieurs nanosecondes à la recherche d'un événement rare, i.e. la torsion du chromophore. Afin de permettre une gestion raisonnable des données, nous utilisons un pas d'écriture de 2 ps. Une nanoseconde de dynamique génère ainsi 500 jeux de cordonnées et vitesses du système, ce qui correspond à 350Mb dans le cas de la GFP. Avec ce pas d'écriture, le temps du premier passage (FPT) du chromophore à une géométrie perpendiculaire (discuté dans le chapitre V), dans une simulation donnée, est déterminé avec une erreur de <2 ps. Ce pas d'écriture permet de suivre la dynamique de la protéine et son interaction avec le chromophore, e.g. le mouvement d'un feuillet β et la formation d'une liaison hydrogène. Le mécanisme de torsion du chromophore est beaucoup plus rapide que le FPT moyen et peut se produire à l'échelle de la picoseconde (calcul théorique du chromophore isolé [84]). Un temps d'écriture du même ordre de grandeur ne pourrait donc pas donner une description précise sur le moment même de torsion. Cependant, le mouvement ou la combinaison de mouvements spécifiques de la protéine qui libèrent la torsion après plusieurs nanosecondes sont susceptibles d'être plus lents que la dynamique interne du chromophore. Il est donc possible d'identifier les mouvements de la protéine juste après le moment de torsion avec ce temps d'écriture.

La température du système est contrôlée par une dynamique de Langevin [130, 128]. Ce type de thermostat imite les conditions d'une particule dans un fluide visqueux. Un terme de bruit, dû au collisions entre la particule et les molécules du fluide, et une force de friction sont utilisés pour

ajuster la température. En utilisant une fréquence de collision de 1 ps^{-1}, relativement petite par rapport à la fréquence physique de $\sim$50 ps^{-1}, la température est mieux stabilisée. La dynamique est propagée par une intégration Leap-Frog ("saut-mouton") en ajustant l'énergie cinétique à la bonne température.

Nous utilisons des conditions périodiques aux limites à pression constante. La pression est contrôlée par un ajustement isotrope de la taille de la boîte de simulation, et ainsi des coordonnées atomiques, basé sur les idées de Berendsen [131]. La pression est couplée à la pression de référence, 1 bar, par une équation de relaxation du premier ordre. Cette méthode de faible couplage, ramène donc la pression moyenne à la valeur de référence mais pas forcement la pression instantanée. L'ajustement dépend de la compressibilité du système, i.e. la protéine en solution, définie comme celle de l'eau (44,6.10^{-6} bar^{-1}) et le temps de relaxation (2 ps).

Les conditions périodiques aux limites permettent également d'utiliser la sommation de Particle Mesh Ewald (PME) pour le traitement des interactions électrostatiques (voir Annexe C.5). Une distance limite (cutoff) de 10 Å est utilisée pour limiter la sommation classique des interactions entre charges ponctuelles.

C.7. Champ de forces de Amber et de CHARMM

L'énergie potentielle du champ de force de Amber et de CHARMM est donnée par l'expression suivante :

$$\begin{aligned} U_{tot} = & \sum_{liaisons} K_r(r - r_{eq})^2 + \sum_{\substack{angles \\ valence}} K_\theta(\theta - \theta_{eq})^2 + \\ + & \sum_{\substack{angles \\ dièdres}} K_\omega \left[1 + cos(n\omega - \gamma)\right] + \sum_{\substack{angles \\ dièdres\, impropres}} K_\varphi(\varphi - \varphi_{eq})^2 \\ + & \sum_{\substack{pairs \\ non-liés}} \frac{q_i q_j}{4\pi\varepsilon_0 r} + \sum_{\substack{pairs \\ non-liés}} \epsilon_m \left[\left(\frac{r_m}{r}\right)^{12} - 2\left(\frac{r_m}{r}\right)^{6} \right] \end{aligned}$$

où les quatre premiers termes représentent l'énergie des interactions liées et les deux derniers l'énergie des interactions non-liées.

Le premier terme représente la déformation d'une liaison. Elle est décrite par un oscillateur harmonique où K_r est la constante de force (donnée en kcal.mol^{-1}.Å^{-2}) et r_{eq} la distance d'équilibre (donnée en Å). Le deuxième terme représente la déformation d'un angle de valence entre atomes liés par liaisons covalentes. Elle est décrite par un oscillateur harmonique où K_θ est la constante de force (donné en kcal.mol^{-1}.radian^{-2}) et θ_{eq} l'angle d'équilibre (donné en radian).

Le troisième terme représente la déformation d'un angle dièdre. Elle est décrite par une fonction cosinus où V_ω est la mi-hauteur de la barrière de torsion (donnée en kcal.mol^{-1}) autour de la liaison centrale de l'angle dièdre ω (donné en radian), n est la multiplicité de la fonction et γ est la phase (donné en radian). La déformation des angles impropres est également décrite par

le terme des angles dièdres. Le quatrième terme représente la déformation d'un angle dièdre impropre. Elle est décrite par un oscillateur harmonique où K_φ est la constante de force (donné en kcal.mol^{-1}.radian^{-2}) et φ_{eq} l'angle d'équilibre (donné en radian). Dans Amber, la déformation d'un angle dièdre impropre peut également être décrite par le potentiel des angles dièdres (le troisième terme).

Le cinquième terme représente les interactions électrostatiques entre atomes non-liés. Elle sont décrites par la loi de Coulomb où q est la charge atomique, r la distance entre les deux noyaux atomiques et ε_0 la permittivité diélectrique dans le vide. Le dernier terme représente les interactions van der Waals entre atomes non-liés. Elles sont décrites par un potentiel de Lennard-Jones où r est la distance entre les deux noyaux atomiques et $-\epsilon_m$ est le minimum du potentiel. Dans Amber, r_m correspond à la distance à laquelle le potentiel atteint le minimum ($-\epsilon_m$) et dans CHARMM, r_m correspond à la distance à laquelle le potentiel coupe l'axe des abscisses.

D. NMA

D.1. Les modes normaux de vibration

L'analyse de modes nomaux (NMA) permet d'étudier les mouvements lents de macromolécules biologiques. L'analyse est faite sur un minimum de la surface d'énergie potentielle (PES) de la molécule. Le puits de la PES est approché par un potentiel harmonique U autour du minimum :

$$U(r) = \frac{1}{2}(r - R) \cdot K(R) \cdot (r - R) \tag{D.1}$$

où R et r sont des vecteur en 3N dimensions (N est le nombre d'atomes) qui décrivent la conformation d'équilibre du minimum et la conformation courante en coordonnées cartésiennes, respectivement. La matrice de constante de force (la matrice Hessienne) K décrit la forme du puits et est définie par la dérivée seconde du potentiel :

$$K = \left[\frac{\partial^2 U}{\partial r_i \partial r_j}\right] r = R_{min} \tag{D.2}$$

Les vecteurs de mode normal sont définis par les vecteurs propres e_i de la matrice K, qui est définie par :

$$K \cdot e_i = \lambda_i e_i \quad i = 1, ..., 3N \tag{D.3}$$

où λ_i sont les valeurs propres de la matrice K. Les modes normaux sont indépendant et on peut ainsi récrire le potentiel harmonique sous la forme suivante :

$$U(c) = \frac{1}{2} c \cdot \Lambda \cdot c \tag{D.4}$$

où la matrice d'interaction Λ est diagonale et les éléments diagonaux sont les valeurs propres λ_i. Les nouvelles coordonnées c sont données par

$$c_i = (r - R)e_i \tag{D.5}$$

Les mouvements collectifs de la protéine correspond à une coût énergétique faible et donc à des petites valeurs propres. Lors d'une analyse de modes normaux d'une molécule isolée, les six premières valeurs propres sont nulles puisqu'elles décrivent la translation et la rotation d'ensemble - des transformations sans coût énergétique.

Afin de décrire les mouvements des atomes sous le potentiel harmonique, on introduit des coordonnées réduites par leur masse :

$$\begin{aligned} \tilde{r} &= \sqrt{M} \cdot r \\ \tilde{R} &= \sqrt{M} \cdot R \\ \tilde{K} &= \sqrt{M}^{-1} \cdot K \cdot \sqrt{M}^{-1} \end{aligned} \tag{D.6}$$

où M est une matrice diagonale de dimension 3N×3N dont les éléments diagonaux sont la masse correspondant à chaque coordonnée. Les lois de mouvements d'un potentiel harmonique sont ainsi donnés par :

$$\ddot{\tilde{r}} = \tilde{K} \cdot (\tilde{r} - \tilde{R}) \tag{D.7}$$

Les solutions, les modes normaux de vibration, ont la forme suivante :

$$\tilde{r}(t) = \tilde{R} + \tilde{A}_i cos(\omega_i t + \delta_i) \quad i = 1, ..., 3N \tag{D.8}$$

où δ_i est un facteur arbitraire de la phase et où $\tilde{A}_i$ et ω_i sont les vecteurs et les valeurs propres de la matrice Hessienne K réduit par la masse :

$$\tilde{K} \cdot \tilde{A}_i = \omega_i \tilde{A}_i \tag{D.9}$$

Les vecteurs de mode normal représentent la direction dans laquelle les atomes se déplacent et à quelle ampleur par rapport aux autres atomes. En revanche, un vecteur de mode normal ne décrit pas l'ampleur absolue du déplacement d'un atome.

Dans le cas d'une protéine il existe un très grand nombre de minima locaux. Le potentiel harmonique est une approximation pour une conformation spécifique de la protéine et uniquement valable pour des petits mouvements autour de ce minimum local. Les mouvements qui séparent les minima locaux, e.g. des rotations de chaînes latérales, sont très différents de ceux qui caractérisent les mouvements de grand ampleur, e.g. des mouvements de domaines des protéines. On peut donc penser que les mouvements lents dans un minimum local soient proches de ceux du potentiel global.

D.2. Application sur la GFP

Dans l'analyse des modes normaux de la GFP, nous avons travaillé dans la suite CHARMM [124].

Le fichier structural de la protéine (*Protein Structure File* - PSF) est construit à partir du champ de forces CHARMM27 et de paramètres spécifiques pour le chromophore, non reconnu par la champ de forces. Nous avons utilisé les charges développées pour le chromophore de la GFP, et adaptées pour le champ de forces de CHARMM, par Reuter et al. [132] et les paramètres de liaisons, d'angles de valence, d'angles dièdres et d'angles impropres que nous avons développés

pour le chromophore de la GFP et utilisé dans le champs de force de Amber. Tous ces paramètres sont adaptés pour l'état fondamental du chromophore.

Nous avons utilisé la struture cristalline [43], contenant 95 molécule d'eau co-cristallisées, comme coordonnées de départ et ensuite effectué une minimisation du système. Le potentiel analytique du champ de force de CHARMM est décrit dans le section C.7. Les cordonnées sont d'abord fixées par des contraintes harmoniques avec une constante de force de 250 kcal/mol. Les contraintes sont diminuées à 100, 50, 25, 10 et 5 kcal/mol au cours de la procédure de minimisation. Dans ces minimisation contraintes la critère de convergence est choisie haut (1.0) et la procédure et donc très courte (maximum 10 cycles de minimisation). Les contraintes sont ensuite enlevées et et le système est minimisé avec une critère de convergence plus stricte (0.01). Dans cette première partie de minimisation nous utilisons l'algorithme *Conjugate Gradient* (CONJ). Dans la dernière minimisation, avec une critère de convergence encore plus stricte (0.00001) nous utilisons l'algorithme *Adopted Basis Newton-Raphson* (ABNR). Le calcul de modes normaux est ensuite effectué sur cette structure minimisé, qui correspond à un minimum local proche de la structure cristalline.