



**HAL**  
open science

# Études structurales de la dynamique de protéines fluorescentes vertes et jaunes utilisées en imagerie cellulaire

Damien Clavel

► **To cite this version:**

Damien Clavel. Études structurales de la dynamique de protéines fluorescentes vertes et jaunes utilisées en imagerie cellulaire. Cristallographie. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLS491 . tel-01466736

**HAL Id: tel-01466736**

**<https://theses.hal.science/tel-01466736>**

Submitted on 13 Feb 2017

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

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

NNT : 2016SACLS491

THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE  
À L'UNIVERSITÉ PARIS-SUD

CO-DIRECTION ENTRE  
LE LABORATOIRE DE CHIMIE PHYSIQUE D'ORSAY  
ET  
L'INSTITUT DE BIOLOGIE STRUCTURALE DE GRENOBLE

Ecole doctorale n°571  
Ecole Doctorale Sciences Chimiques : Molécules, Matériaux,  
Instrumentation et Biosystèmes

Spécialité chimie

par

**M. Damien Clavel**

Etudes structurales de la dynamique de protéines fluorescentes  
vertes et jaunes utilisées en imagerie cellulaire

Thèse présentée et soutenue à Orsay, le 20 décembre 2016.

Composition du Jury :

Mme	ISABELLE DEMACHY	Pr	Université Paris-Sud	(Présidente du jury)
Mme	CLAUDINE MAYER	Pr	Université Paris Diderot	(Rapporteur)
Mme	SOPHIE SACQUIN-MORA	CR	IBPC	(Rapporteur)
Mme	ISABELLE NAVIZET	Pr	Université Paris-Est	(Examinatrice)
M.	SERGE PIN	DR	CEA Saclay	(Examinateur)
Mme	HÉLÈNE PASQUIER	MCF	Université Paris-Sud	(Directrice de thèse)
M.	ANTOINE ROYANT	CR	Institut de Biologie Structurale	(Co-directeur de thèse)



*A mes grands-parents.*

# Remerciements

Je tiens tout d'abord à remercier mes directeurs de thèse qui ont œuvré à leur manière tout au long de cette aventure avec Hélène Pasquier, pour l'opportunité de réaliser un doctorat, et Antoine Royant, pour sa patience et son accueil dans sa jeune équipe.

Je tiens à remercier chaleureusement mes rapporteurs, Claudine Mayer et Sophie Sacquin-Mora pour leurs commentaires ainsi que l'ensemble des membres de mon jury qui ont permis d'améliorer la qualité de mon manuscrit.

J'ai eu la chance d'effectuer ma thèse à cheval sur deux laboratoires ; le Laboratoire de Chimie Physique à Orsay et l'Institut de Biologie Structurale à Grenoble. Ce grand écart géographique n'a pas été facile à gérer tous les jours, mais j'en ai retiré une richesse tant dans les rencontres humaines que scientifiques, dont une activité intense autour des protéines fluorescentes, avec Marie, Virgile, Romain, Jérémy, Mariam, Cornélia et Joyce.

Je remercie tout particulièrement Jacqueline Ridard et Bernard Levy pour l'accueil qu'ils m'ont réservé au sein du groupe ThéoSim. Leurs contributions ont été essentielles à ce travail tant pour les champs de force des chromophores que pour les cours intensifs de codage.

Je remercie par la même occasion mes différents voisins de paillasse que j'ai pu avoir dans l'équipe *icOS*, avec le *feu follet* Céline et l'*ours* Sylvain. Je n'oublie pas Guillaume, alias le *tonton flingueur* de thèse, qui a passé du temps au chevet de *mNeonGreen* et de mon manuscrit.

Je remercie tous ceux qui m'ont permis de décompresser au cours de ces trois ans. Il y a bien sûr la bande des "vrais mecs" avec qui nous avons partagé quelques soirées mémorables, ainsi que les genevois. Bien entendu, j'ai grandement apprécié de voir les américains Trevor, Brittany et Rachel qui ont pas mal voyagé pour échouer à Grenoble. I would like to acknowledge with a

---

special citation Vitalijs and Alexandra for their spirit of adventure and their hospitality. Et puis tous les strasbourgeois, les habitués de la rue Sédillot et de la place des Halles, qui ont contribué à leur manière à ma thèse. Petit clin d'oeil à Dr. Jérémy qui m'a montré pour la première fois une thèse, la sienne. Sans oublier, un détour par les montagnes avec un remerciement sincère et profond pour Clauclau, Amande, Ju et Guigui pour leur camaraderie inoxydable, et ce malgré la distance.

Sans oublier les membres de ma famille qui ont contribué à mon travail grâce leur présence et leur soutien sans faille.

Enfin, j'exprime toute ma gratitude à Joan qui m'a soutenu, épaulé et aidé tout au long de cette aventure et sans qui je ne me serai jamais lancé sur les routes escarpées du doctorat.



# Table des matières

<b>Remerciements</b>	<b>i</b>
<b>Table des matières</b>	<b>viii</b>
<b>Abréviations</b>	<b>x</b>
<b>Table des figures</b>	<b>xx</b>
<b>Liste des tableaux</b>	<b>xxiii</b>
<b>I Avant-propos</b>	<b>1</b>
<b>II Les protéines fluorescentes</b>	<b>4</b>
A Historique . . . . .	4
A.1 Découverte de AvGFP . . . . .	4
A.2 Découverte d'homologues de AvGFP . . . . .	5
A.3 Ingénierie des protéines fluorescentes . . . . .	6
B Phylogénèse des protéines fluorescentes . . . . .	8
B.1 Le gène GFP . . . . .	8
B.2 Provenance des protéines fluorescentes étudiées en cours de thèse . . . . .	9
B.2.a La méduse <i>Aequorea victoria</i> . . . . .	10
B.2.b Les vers de mer du genre <i>Branchiostoma</i> . . . . .	10
C Structure tridimensionnelle des protéines fluorescentes . . . . .	12
C.1 Structure tertiaire . . . . .	12
C.2 Le chromophore . . . . .	15
D La GFP sauvage ( <i>Green Fluorescent Protein</i> ) . . . . .	17
D.1 Photophysique de AvGFP . . . . .	17
D.2 Le transfert de proton à l'état excité chez AvGFP . . . . .	19
D.3 Le glutamate 222 : accepteur final de proton . . . . .	21
D.4 Modification du transfert de proton à l'état excité par mutagenèse . . . . .	21
E Les protéines fluorescentes jaunes . . . . .	24
E.1 Protéines fluorescentes jaunes dérivées de AvGFP . . . . .	24
E.2 Protéines fluorescentes jaunes d'autres organismes . . . . .	27
F Protéines fluorescentes du genre <i>Branchiostoma</i> . . . . .	28
F.1 <i>lanYFP</i> de <i>Branchiostoma lanceolatum</i> . . . . .	28
F.2 Autres protéines fluorescentes du genre <i>Branchiostoma</i> . . . . .	30
G Applications méthodologiques des protéines fluorescentes en biologie cellulaire . . . . .	33
G.1 Applications des protéines fluorescentes en microscope de fluorescence . . . . .	33
G.2 Principe du FRET . . . . .	34
G.3 Donneur et accepteur de FRET : un cahier des charges . . . . .	36
G.4 Donneurs et accepteurs de FRET . . . . .	36



	G.4.a	Donneurs de FRET à la fluorescence cyan . . . . .	36
	G.4.b	Accepteurs de FRET pour dérivés de la <i>Cyan Fluorescent Protein</i> . . . . .	38
<b>III</b>		<b>Etude par dynamique moléculaire des réseaux de liaisons hydrogène de AvGFP</b>	<b>39</b>
A		Les liaisons hydrogène dans les systèmes biologiques . . . . .	39
	A.1	Rôle et définition . . . . .	39
	A.2	Réseaux de liaisons hydrogène et chemin de translocation de protons . .	41
B		Recherche des réseaux de liaisons hydrogène dans les protéines . . . . .	43
	B.1	Contexte . . . . .	43
	B.2	Algorithme de recherche des réseaux de liaisons hydrogène au sein d'une protéine . . . . .	45
	B.2.a	Structure générale de l'algorithme . . . . .	45
	B.2.b	Entrée des données . . . . .	46
	B.2.c	Typologie des atomes : nature donneur/accepteur de proton .	47
	B.2.d	Recherche des réseaux de liaisons hydrogène . . . . .	48
	B.2.d i	Pavage tridimensionnel de la boîte de simulation . .	48
	B.2.d ii	Recherche des liaisons hydrogène . . . . .	50
	B.2.d iii	Agrégation des liaisons hydrogène en réseau de liaisons hydrogène . . . . .	51
	B.2.d iv	Réseau global de liaisons hydrogène connecté à un atome d'intérêt . . . . .	51
	B.2.d v	Fichiers résultats . . . . .	51
	B.3	Représentations graphiques des réseaux de liaisons hydrogène . . . . .	52
	B.3.a	Représentation sagittale (2D) . . . . .	52
	B.3.b	Représentation 3D . . . . .	53
	B.3.c	Représentation matricielle . . . . .	54
C		Analyse des dynamiques moléculaires . . . . .	58
	C.1	Sélection des structures cristallographiques . . . . .	58
	C.1.a	Structures cristallographiques . . . . .	58
	C.1.b	Différences structurales majeures au sein des protéines étudiées	62
	C.2	Choix des états de protonation et des conformations locales de départ .	64
	C.2.a	Etats de protonation . . . . .	64
	C.2.b	Choix des conformations locales . . . . .	66
	C.3	Stabilité de la structure protéique au cours des simulations . . . . .	66
	C.3.a	RMSd . . . . .	66
	C.3.b	RMSf . . . . .	69
	C.4	Première analyse des dynamiques moléculaires : aspects structuraux . .	70
	C.4.a	Phénol du chromophore . . . . .	70
	C.4.b	Résidu 65 . . . . .	71
	C.4.c	Histidine 148 . . . . .	74
	C.4.d	Résidu 203 . . . . .	75
	C.4.e	Résidu 205 . . . . .	76
	C.4.f	Résidu 222 . . . . .	77
D		Dynamique des réseaux de liaisons hydrogène et chemins de translocation de proton . . . . .	79
	D.1	Performances de l'algorithme . . . . .	79
	D.1.a	Suivi du pavage tridimensionnel . . . . .	79

D.1.b	Répartition des molécules d'eau par sphère . . . . .	81
D.1.c	Distribution des liaisons hydrogène . . . . .	82
D.1.d	Distribution des réseaux de liaisons hydrogène . . . . .	83
D.2	Réseau global de liaisons hydrogène connecté au phénol du chromophore	85
D.2.a	AvGFP (1W7S) . . . . .	85
D.2.a i	Réseau global de liaisons hydrogène . . . . .	85
D.2.a ii	Chemin de translocation de protons impliquant le phénol du chromophore . . . . .	88
D.2.b	AvGFP-S205V (2QLE) . . . . .	93
D.2.b i	Réseau global de liaisons hydrogène . . . . .	93
D.2.b ii	Chemin de translocation de protons impliquant le phénol du chromophore. . . . .	95
D.2.c	AvGFP-T203V-S205A (4OGS) . . . . .	102
D.2.c i	Réseau global de liaisons hydrogène . . . . .	102
D.2.c ii	Chemin de translocation de protons impliquant le phénol du chromophore. . . . .	104
D.2.d	AvGFP-T203V-S205V (4OGS-S205V) . . . . .	111
D.2.d i	Réseau global de liaisons hydrogène . . . . .	111
D.2.d ii	Chemin de translocation de protons impliquant le phénol du chromophore . . . . .	112
D.2.e	AvGFP-S65T (1C4F) . . . . .	115
D.2.e i	Réseau global de liaisons hydrogène . . . . .	115
D.2.e ii	Chemin de translocation de protons impliquant le phénol du chromophore . . . . .	117
E	Discussion . . . . .	123
F	Conclusion . . . . .	127

#### IV Etude structurale de protéines fluorescentes de *Branchiostoma lanceolatum* par cristallographie aux rayons X 128

A	Contexte structural . . . . .	128
A.1	Diversité des organismes contenant des protéines fluorescentes . . . . .	128
A.2	Structures de protéines fluorescentes du phylum <i>Arthropoda</i> . . . . .	129
A.3	Structures de protéines fluorescentes du phylum <i>Chordata</i> . . . . .	130
A.4	Objectifs . . . . .	132
B	Structure de protéines fluorescentes de <i>Branchiostoma lanceolatum</i> . . . . .	135
B.1	Cristallisation . . . . .	135
B.1.a	Cristallisation de mNeonGreen à pH acide . . . . .	135
B.1.b	Cristallisation de lanYFP à pH physiologique . . . . .	136
B.1.c	Cristallisation de mNeonGreen à pH physiologique . . . . .	136
B.2	Structures cristallographiques de lanYFP et mNeonGreen . . . . .	139
B.2.a	Collecte des données de diffraction et résolution des structures	139
B.2.b	Structure de mNeonGreen à pH acide . . . . .	141
B.2.c	Structure de lanYFP à pH physiologique . . . . .	142
B.2.d	Structure de mNeonGreen à pH physiologique . . . . .	144
B.2.e	Comparaison des structures de lanYFP et mNeonGreen à pH physiologique . . . . .	144
B.2.f	Comparaison des structures de mNeonGreen . . . . .	150
B.3	Domages spécifiques aux rayons X subis par mNeonGreen . . . . .	151

	B.3.a	Spectroscopie d'absorption UV-visible <i>in crystallo</i> . . . . .	151
	B.3.b	Spectroscopie Raman sur ligne de lumière . . . . .	153
	B.3.c	Protocole de collecte de données de diffraction . . . . .	155
	B.3.d	Discussion . . . . .	160
C		Analyse de l'environnement du chromophore par calcul d'énergie d'interaction	162
	C.1	But de l'analyse . . . . .	162
	C.2	Principe de l'analyse . . . . .	162
	C.3	Principe d'extraction des énergies d'interaction de van der Waals . . . . .	163
	C.4	Interprétation . . . . .	164
D		Conclusion . . . . .	166
<b>V Perspectives</b>			<b>167</b>
<b>VI Méthodes</b>			<b>170</b>
A		Expression et purification . . . . .	170
	A.1	Expression de protéine recombinante . . . . .	170
		A.1.a Description du gène des protéines fluorescentes de <i>Branchiostoma lanceolatum</i> . . . . .	170
		A.1.b Expression des protéines recombinantes . . . . .	171
	A.2	Purification de protéine recombinante . . . . .	172
	A.3	Protocole de digestion à la trypsine . . . . .	173
B		Détermination de la structure de protéines par cristallographie aux rayons X . . . . .	175
	B.1	La cristallisation des protéines fluorescentes . . . . .	175
		B.1.a Généralités concernant les cristaux et la cristallisation . . . . .	175
		B.1.b Principe de cristallisation . . . . .	176
		B.1.c Méthode de diffusion de vapeur . . . . .	177
		B.1.d Optimisation des cristaux par ensemencement homogène . . . . .	179
	B.2	Diffraction aux rayons X . . . . .	179
		B.2.a La loi de Bragg . . . . .	179
		B.2.b L'espace direct et l'espace réciproque . . . . .	180
		B.2.c Représentation de la sphère d'Ewald . . . . .	181
		B.2.d Collecte des données de diffraction aux rayons X . . . . .	181
		B.2.e Traitement des données de diffraction aux rayons X . . . . .	182
		B.2.f Calcul de la carte de densité électronique et affinement de structure . . . . .	183
		B.2.f i Le problème de la phase . . . . .	183
		B.2.f ii Le remplacement moléculaire . . . . .	183
		B.2.f iii Affinement de structure et facteurs R . . . . .	183
	B.2.g	Logiciels de cristallographie des protéines . . . . .	185
C		Spectroscopie optique . . . . .	186
	C.1	Généralités . . . . .	186
	C.2	Spectroscopie d'émission de fluorescence . . . . .	187
	C.3	Spectroscopie Raman . . . . .	187
	C.4	Protocoles expérimentaux de détermination de constantes spectroscopiques . . . . .	188
		C.4.a Détermination du coefficient d'extinction molaire . . . . .	188
		C.4.b Détermination du rendement quantique par comparaison . . . . .	188
		C.4.c Détermination de la brillance . . . . .	189

C.5	Spectroscopie optique <i>in crystallo</i> . . . . .	189
D	Simulations par dynamique moléculaire . . . . .	191
D.1	Principe . . . . .	191
D.1.a	Equations du mouvement de Newton . . . . .	191
D.1.b	Conditions aux limites périodiques . . . . .	192
D.1.c	Contrôle de température et de pression . . . . .	193
D.1.d	Champs de force . . . . .	194
D.1.d i	Définition . . . . .	194
D.1.d ii	Champ de force <i>ff99</i> de la suite de programmes <i>AMBER</i> : résidus standards . . . . .	195
D.1.d iii	Champ de force spécifique au chromophore : un résidu non-standard . . . . .	196
D.1.d iv	Le modèle de molécule d'eau . . . . .	197
D.2	De la structure cristallographique à la dynamique moléculaire . . . . .	197
D.2.a	Préparation du système . . . . .	197
D.2.a i	Etat de protonation des résidus acido-basiques de la protéine . . . . .	198
D.2.a ii	Préparation du système pour la dynamique molé- culaire . . . . .	199
D.2.b	Minimisation de l'énergie du système et mise à l'équilibre . .	199
D.2.c	Phase de production . . . . .	200
<b>Bibliographie</b>		<b>201</b>
<b>VII Annexes</b>		<b>215</b>
A	Longueur d'onde associée aux couleurs . . . . .	215
B	Compartiments cellulaires et pH chez un archétype de cellule de mammifère . .	215
C	Table des acides aminés . . . . .	216
D	Groupe d'espace . . . . .	217
E	Séquence de protéines de <i>Branchiostoma</i> . . . . .	218
F	Etude des interfaces des structures cristallines . . . . .	219
G	Considérations sur l'exploitation des cartes de différence Fourier . . . . .	221
H	Article <i>Acta Crystallographica section D: Structural Biology</i> . . . . .	222



# Abréviations

**ADN** acide désoxyribonucléique.

**CFP** *Cyan Fluorescent Protein*, protéine fluorescente cyan dérivé de AvGFP.

**EMBL** *European Molecular Biology Laboratory*, Laboratoire Européen de Biologie Moléculaire.

**ESPT** *Excited State Proton Transfer*, transfert de proton à l'état excité.

**ESRF** *European Synchrotron Radiation Source*, Installation Européenne de Rayonnement Synchrotron.

**FRET** Transfert d'énergie résonant de Förster.

**GFP** *Green Fluorescent Protein* ou AvGFP.

**HTX** *High Throughput Crystallisation Laboratory*, plateforme de cristallisation haut débit de l'EMBL.

**IBS** Institut de Biologie Structurale à Grenoble.

**IMAC** *Immobilized Metal Affinity Chromatography*, chromatographie d'affinité par chélation de métaux immobilisés.

**MR** *Molecular replacement*, remplacement moléculaire.

**NCBI** *National Center for Biotechnology Information*, centre américain pour les informations biotechnologiques.

**PDB** *Protein Data Bank*, banque de données mondiale de structures de protéine.

**ROS** *reactive oxygen species*, espèces réactives de l'oxygène.

**SDS-PAGE** *Sodium dodecyl sulfate polyacrylamide gel electrophoresis*, électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium.

**SOC** Milieu riche en glucose.

**vdW** van der Waals.

**YFP** *Yellow Fluorescent Protein*, protéine fluorescente jaune.

# Liste des figures

II.1	Couleur et longueur d'onde (voir tableau VII.A) . . . . .	5
II.2	Diversité spectrale des PF obtenue dans le laboratoire de Tsien [Wang et al., 2004, Shaner et al., 2004, Tsien, 2009b] . . . . .	6
II.3	Chronologie des événements pour le champ de recherche des protéines fluorescentes . . . . .	7
II.4	Arbre phylogénétique des organismes vivants qui possèdent un gène codant pour un domaine GFP [Alieva et al., 2008, Geer et al., 2009, Shagin et al., 2004] . Les clades regroupés sont signalés par un gros point et les genres étudiés au cours de cette thèse sont surlignés en jaune et vert. . . . .	8
II.5	Arbre phylogénétique schématique indiquant la provenance de protéines fluorescentes, (a) <i>Aequorea victoria</i> [Haddock, 2016], (b) <i>Discosoma striata</i> [Matz et al., 1999], (c) <i>Chiriduis poppei</i> [Masuda et al., 2006] et (d) <i>Branchiostoma lanceolatum</i> et <i>Branchiostoma floridae</i> [Shaner et al., 2013, Bomati et al., 2014] (adapté de [Lukyanov, 2013]) . . . . .	9
II.6	Méduse <i>A. victoria</i> [Haddock, 2016] . . . . .	10
II.7	Les photo-organes d' <i>A. victoria</i> situés sur le pourtour de son ombrelle sont stimulés par de la lumière bleue [Zimmer, 2015] . . . . .	10
II.8	planche d'anatomie de Pallas classant <i>B. lanceolatum</i> dans l'embranchement des mollusques et de Poll [Pallas, 1774, Poll, 1947] . . . . .	11
II.9	Structure tridimensionnelle de AvGFP en forme de tonneau de brins $\beta$ (vert), avec son hélice $\alpha$ coaxiale (bleu) et ses boucles (jaune) [Ormö et al., 1996, Yang et al., 1996] . . . . .	12
II.10	Alignement de séquence de protéines fluorescentes représentatives des embranchements cnidaire, arthropode et chordé réalisé avec le serveur Clustal Omega [Sievers and Higgins, 2014] et sa représentation avec le serveur ESPript 3.0 [Robert and Gouet, 2014]. . . . .	13
II.11	Mécanisme de formation du chromophore de AvGFP [Cubitt et al., 1995, Miyawaki et al., 2003] . . . . .	15
II.12	Représentation du chromophore protoné cis-plan et des angles diédraux $\tau$ et $\varphi$ .	16
II.13	Typologie des chromophores chez les PF homologues de AvGFP en fonction de leur longueur d'onde d'émission. . . . .	16
II.14	Spectre d'absorption et d'émission de fluorescence de AvGFP. Les deux bandes du spectre d'absorption dénote l'équilibre acido-basique entre la forme A, neutre, et B, anionique, du chromophore [Tsien, 1998]. . . . .	17
II.15	Processus mis en jeu lors de la photo-excitation de AvGFP (adapté de [Day and Davidson, 2014]) . . . . .	19
II.16	Chemin de translocation de proton supportant l'ESPT chez AvGFP et ses mutants [Shu et al., 2007] . . . . .	20
II.17	Représentation tridimensionnelle du chemin de translocation de proton supportant l'ESPT chez AvGFP et ses mutants . . . . .	20

II.18	Absorption et blanchiment des pics IR qui sont observés par spectroscopie TIR et qui ont été associées à l'étirement antisymétrique du carboxylate ( $1565\text{ cm}^{-1}$ carrés pleins) et à la forme protonée du groupe carboxyle du résidu Glu 222 [Stoner-Ma et al., 2005] . . . . .	21
II.19	Chromophore de type GYG présent chez les protéines fluorescentes jaunes dérivées de AvGFP montrant l'interaction de type $\pi$ - $\pi$ avec la tyrosine 203 . . . . .	24
II.20	Spectre d'absorption à différents pH (trait pleins colorés) et d'émission de fluorescence à pH 7,4 (pointillés noirs, excitation à 514 nm) pour la protéine fluorescente jaune YFP . . . . .	25
II.21	Formalisme décrivant les interfaces pour les tétramères naturelles de protéines fluorescentes, (a) protéines naturelles rouges DsRed (code PDB 1G7K), (b) représentation schématique des interactions entre monomères (adapté de [Shen et al., 2015]) . . . . .	28
II.22	Schéma de principe de la monomérisation de <i>lanYFP</i> en mNeonGreen avec la protéine fluorescente jaune dimérique <i>dlanYFP</i> . . . . .	30
II.23	Arbre phylogénétique des séquences codantes pour de PF du genre <i>Branchiostoma</i> . <i>blFP</i> : <i>B. lanceolatum</i> , <i>bflo</i> : <i>B. floridae</i> , <i>bder</i> : <i>Branchiostoma</i> [Mushegian et al., 2014, Israelsson, 2013, Putnam et al., 2008] . . . . .	31
II.24	Evolution de l'efficacité de FRET ( $E$ ) en fonction du rapport $r/R_0$ [Lakowicz, 2006] . . . . .	35
III.1	Critères qui définissent une liaison hydrogène (LH) avec une distance $d$ entre D et A et l'angle $\Omega$ ( $\widehat{DHA}$ ). Les valeurs les plus souvent utilisées sont telles que $d$ est inférieur à $3,2\text{ \AA}$ et que $\Omega$ est compris entre $120^\circ$ et $180^\circ$ . . . . .	40
III.2	Structure de l'algorithme de recherche des liaisons hydrogène, de leur agrégation et de leur analyse. $n$ correspond au numéro séquentiel de l'instant analysé et $N$ nombre total d'instantanées constituant la trajectoire de dynamique moléculaire à analyser. . . . .	46
III.3	Représentation schématique de la typologie des sous boîtes qui composent le pavage tridimensionnel mis en place pour l'algorithme. Les atomes de la protéine sont présents dans les sous-boîtes de type 1 et 2. Les sous boîtes de type 3 et 4 sont uniquement constituées de molécules d'eau. La typologie des sous-boîtes est déterminée par leur composition et leur voisinage. . . . .	50
III.4	Représentation sagittale des réseaux de LH connectés à l'atome OH (donneur de proton) du phénol du chromophore dans le cas de la AvGFP. Ces réseaux ont été obtenus lors de l'analyse par l'algorithme d'une dynamique moléculaire de 30 ns (soit $N = 15000$ instantanés) effectuée sur la structure protéique 1W7S. L'atome d'intérêt sélectionné, ici Chrom-66/OH, est mis en exergue (entouré d'un ovale rouge) grâce à la connaissance de son numéro séquentiel. Seules les LH avec une densité de probabilité supérieure à $T = 10\%$ sont représentées. . . . .	53
III.5	Représentation tridimensionnelle du réseau de LH connecté à l'atome OH (donneur de proton) du phénol du chromophore dans le cas de la AvGFP (structure 1W7S). Seuls les résidus dont au moins un atome est impliqué dans une LH sont représentés. Les LH sont représentées par des traits jaunes dont l'épaisseur est corrélée à leur densité de probabilité calculée sur une MD de 30 ns ( $N=15000$ ). Ici, seules les LH avec une densité de probabilité supérieure à $T=10\%$ sont représentées. . . . .	54



III.6	Représentation matricielle du réseau global de liaisons hydrogène connecté à l'atome OH, donneur de proton, du phénol du chromophore de AvGFP, en fonction de (a) la distance $d$ , (b) de l'angle $\Omega$ . Les noms des atomes entre lesquels les LH s'établissent sont spécifiés sur l'axe des ordonnées du graphe central. Ces LH sont représentées par ordre décroissant de leur densité de probabilité. Les résultats sont issus de l'analyse d'une dynamique moléculaire d'une durée de 30 ns effectuée à partir de la structure protéique 1W7S (N = 15000 instantanés, T = 10%). En moyenne, le réseau global connecté à OH du phénol comporte 12 liaisons hydrogène par instantané. De plus, la liaison hydrogène la plus probable présente une densité de probabilité de plus de 90% dans cette protéine. . . . .	56
III.7	Représentation matricielle du chemin de translocation de proton partant de l'OH du phénol du chromophore de AvGFP et arrivant à l'atome d'oxygène OE2, de l'accepteur final Glu222. Ce chemin fait intervenir 3 LH, représentées dans le sens de la translocation, de $D_i$ à $A_f$ , impliquant le groupe OH du chromophore, l'oxygène de la molécule d'eau 215, l'hydroxyle de Tyr145 et enfin OE2 du Glu222. Il s'agit d'un chemin de translocation de proton alternatif au réseau usuel peu probable. Structure protéique : 1W7S (N = 15000 instantanés, T = 10%). . . . .	57
III.8	Rotamère t de la valine choisi pour la mutation <i>in silico</i> de la position 205 à partir de la structure cristallographique de 4OGS pour générer le mutant AvGFP-T203V-S205V. . . . .	60
III.9	Différences structurales entre les structures choisies pour l'étude des réseaux de liaisons hydrogène AvGFP(vert), AvGFP-S205V (orange), AvGFP-S65T (jaune), AvGFP-T203V-S205V (blue), AvGFP-T203V-S205A (mauve). Les valeurs de RMSd ont été calculées sur la base d'un alignement tridimensionnelle avec la structure de 1W7S et des atomes de la chaîne principale (CA,C,N,O) et le terme u.a. signifie unité asymétrique. . . . .	61
III.10	Angle dièdre de $\chi_{65}$ (OG1/C1/CA1/CB1) permettant de décrire les conformations 65OH-down et 65OH-up . . . . .	63
III.11	Angle dièdres $\chi_{203}$ et $\chi_{205}$ décrit par les atomes OG1/C $\beta$ /C $\alpha$ /C dans les résidus thréonine et sérine . . . . .	63
III.12	Valeurs du $pK_{1/2}$ calculées par le logiciel $H^{++}$ pour le groupe cible de protéines étudiées au cours de ces travaux avec AvGFP en vert, AvGFP-S65T en orange, AvGFP-S205V en jaune et AvGFP-T203V-S205A en bleu . . . . .	65
III.13	RMSd des structures secondaires feuillet $\beta$ (en rouge) et hélices $\alpha$ (en bleu) avec leur moyenne respective représentée en pointillés, (a) AvGFP, (b) AvGFP-S205V, (c) AvGFP-S65T, (d) AvGFP-T203V-S205V, (e) AvGFP-T203V-S205A, (f) contribution de l'hélice $\alpha_1$ à la valeur du RMSd des hélices de AvGFP-T203V-S205A . . . . .	68
III.14	Facteur de température B en fonction du numéro de résidu au cours des dynamiques de AvGFP (trait vert), AvGFP-S65T (trait jaune), AvGFP-S205V (trait orange), AvGFP-T203V-S205A (trait bleu) et AvGFP-T203V-S205V (trait mauve). Les structures secondaires des mutants de AvGFP sont localisées avec un fond vert pour les brins $\beta$ et un fond bleu pour les hélices $\alpha$ . La position du chromophore est représentée par les pointillés rouges. . . . .	69

III.15	Trajectoire de l'angle dièdre $\chi(\text{HH/OH/CZ/CE1})$ du phénol du chromophore au cours des dynamiques moléculaires de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve). . . . .	70
III.16	Trajectoires de l'angle dièdre $\chi_{65}$ caractérisant la conformation de l'hydroxyle en position 65 dans AvGFP(vert), AvGFP-S65T (jaune) et AvGFP-S205V (orange). Les conformations 65OH- <i>down</i> , 65OH- <i>up</i> , et 65OH- <i>opposée</i> de l'hydroxyle sont respectivement caractérisées par $\chi_{65} \sim +65^\circ$ , $\chi_{65} \sim -65^\circ$ et $\chi_{65} \sim -180^\circ$ ou $\chi_{65} \sim +180^\circ$ . . . . .	72
III.17	Trajectoires de l'angle dièdre $\chi_{65}$ caractérisant la conformation de l'hydroxyle en position 65 pour les simulations de AvGFP-T203V-S205A (a) et AvGFP-T203V-S205V (b). . . . .	72
III.18	Trajectoires de l'angle dièdre $\chi_{148}$ caractérisant la conformation de la chaîne latérale de l'histidine 148 au cours des dynamiques de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-S205A (bleu) et AvGFP-T203V-S205V (mauve). . . . .	74
III.19	Trajectoires de l'angle dièdre $\chi_{T203}$ (a) pour AvGFP(vert), AvGFP-S65T (jaune) et AvGFP-S205V (orange) ainsi que $\chi_{V203}$ (b) pour AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve) en fonction du temps des dynamique. . . . .	75
III.20	Trajectoires de l'angle dièdre $\chi_{205}$ en fonction du temps et du résidu en position 205, (a) sérine avec les mutants AvGFP (vert) et AvGFP-S65T (jaune) et (b) valine avec AvGFP-S205V (orange) et AvGFP-T203V-S205V (mauve). . . . .	76
III.21	Trajectoires de l'angle dièdre $\chi_{222}$ caractérisant l'échange des atomes OE1 et OE2 du Glu222 au cours des dynamiques, (a) AvGFP(vert) et AvGFP-S65T (jaune), (b) AvGFP-S205V (orange), (c) AvGFP-T203V-S205A (bleu) et (d) AvGFP-T203V-S205V (mauve). . . . .	78
III.22	Variation du nombre de sous-boîtes des sphères internes et de bord de la protéine et de la première sphère externe à la protéine lors du pavage de la boîte de simulation au cours de la dynamique moléculaire de AvGFP (1W7S) (trait fin gris). La moyenne glissante des sous-boîtes (100 instantanés) est représentée en trait gras avec de couleurs différentes pour les sphères interne (cyan), de bord (bleu) et externe (mauve). . . . .	80
III.23	Distribution des molécules d'eau entre les différentes catégories de sphères considérées pour la recherche des LH le long de la dynamique pour AvGFP (1W7S) : sphères interne (cyan) et de bord (bleu) de la protéine, première sphère d'hydratation (mauve) . . . . .	82
III.24	Distribution des LH, connectées au groupe hydroxyle du phénol du chromophore, détectées par l'algorithme lors de l'analyse de la dynamique de AvGFP (1W7S). Les deux distributions obtenues pour une distance d'environ 2,2 Å et des angles de 110° et 125° correspondent aux LH intra-résidus (Asp, Glu, His). La distribution la plus large centrée à une distance de 2,8 Å et un angle d'environ 165° correspond aux LH entre résidus, entre un résidu et une molécule d'eau ou encore entre deux molécules d'eau. . . . .	83
III.25	Trajectoires du nombre de réseaux de LH au cours de la simulation pour les simulations de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve) . . . . .	84

III.26	Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP (1W7S). Seules les LH dont la densité de probabilité est supérieure à $T = 0,5\%$ sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge. . . . .	87
III.27	Distribution de distances de la liaison hydrogène entre Glu222/OE2 et Ser205/OG au cours de la dynamique de AvGFP (1W7S). . . . .	88
III.28	Représentation matricielle du chemin de translocation usuellement décrit dans la littérature entre Chrom.66/OH et Glu222/OE2 dans le cas de AvGFP (1W7S)	89
III.29	Représentation matricielle du chemins de translocation alternatif entre Chrom.66/OH et Glu222/OE2 dans le cas de AvGFP (1W7S) . . . . .	90
III.30	Représentation tridimensionnelle (a) du chemin supportant l'ESPT usuellement décrit dans la littérature, (b) rupture transitoire du chemin de translocation ( $t = 2,09$ ns) et (c) chemin alternatif utilisant Tyr145/OH comme relais pour les chemins de translocation entre Chrom.66/OH et Glu222/OE2 retrouvés avec l'algorithme ( $TH > 0,5\%$ ), . . . . .	91
III.31	Schéma récapitulant les chemins de translocation entre Chrom66./OH et Glu222/OE2 trouvés dans la dynamique de AvGFP (1W7S), (a) Densité de probabilité des différentes LH constituant les chemins qui sont connectés au phénol du chromophore (Chrom.66/OH) et (b) Réarrangement du résidu Tyr145 pour l'établissement du chemin alternatif. Les instantanés gris et vert correspondent respectivement à $t = 2$ ns et $t = 2,5$ ns. . . . .	92
III.32	Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-S205V (2QLE). Seules les liaisons hydrogène dont la densité de probabilité est supérieure à $T = 1\%$ sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge. . . . .	94
III.33	Représentations matricielles du chemin de translocation principal entre Chrom.66/OH et Glu222, (a) chemin entre Chrom.66/OH et Glu222/OE1 et (b) a) Chemin entre Chrom.66/OH et Glu222/OE2 pour la dynamique de AvGFP-S205V (2QLE). . . . .	96
III.34	Longueur du chemin de translocation de proton entre Chrom66./OH et Glu222/OE2 dans AvGFP-S205V au cours de la simulation. Les zones de temps de rupture du chemin sont représentées en cyan et les temps de début et fin de la rupture sont précisés. . . . .	97
III.35	Représentation tridimensionnelle du chemin de translocation principal entre Chrom.66/OH et Glu222 pour AvGFP-S205V. . . . .	97
III.36	Représentation tridimensionnelle du chemin de translocation secondaire entre Chrom66/OH et Glu222 faisant intervenir une molécule d'eau cristallographique pour la dynamique de AvGFP-S205V (2QLE), (a) chemin entre Chrom.66/OH et Glu222/OE1, (b) chemin entre Chrom.66/OH et Glu222/OE2 . . . . .	99

III.37	Longueur des chemins de translocation entre Chrom.66/OH et Glu222/OEx dans AvGFP-S205V (2QLE). Points rouges : chemin prépondérant avec Glu222/OE2 comme accepteur final (longueur moyenne du chemin, $\langle l_{chemin} \rangle = 8,45 \pm 0,25$ Å). Points oranges : même chemin avec Glu222/OE1 comme accepteur final ( $\langle l_{chemin} \rangle = 8,43 \pm 0,25$ Å). Les chemins secondaires impliquant la molécule d'eau issue de la dynamique, Wat4525, sont représentés en bleu et violet suivant la nature de l'atome accepteur final, respectivement Glu222/OE1 ( $\langle l_{chemin} \rangle = 8,43 \pm 0,08$ Å) ou Glu222/OE2 ( $\langle l_{chemin} \rangle = 8,34 \pm 0,23$ Å). . . . .	100
III.38(a)	Densité de probabilité des chemins de translocation entre Chrom.66/OH et Glu222/OEx obtenus lors de l'analyse de la dynamique moléculaire de AvGFP-S205V. (b) Echange de molécules d'eau observé à 28,8 ns de la dynamique entre Wat62 (jaune) et Wat4525 (violet) dans AvGFP-S205V. . . . .	101
III.39	Coupe sagittale présentant le réseau global de LH connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-T203V-S205A (4OGS). Seules les LH dont la densité de probabilité est supérieure à $T = 2\%$ sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge. . . . .	103
III.40	Composition des chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx dans AvGFP-T203V-S205A (4OGS) . . . . .	104
III.41	Trajectoire des longueurs des 7 types de chemin de translocation observés entre Chrom.66/OH et Glu222/OEx dans la dynamique de AvGFP-T203V-S205A (4OGS). La longueur des chemins en fonction du temps permet de les classer selon le nombre de LH qui les composent avec 2 LH, les types 6 (cyan),8 (bleu),9 (bleu foncé), 3 LH les types 4 (rouge),5 (magenta),7 (mauve) et 4 LH le type 1 (jaune). Les temps d'apparition des événements sont reportés en rouge (ns). . . . .	106
III.42	Passage entre les chemins de translocation, (a) de type 1 (en jaune) et (b) de type 4 ou 5 (en rouge type 4 uniquement) au cours de la dynamique de AvGFP-T203V-S205A. . . . .	107
III.43(a)	Changement conformationnel de la chaîne latérale du résidu Tyr145 permettant l'établissement des chemins de type 1, 4/5 et 6 dans AvGFP-T203V-S205A.	
(b)	Chemin de translocation de type 6 (en cyan). . . . .	108
III.44	Densité de probabilité (a) et caractéristiques (distance $d$ et angle $\Omega$ ) (b) des LH intervenant dans le chemin de translocation de type 1 trouvé dans AvGFP-T203V-S205A. Le chemin de type 1 est représenté en jaune, celui de type 1.bis en bleu. . . . .	109
III.45	Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 4, 5 et 7 trouvés dans AvGFP-T203V-S205A. Les chemins de type 4 sont indiqués en rouge, 4.bis en orange, 5 en violet, 5. bis en magenta, 7 en bleu et 7.bis en cyan. . . . .	109
III.46	Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 6 et 8 trouvés dans AvGFP-T203V-S205A. Le chemin de type 6 est représenté en cyan, le 6.bis en bleu pointillé et le 8 en magenta. . . . .	110

III.47	Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-T203V-S205V. Seules les LH dont la densité de probabilité est supérieure à $T = 0,1\%$ sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge. . . . .	112
III.48	Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 6 et 8 trouvés dans AvGFP-T203V-S205V. . . . .	113
III.49	Mécanisme possible de blocage de la chaîne latérale de Tyr145 (flèche cyan) par les résidus apolaires Val58 et Val205 en présence d'une valine en position 205 dans AvGFP-T203V-S205V. . . . .	114
III.50	Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-S65T (1C4F). Seules les LH dont la densité de probabilité est supérieure à $T = 2\%$ sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge. . . . .	116
III.51	Composition des chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx dans AvGFP-S65T (1C4F) . . . . .	117
III.52	Trajectoire des longueurs des 3 types de chemins de translocation observés entre Chrom66/OH et Glu222/OEx dans la dynamique de AvGFP-S65T (1C4F). Le chemin 1 est indiqué en jaune, les chemins de type 2 en rouge (Glu222/OE2 accepteur final) et magenta (Glu222/OE1 accepteur final) et le chemin de type 3 en bleu. Les temps d'apparition des événements sont reportés en rouge (ns). . . . .	119
III.53	Mécanisme de passage des chemins 1 à 3 avec le sens de déplacement de Wat48 numéroté et fléché ainsi que la rotation de Ser205. Instantané pris à 6,2 (doré), 7,2 (bleu léger) et 8 ns (bleu intense). . . . .	120
III.54	Mécanisme de passage des chemins 1 à 2 avec le sens de déplacement de Ser205 (magenta) et le mécanisme de rotation de Tyr145 (rouge). . . . .	121
III.55(a/b)	Densité de probabilité et (c/d) caractéristiques (longueur $d$ et angle $\Omega$ ) des LH intervenant dans les chemins de translocation trouvés dans AvGFP-S65T. Le chemin 1 est représenté en jaune, le chemin 2 en rouge, le chemin 2 bis en violet et le chemin 3 en bleu. . . . .	122
III.56	Schéma synoptique des divers chemins de translocation trouvés par l'algorithme pour AvGFP et ses mutants (chemins de la littérature en trait plein et ceux mis en évidence par l'algorithme en trait pointillé). . . . .	125
III.57	Valeur de $k_{PT}$ relative à AvGFP et densité de probabilité des chemins de translocation (Chrom-66/OH à Glu222/OEx) pour chacune des dynamiques étudiées. . . . .	126
IV.1	Structure chimique du chromophore de <i>lan</i> RFP présentant une liaison covalente entre Tyr59/C $\beta$ (vert) et Tyr62/O (bleu) [Pletnev et al., 2013] . . . . .	131
IV.2	Décomposition du processus d'évolution dirigée de <i>lan</i> YFP en mNeonGreen . . . . .	132
IV.3	Représentation tridimensionnelle du processus d'évolution de <i>lan</i> YFP en mNeonGreen décrivant les étapes de réduction du degré d'oligomérisation (étapes 1 [violet] et 3 [bleu]) et de restauration des propriétés d'émission de fluorescence (étapes 2 [jaune] et 4 [vert]). . . . .	133
IV.4	Cristaux de mNeonGreen à pH 4,50. . . . .	136

IV.5	Cristaux de <i>lanYFP</i> entremêlés. . . . .	136
IV.6	Cristaux de mNeonGreen à pH physiologique photographiés après 9 jours sur la plateforme du HTX . . . . .	137
IV.7	Cristaux de mNeonGreen obtenus avec trypsinisation de la protéine effectuée (a) avant et (b) pendant la cristallisation . . . . .	137
IV.8	Spectres d'absorption UV-visible sur cristaux de mNeonGreen à pH acide (pH 4,5 - rouge foncé) et physiologique (pH 8,0 - vert) . . . . .	138
IV.9	Arrangement hexamérique de mNeonGreen dans l'unité asymétrique des cristaux obtenus à pH acide. . . . .	141
IV.10	Comparaison structurale entre AvGFP code PDB 1EMA [Ormö et al., 1996] (bleu clair) et de mNeonGreen à pH acide (vert). . . . .	142
IV.11	Arrangement du 16-mère de <i>lanYFP</i> dans l'unité asymétrique avec zoom sur un des tétramères. . . . .	143
IV.12	Description des interfaces de <i>lanYFP</i> entre les tétramères ABCD, EFGH, IJKL et MNOP, (a) surface d'interaction (magenta/rouge pour l'interface A/B et bleu/cyan pour l'interface A/C), (b) nombre d'interaction hydrophiles à chaque interfaces A/B et A/C (vert : liaisons hydrogène, orange : ponts salins) . . . . .	144
IV.13	Interfaces (a) A/B et (b) A/C de <i>lanYFP</i> (chaîne latérale orange) représentant deux mutations cruciales pour la monomérisation de mNeonGreen (chaîne latérale vertes) . . . . .	145
IV.14	Représentation des surfaces électrostatiques (positive en rouge et négative en bleu) de <i>lanYFP</i> et mNeonGreen à l'interface A/C (cerclée de vert), (a) vue éclatée du dimère de <i>lanYFP</i> , (b) face du monomère A de <i>lanYFP</i> à l'interface avec C et (c) face du monomère de mNeonGreen . . . . .	146
IV.15	Densité électronique autour du résidu 143 chez mNeonGreen à pH physiologique. (a) Représentation classique de la densité en maille contourée à un niveau de $\sigma$ de 5,0, 2,0 et 1,0 (de haut en bas) montrant progressivement la présence de l'ion chlore (dans la carte à 5,0 $\sigma$ ), du groupement carboxylate (carte à 2,0 $\sigma$ ) et de la conformation non carboxylée de la lysine (carte à 1,0 $\sigma$ ). (b) Représentation volumique des trois cartes superposées (magenta : 5,0 $\sigma$ , vert : 2,0 $\sigma$ , bleu : 1,0 $\sigma$ ). . . . .	147
IV.16	Comparaison de l'environnement du chromophore pour <i>lanYFP</i> (orange) et mNeonGreen (vert), (a) site de fixation de l'ion chlorure de <i>lanYFP</i> et de la lysine carboxylée pour mNeonGreen à pH physiologique, (b) différence localisée sur la face opposée du chromophore. La liaison hydrogène entre Tyr102 et le groupement carbonyle de Lys56 est représentée par des pointillés jaune. L'interaction doublet non-liant / $\pi$ entre le carbonyle de Pro55 et le cycle de l'imidazolinone du chromophore est représenté par des pointillés rouges pour <i>lanYFP</i> et cyans pour mNeonGreen. . . . .	149
IV.17	Environnement du chromophore de mNeonGreen à pH acide (a) et physiologique (b) . . . . .	150
IV.18	Cristal de mNeonGreen observé sur la caméra de microdiffractomètre MD2 de la ligne de lumière de l'ESRF après une collecte de données diffraction (715 kGy) [de Sanctis et al., 2012]. Le faisceau de rayons X semble laisser une trace orange sur le cristal de mNeonGreen et bleu foncé dans le solvant, y signalant la présence d'électrons solvatés. . . . .	151
IV.19	Spectres d'absorption UV-visible des volumes de cristal de mNeonGreen irradié (orange) et non irradié (vert) par les RX . . . . .	152

IV.20	Série de spectres Raman enregistrés sur un cristal de mNeonGreen avec une dose croissante de rayons X déposée. Le spectre de différence avec le spectre enregistré à 0 kGy ( $S_{0 \text{ kGy}} - S_{x \text{ kGy}}$ ) est représenté en gris. . . . .	153
IV.21	Dose de rayons X accumulée en fonction du jeu de donnée. Les rectangles verts représentent les jeux de données exploités. . . . .	155
IV.22	Cartes de différence Fourier représentées au voisinage du chromophore entre 50 et 844 kGy. les doses de radiation sont modélisées par un niveau de contour de $7,0 \times 10^{-2} \text{ e}^{-} \cdot \text{\AA}^3$ ce qui correspond à $6,0 \sigma$ pour le jeu n°2 (rose : négatif, jaune : positif). . . . .	158
IV.23	Gain et perte d'électrons induit en $\text{e}^{-} \cdot \text{\AA}^3$ en fonction de la dose de radiations pour différents groupe de la cavité du chromophore : en rouge, les trois pics les plus forts, en vert, pics associés au pont méthylène, en bleu, les autres pics forts. . . . .	159
IV.24	Changements structuraux induits par les rayons X dans l'environnement du chromophore de mNeonGreen, avec une modélisation à un niveau de $4,0 \sigma$ après une irradiation de 412 kGy. Les flèches en bleu indiquent le déplacement du chromophore et du mouvement opposé du résidu Cys139 tandis que la flèche en rouge suggère le passage transitoire de l'hybridation $\text{sp}^2$ à $\text{sp}^3$ du pont méthylène. . . . .	161
IV.25	Energie d'interaction de van der Waals avec l'effet des paramètres de coupure (ligne pointillé bleue) et de lissage (ligne pointillée orange) . . . . .	163
IV.26	Matrice des paires d'interaction selon la valeur moyenne des énergies d'interaction de van der Waals, (a) <i>lanYFP</i> et (b) mNeonGreen . . . . .	164
IV.27	Matrice des paires d'interaction selon la densité de probabilité des énergie d'interaction déstabilisantes de van der Waals (a) <i>lanYFP</i> (b) mNeonGreen . . . . .	164
V.1	Schéma synoptique d'amélioration des protéines fluorescentes jaunes intégrant les méthodes discutées au cours de la thèse. . . . .	168
VI.1	Plasmide pNCS- <i>lanYFP</i> codant pour la protéine fluorescente jaune <i>lanYFP</i> . . . . .	171
VI.2	image boîte de petri sous lumière bleue . . . . .	172
VI.3	SDS-PAGE de contrôle de la digestion trypsique pour <i>lanYFP</i> (a) et mNeonGreen (b). CT correspond au puit contrôle avec la protéine sans ajout de trypsine, les ratios trypsine/protéine sont en masse et le temps de coupure est indiqué au dessus des puits. . . . .	174
VI.4	Arrangement microscopique des atomes dans un matériau amorphe (a), des polycristaux (b) et un monocristal (c). . . . .	175
VI.5	Représentation des paramètres de maille . . . . .	176
VI.6	Diagramme de phase et les conséquences empiriques de la cristallisation : précipitation (bleu), accroissement cristallin par diffusion en phase vapeur (rouge), dialyse (violet), cristallisation après précipitation partielle (orange), ensemencement (jaune), cristallisation en batch (vert) et pas de cristallisation (gris) [Haas and Drenth, 1999, Bergfors, 2003]. . . . .	177
VI.7	Schéma expérimentaux de diffusion en phase vapeur par la technique de goutte suspendue (a) et goutte assise (b). La graisse silicone scellant les dispositifs apparaît en orange. . . . .	178
VI.8	Illustration de la loi de Bragg avec l'interaction du faisceau de rayons X et les plans réticulaires d'un cristal. . . . .	180

VI.9 Représentation de la construction d'Ewald. A chaque intersection entre la sphère d'Ewald de rayon $1/\lambda$ (cercle gris) et un point du réseau réciproque contenu dans la sphère de résolution de rayon $1/d$ (cercle rouge), une interférence constructive est visible par diffraction . . . . .	181
VI.10 cliché expérimental de diffraction par rayons X obtenu pour un cristal de mNeonGreen. L'ombre sur le cliché correspond à la position du capteur de faisceau. . . . .	182
VI.11 Diagramme de Perrin-Jablonski représentant les processus de désexcitation moléculaire radiatifs en non radiatifs [Perrin, 1926, Jabłoński, 1935, Valeur and Berberan-Santos, 2012] . . . . .	186
VI.12 Montage expérimental <i>cryobench</i> (ESRF-ID29S) et les configurations des objectifs utilisés pour (a) l'absorption UV-visible avec le mode transmission (géométrie $180^\circ$ ), (b) l'émission de fluorescence avec le mode réflexion (géométrie $90^\circ$ ) et (c) la spectroscopie Raman avec le mode rétro-diffusion (géométrie $0^\circ$ ) [von Stetten et al., 2015]. . . . .	190
VI.13 Montage de la sonde Raman sur le diffractomètre de la ligne de lumière ID29. (a) Position de la sonde hors du faisceau de rayons X pour collecte de données de diffraction. (b) Position de la sonde en face de l'échantillon pour collecte de données de de spectroscopie Raman. . . . .	190
VI.14 Vue bidimensionnelle d'une cellule de simulation répliquée dans les trois dimensions de l'espace. Lors de l'utilisation des conditions aux limites périodiques, si une molécule quitte la boîte centrale ses images dans les cellules fantômes voisines se déplacent de manière similaire. . . . .	193
VI.15 Structures chimiques des différents états de protonation des acides aminés acide glutamique (GLU, GLH), acide aspartique (ASP, ASH) et histidine (HID, HIE, HIP). Le nom indiqué sous chaque structure correspond au code utilisé dans la suite <i>AMBER</i> . . . . .	196
VI.16 modèle TIP3P à trois sites d'interaction pour les molécules d'eau. (a) Description géométrique, (b) paramètres physiques . . . . .	197
VII.1 pH des différents compartiments cellulaires chez un archétype de cellule de mammifère [Casey et al., 2010] . . . . .	215
VII.2 Diagramme des groupes d'espace utilisés pour décrire la maille élémentaire des protéines fluorescentes mNeonGreen et <i>lanYFP</i> , (a) $P12_11$ , (b) $P6_522$ , (c) $P2_12_12_1$ , (d) $P2_12_12_1$ . . . . .	217
VII.3 Intensité des pics de densité provenant des cartes de différence Fourier en fonction (a) du niveau de $\sigma$ et (b) de la fraction d'électrons par unité de volume ( $e^- \cdot \text{\AA}^3$ ) . . . . .	221



# Liste des tableaux

II.1	Matrice d'identité et de similarité de séquence pour des protéines fluorescentes représentatives des embranchements cnidaire, arthropode et chordé (ESPrpt 3.0 [Robert and Gouet, 2014]). . . . .	14
II.2	Propriétés photophysiques essentielles des formes A (neutre) et B (anionique) du chromophore de AvGFP . . . . .	18
II.3	Vitesse de transfert de proton à l'état excité $k_{PT}$ rapportée dans la littérature pour différents mutants de AvGFP. Le ratio $k_{PT}$ est calculé comme $k_{PT}^{AvGFP}/k_{PT}^{mut.}$ . [van Thor et al., 2006] <sup>a</sup> , [Erez et al., 2011] <sup>b</sup> , [Shu et al., 2007] <sup>c</sup> , [Wineman-Fisher et al., 2014] <sup>d</sup> , . . . . .	23
II.4	Comportement des propriétés photophysiques des YFP suivant le compartiment intra-cellulaire, ✗ : inutilisable ( $pH < pH_{1/2} + 1$ ), ✓ : invariabilité ( $pH > pH_{1/2} + 1$ ). (a) [Casey et al., 2010], (b) table II.5 . . . . .	26
II.5	Principales propriétés de AvGFP et de ses principaux dérivés émettant dans la partie jaune du spectre du visible. $\epsilon$ désigne le coefficient d'absorption molaire ( $mM^{-1}.cm^{-1}$ ), $\phi_{fl}$ le rendement quantique d'émission de fluorescence et B la brillance ( $mM^{-1}.cm^{-1}$ ), avec $B = \phi_{fl} \cdot \epsilon$ . Les mutations F64L, M153T, V163A ou S175G ont été introduites afin d'améliorer les propriétés de repliement et de maturation de ces protéines. La protéines dont l'acronyme est sEYFP est appelé super-EYFP. [van Thor et al., 1998] <sup>a</sup> , [Tsien, 1998, Wachter et al., 1998] <sup>b</sup> , [Ormö et al., 1996] <sup>c</sup> , [Nagai et al., 2002] <sup>d</sup> , [Kremers et al., 2006] <sup>e</sup> , [Griesbeck et al., 2001] <sup>f</sup> . . . . .	26
II.6	PF jaunes naturelles et leurs propriétés photophysiques, * le coefficient d'extinction molaire $\epsilon$ et la brillance B. sont exprimés en $mM^{-1}.cm^{-1}$ . [Matz et al., 1999] <sup>a</sup> , [Shaner et al., 2013] <sup>b</sup> , [Pletneva et al., 2013] <sup>c</sup> . . . . .	27
II.7	Propriétés photophysiques de <i>lanYFP</i> et de ses mutants [Shaner et al., 2013] . . . . .	29
II.8	PF dérivées d'organisme du genre <i>Branchiostoma</i> et leurs principales caractéristiques photophysiques . . . . .	32
II.9	Propriétés photophysiques des donneurs de FRET dérivés de la ECFP. $\epsilon$ (a) et $B_{rel}$ . (b) sont respectivement exprimés en $mM^{-1}.cm^{-1}$ et en pourcentage brillance relative à EGFP. . . . .	37
III.1	Mutants dérivés de AvGFP étudiés. Sont indiqués le code PDB de la structure cristallographique choisie, le pH de cristallisation (pH), la vitesse de transfert de proton à l'état excité $k_{PT}$ et la longueur du chemin de translocation de protons proposé dans la littérature. Cette longueur a été estimée à partir des structures cristallographiques en mesurant la distance entre les atomes lourds du chemin (moyenne établie à partir de tous les monomères des structures cristallographiques). . . . .	44

III.2	Définition des atomes donneur et accepteur de proton pour les chaînes latérales de résidu. La nomenclature adoptée spécifie le nom de l'atome ainsi que le nom du proton entre parenthèse. Les atomes en rouge sont dit spéciaux puisqu'ils établissent les LH intra-moléculaires, par conséquent l'atome entre parenthèses n'est pas un proton mais un atome qui mime la résonance d'un groupe chimique.	48
III.3	Principales caractéristiques des structures cristallographiques étudiées : facteurs de température moyens par chaîne (B), nombre de résidus et d'atomes et nombre de molécules d'eau pour chaque chaîne.	60
III.4	Séquence consensus de AvGFP et ses mutants mettant en évidence les états de protonation des différentes histidines qui sont surlignées de jaune. Les sites dont les états de protonation sont différents dans la série de mutants de AvGFP étudiés sont désignés par le type HIX. Le chromophore et Glu222 et les positions mutées dans la série de mutants de AvGFP sont respectivement surlignés en vert, bleu et rouge.	65
III.5	RMSd des atomes du squelette (N, C $\alpha$ , C, O) des résidus impliqués dans les brins $\beta$ ou les hélices $\alpha$ pour les différentes protéines étudiées	67
III.6	Principales caractéristiques issues du pavage de la boîte de simulation pour les 5 protéines étudiées. (*) Le volume de la protéine est défini ici comme le volume des sous-boîtes constituant les sphères interne et de bord de la protéine.	80
III.7	Répartition des molécules d'eau entre les trois sphères considérées pour la recherche des liaisons hydrogène pour les différentes protéines étudiées.	81
III.8	Nombre moyen de réseaux de liaisons hydrogène par instantané pour AvGFP et quatre de ses mutants étudiés.	84
III.9	Caractéristiques (densité de probabilité et longueur $\langle l_{chemin} \rangle$ ) des 18 chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx dans AvGFP-T203V-S205A (4OGS).	105
III.10	Caractéristiques des 3 classes de chemins de translocation observés entre Chrom.66/OH et Glu222/OEx dans avGFP-T203V-S205A. Ces chemins sont classés selon le nombre de LH qui les constitue.	107
III.11	Caractéristiques (densité de probabilité et longueur $\langle l_{chemin} \rangle$ ) des huit chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx et de leur densité de probabilité dans AvGFP-S65T (1C4F). Les chemins 1 bis et 4 bis peuvent être considérés comme inexistantes.	118
IV.1	Listes des structures de PF du phylum <i>Arthropoda</i> , (a) [Wilmann et al., 2006], (b) [Evdokimov et al., 2006], (c) [Suto et al., 2009]	129
IV.2	Listes des structures de PF du phylum <i>Chordata</i> , (a) [Pletnev et al., 2013], (b) [Bomati et al., 2014]	130
IV.3	Table des données cristallographiques. les valeurs entre parenthèse correspondent aux coquilles de plus haute résolution), (#) la coupure en résolution est basée sur le $CC_{1/2}$ , (†) $R_{meas.} = R_{merge} \times [N/(N - 1)]^{1/2}$ , où N est la multiplicité. (*) u.a. : unité asymétrique	140
IV.4	Description des principaux modes de vibration en spectroscopie Raman pour les protéines fluorescentes, [Schellenberg et al., 2001] <sup>a</sup> , [von Stetten et al., 2015] <sup>b</sup> , [Bell et al., 2000] <sup>c</sup> , [He et al., 2002] <sup>d</sup>	154

---

IV.5	Table de réduction des données cristallographiques collectées lors de l'expérience de dommages aux radiations menée sur un cristal de mNeonGreen. Les valeurs entre parenthèse sont pour les coquilles de plus haute résolution), (#) la coupure en résolution est basée sur le $CC_{1/2}$ , (†) $R_{meas.} = R_{merge} \times [N/(N - 1)]^{1/2}$ , où N est la multiplicité. . . . .	156
IV.6	Intensité des pics observés en $e^{-} \cdot \text{Å}^3$ dans les cartes de différence Fourier où n est le numéro du jeu de donnée enregistré par diffraction aux rayons X au cours de l'expérience de dommages aux radiations. . . . .	159
VI.1	Nomenclature des champs de force décrivant les différents types de chromophore utilisés. Le champ de force utilisé dépend de l'état de protonation du groupe phénol du chromophore . . . . .	197
VII.2	Interfaces dans le dimère de trimère de mNeonGreen à pH acide [ <a href="#">Krissinel and Henrick, 2007</a> ] . . . . .	219
VII.3	Interfaces de mNeonGreen avec ses symétriques à pH physiologique [ <a href="#">Krissinel and Henrick, 2007</a> ] . . . . .	219
VII.4	Interfaces dans le tetramère de tetramère de <i>lanYFP2</i> à pH physiologique [ <a href="#">Krissinel and Henrick, 2007</a> ] . . . . .	220



# I CHAPITRE

## Avant-propos

Le sujet de la thèse est l'étude de protéines fluorescentes vertes (*Green Fluorescent Protein* de la méduse *Aequorea victoria*, ou GFP) et jaunes (*Yellow Fluorescent Proteins*, ou YFP) présentant un intérêt en imagerie de fluorescence. Ce travail s'inscrit dans le contexte des travaux de deux équipes, celle de photobiologie au Laboratoire de Chimie Physique d'Orsay et l'équipe *icOS* à l'Institut de Biologie Structurale de Grenoble, qui visent à mettre au point des protéines fluorescentes (PF) de haute performance pour les techniques d'imagerie de fluorescence. Les travaux antérieurs et indépendants de ces deux équipes ont mené à l'élaboration de deux PF cyans aux propriétés photophysiques fortement améliorées : Aquamarine et mTurquoise2. Hélène Pasquier, d'Orsay, et Antoine Royant, de Grenoble, ont choisi d'unir leurs efforts pour proposer des voies d'amélioration de PF jaunes. En effet, les PF jaunes les plus connues sont des PF dérivées de la GFP présentant une forte sensibilité environnementale (sensibilité au pH, aux halogénures) et des propriétés photophysiques suboptimales (rendement quantique, extinction molaire). Ils m'ont proposé deux approches distinctes, mais complémentaires. D'un côté, j'ai effectué une étude de la dynamique de mutants de la GFP, comprenant notamment des YFP, par simulation de dynamique moléculaire. De l'autre, j'ai réalisé une étude structurale par cristallographie aux rayons X de nouvelles PF jaunes performantes provenant d'un organisme appartenant à un phylum différent de celui des cnidaires, le phylum des méduses, anémones de mer et coraux dans lequel on trouve la grande majorité des PF utilisées actuellement en imagerie, afin de rationaliser leurs déterminants de fluorescence.

La première partie de la thèse (chapitre III) vise à comprendre la dynamique du réseau de liaisons hydrogène (LH) qui maintient le chromophore de la GFP et de ses mutants. La dépendance en pH de ces protéines est liée à l'équilibre acido-basique du groupement phénol du chromophore. La réduction de la sensibilité en pH de ces protéines implique donc de stabiliser la forme phénolate du chromophore. Nous avons souhaité par une approche théorique mieux comprendre et caractériser les paramètres conduisant à une stabilisation de cette forme. Le dé-

veloppement d'un algorithme de recherche des réseaux LH au cours de dynamique moléculaire est le cœur de cette partie. Dans la forme protonée du chromophore, il existe un transfert de proton à l'état excité (ESPT) entre le groupement phénol du chromophore et le résidu glutamate 222 situé à proximité. Afin de déterminer l'accord entre notre approche théorique et les données spectroscopiques expérimentales de l'ESPT, nous avons étudié des simulations par dynamique moléculaire de la GFP et de 4 de ses mutants. Les résultats de notre approche théorique suggèrent un accord avec les vitesses de transfert de proton à l'état excité,  $k_{PT}(\text{ESPT})$ , mesurées expérimentalement. L'utilisation de cet algorithme de recherche des réseaux de LH est maintenant disponible pour l'étude des YFP ainsi que d'autres systèmes biologiques utilisant des transferts de protons tels que la bactériorhodopsine, une pompe à proton transmembranaire d'halobactérie, le complexe IV de la chaîne respiratoire mitochondriale ou le centre réactionnel bactérien.

Dans la deuxième partie (chapitre IV), j'ai effectué une étude structurale de deux PF jaune/jaune-vert issues du ver marin *Branchiostoma lanceolatum*, aux propriétés photophysiques exceptionnelles. La première, *lanYFP*, est naturelle, et présente une structure quaternaire tétramérique ; la seconde, mNeonGreen, a été obtenue par évolution dirigée à partir de la première, et est monomérique. mNeonGreen présente toutes les qualités requises pour être la protéine fluorescente jaune de référence en microscopie de fluorescence. Grâce à notre analyse structurale, nous avons rationalisé le processus de monomérisation de la protéine fluorescente naturelle *lanYFP* en mNeonGreen. Les structures de *lanYFP* et mNeonGreen ont été réalisées à pH physiologique et à pH acide pour mNeonGreen. Nous avons également cherché à déterminer les raisons du changement de couleur remarquable subi par les cristaux de mNeonGreen sous l'effet des rayons X. L'étude de ce phénomène par spectroscopies d'absorption de lumière UV-Visible et Raman *in crystallo* en corrélation avec la cristallographie aux rayons X a permis de proposer un mécanisme d'endommagement spécifique du chromophore et de son environnement. Ces travaux ont été rassemblés dans un manuscrit qui a été soumis au journal *Acta Crystallographica section D: Structural Biology*.

L'ensemble des résultats obtenus au cours de la thèse va permettre de proposer un cadre à la fois théorique et expérimental permettant de comprendre des déterminants structuraux cruciaux contrôlant les propriétés de fluorescence des PF jaunes. Le manuscrit est divisé en 4 parties. Dans la première sont présentées les PF sous un angle bibliographique en mettant l'accent sur les propriétés photophysiques de AvGFP et des PF jaunes de *Branchiostoma lanceolatum*, ainsi

que leur application en imagerie de fluorescence. La deuxième partie expose les résultats de l'étude des réseaux de LH de AvGFP et de ses mutants par dynamique moléculaire. La troisième partie traite de l'étude structurale des PF jaunes de l'espèce *Branchiostoma lanceolatum*, en mettant l'accent sur la compréhension du processus d'évolution dirigée. Enfin, la quatrième partie décrit les méthodologies employées.

# II CHAPITRE

## Les protéines fluorescentes

Cette partie retrace l’historique de la découverte de la GFP (*Green Fluorescent Protein*) d’*Aequorea victoria* ainsi que ses principales caractéristiques structurales et photophysiques.

### A Historique

#### A.1 Découverte de AvGFP

Les travaux de Osamu Shimomura concernant la compréhension de l’émission de bioluminescence verte par la méduse *A. victoria* ont débuté au début des années 60 [Zimmer, 2015] avec l’identification, dans un premier temps, de l’aequorine, une protéine émettant de la lumière bleue en présence de calcium [Shimomura et al., 1962]. Cherchant à comprendre l’origine de l’émission verte des méduses, Shimomura a isolé une seconde protéine émettant de la lumière verte : il s’agit de la GFP (*Green Fluorescent Protein*, AvGFP). Une décennie de travail a été nécessaire pour comprendre notamment le mécanisme d’émission de AvGFP [Shimomura and Johnson, 1975] et pour résoudre la structure de son chromophore [Shimomura et al., 1963, Shimomura and Johnson, 1969, Shimomura and Johnson, 1970, Shimomura and Johnson, 1972, Shimomura, 1979].

Prasher réussit à cloner le gène codant pour AvGFP [Prasher et al., 1992]. En 1994, Chalfie récupérant la séquence codante pour AvGFP réussit à montrer qu’il était possible d’exprimer AvGFP dans une cellule (*Escherichia coli*) sans nécessité de co-facteur et de la fusionner à une protéine d’intérêt. Ces développements ont permis les premières applications en biotechnologie de l’AvGFP comme marqueur non invasif de protéines de la mouche du vinaigre (*Drosophila melanogaster*) et l’étude du système nerveux du ver transparent *Caenorhabditis elegans* [Chalfie et al., 1994, Wang and Hazelrigg, 1994].

D’un autre côté, Roger Tsien s’est intéressé à l’AvGFP pour développer les techniques



d'imagerie. L'équipe de Tsien a montré qu'il était possible de générer des PF avec des propriétés spectrales différentes grâce à l'introduction de mutations spécifiques dans le gène codant de AvGFP. Cela a conduit à l'élaboration des variants spectraux de l'AvGFP émettant du bleu au jaune [Heim et al., 1994, Heim and Tsien, 1996] (figure II.1). En parallèle de cette extension spectrale des PF, leurs travaux ont permis d'élucider le mécanisme de formation du chromophore [Zacharias and Tsien, 2006], d'améliorer le repliement tertiaire des PF [Tsien, 1998], de rendre monomérique ces protéines [Zacharias et al., 2002] et d'améliorer leurs propriétés de brillance [Heim and Tsien, 1996] (voir paragraphe VI.C.4.c).

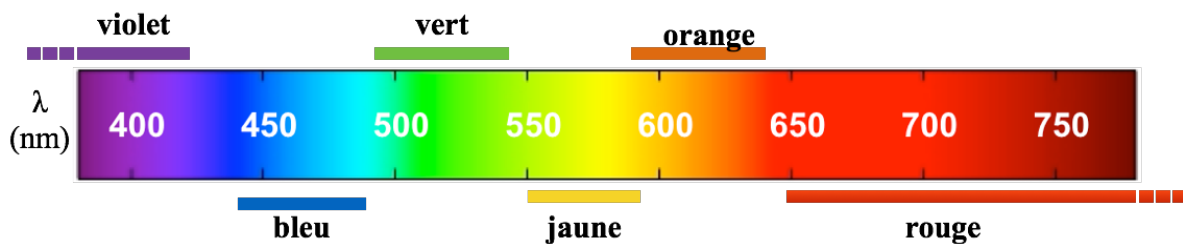


FIGURE II.1 : Couleur et longueur d'onde (voir tableau VII.A)

## A.2 Découverte d'homologues de AvGFP

Les techniques d'imagerie cellulaires *in vivo* requièrent l'usage de protéines fluorescentes émettant dans la fenêtre optique (600 à 1200 nm), où l'absorption des tissus biologiques est minimale [Anderson and Parrish, 1981]. Cette considération technique a ouvert "une course vers le rouge" dans l'ingénierie des protéines fluorescentes [Tsien, 1999]. Jusqu'en 1999, la YFP dérivée de la AvGFP est la PF avec les propriétés d'émission les plus déplacées vers le rouge. Cependant, les différentes tentatives pour étendre la palette spectrale des dérivés de AvGFP vers le rouge ont été peu récompensées [Mishin et al., 2008].

En 1999, l'équipe de Sergey Lukyanov a découvert une protéine fluorescente rouge (*DsRed*) chez un corail non bioluminescent *Discosoma sp.*, un organisme de la classe des anthozoaires [Matz et al., 1999]. Cette découverte a été cruciale puisqu'elle a permis de montrer que des organismes non bioluminescents renferment des PF. Le séquençage d'ADN haut débit a permis la découverte de nouveaux organismes principalement non bioluminescents possédant des gènes endogènes codant pour des homologues de AvGFP comme des protéines fluorescentes naturelles bleues, vertes, jaunes et rouges dans les embranchements anthropode, chordé et cnidaire [Deheyn and Latz, 2007, Baumann et al., 2008, Hunt et al., 2010] tel que *Zoanthus sp.* et *Branchiostoma lanceolatum* [Zagranichny et al., 2004, Deheyn et al., 2007].

La recherche de nouveaux gènes codants pour les homologues de AvGFP est un domaine en plein essor avec l'avènement des techniques de séquençage haut débit (expédition scientifique TARA [Bork et al., 2015]).

### A.3 Ingénierie des protéines fluorescentes

Suite à la découverte de Lukianov, Tsien et son équipe se sont investis dans le développement de variants spectraux des homologues de AvGFP afin d'élargir la palette spectrale grâce de l'utilisation de méthodes de mutagenèse dirigée et ou aléatoire. Ces méthodes ont permis de créer la série mFruits [Shaner et al., 2004] avec des mutatsnts monomériques dérivés de DsRed. Les travaux pionniers de Shimomura, Tsien et Chalfie ont été récompensés par un prix Nobel de chimie en 2008 pour "la découverte et le développement de la protéine fluorescente verte" [Tsien, 2009a, Shimomura, 2009, Chalfie, 2009].

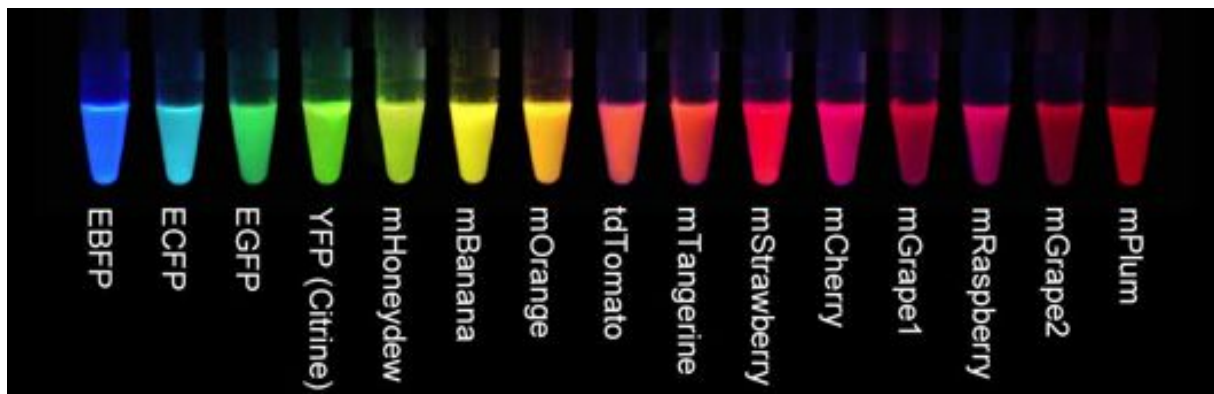


FIGURE II.2 : Diversité spectrale des PF obtenue dans le laboratoire de Tsien [Wang et al., 2004, Shaner et al., 2004, Tsien, 2009b]

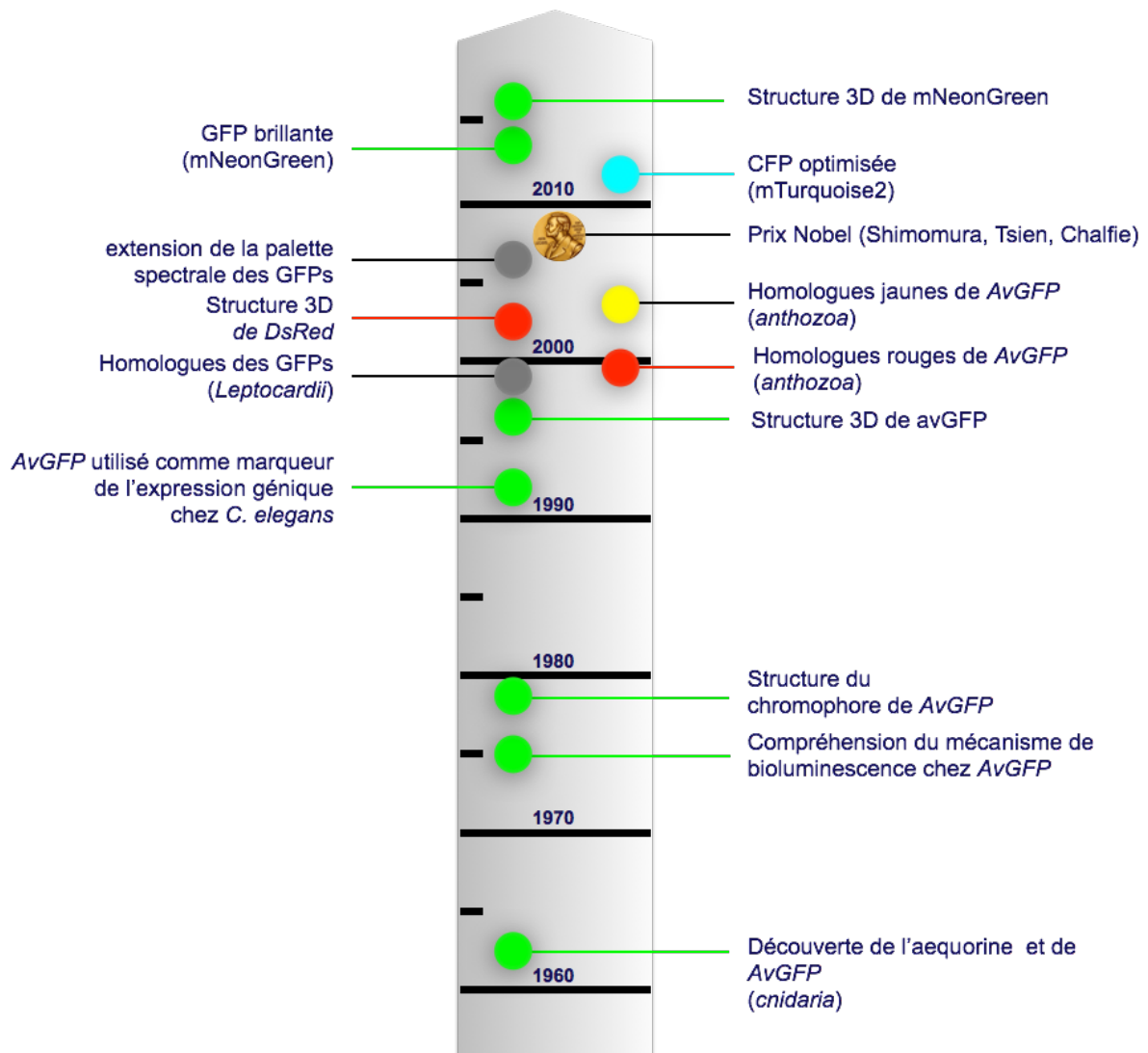


FIGURE II.3 : Chronologie des événements pour le champ de recherche des protéines fluorescentes

## B Phylogénèse des protéines fluorescentes

### B.1 Le gène GFP

Les protéines fluorescentes découvertes jusqu'à ce jour proviennent uniquement d'animaux marins. A l'état naturel, des gènes codant pour des protéines colorées de type PF et/ou chromoprotéines ont été détectées environ 300 fois dans des organismes des embranchements arthropode, chordé et cnidaire (NCBI databank, [Geer et al., 2009]). Sur la base des données du NCBI, un arbre élaboré par cladistique a été élaboré afin de visualiser la répartition des protéines naturelles colorées de type GFP au sein des organismes vivants.

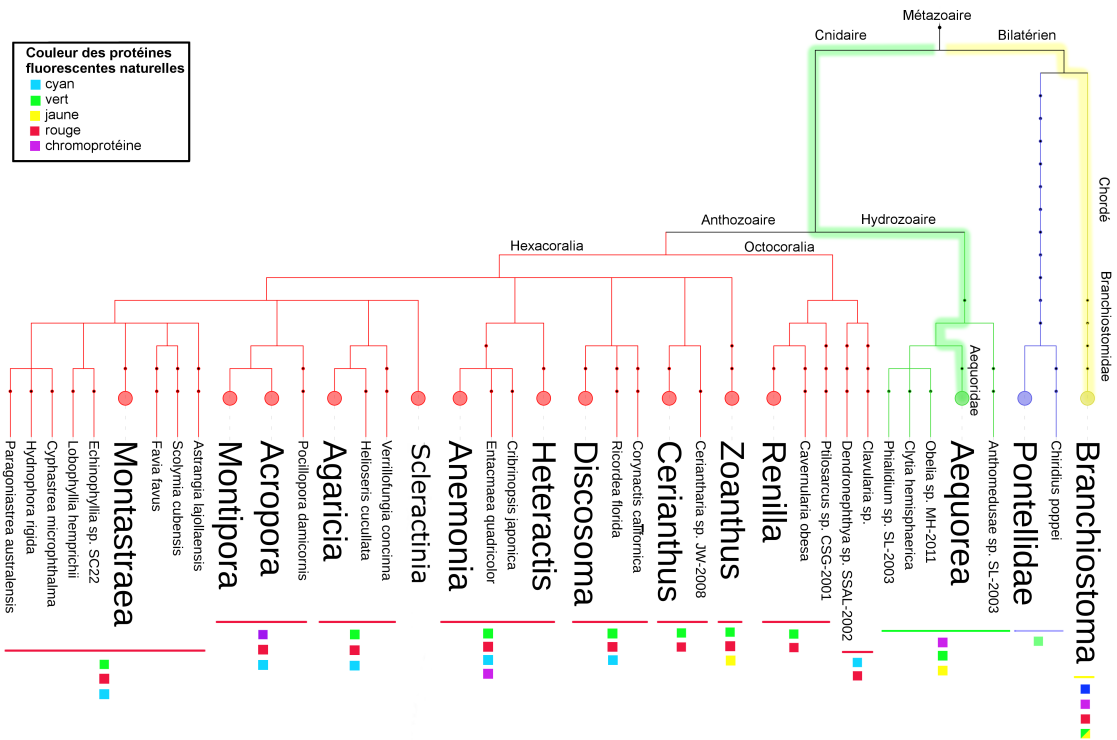


FIGURE II.4 : Arbre phylogénétique des organismes vivants qui possèdent un gène codant pour un domaine GFP [Alieva et al., 2008, Geer et al., 2009, Shagin et al., 2004]. Les clades regroupés sont signalés par un gros point et les genres étudiés au cours de cette thèse sont surlignés en jaune et vert.

Outre des PF de couleur rouge, il est intéressant de noter une diversité spectrale avec la découverte de PF naturelles vertes, cyans, jaunes et des chromoprotéines. La classe des hydrozoaires est la première à avoir été étudiée dans le cadre de travaux sur les protéines fluorescentes [Shimomura et al., 1962] dont la méduse *A. victoria* qui contient le gène codant

pour l'AvGFP [Shimomura, 2009]. A l'état naturel, la classe des hydrozoaires ne possède que des protéines de couleur verte et jaune. La classe des anthozoaires est largement représentée dans l'arbre phylogénétique de la figure II.4 avec 42 embranchements renfermant des protéines fluorescentes, dont celui de *Discosoma sp.* premier corail découvert avec le gène codant pour la protéine fluorescente rouge, DsRed [Shagin et al., 2004].

## B.2 Provenance des protéines fluorescentes étudiées en cours de thèse

Ma thèse de doctorat s'est tout particulièrement intéressée à trois organismes marins des embranchements cnidaire et chordé qui sont *Aequorea victoria*, *Branchiostoma lanceolatum* et *Branchiostoma floridae*. L'embranchement des cnidaires regroupe des organismes marins qui présentent des cellules urticantes. Cet embranchement est présent dans la nature sous deux formes, une forme fixée avec des coraux et des anémones de mer (anthozoaires) et une forme libre et mobile avec des méduses (hydrozoaires).

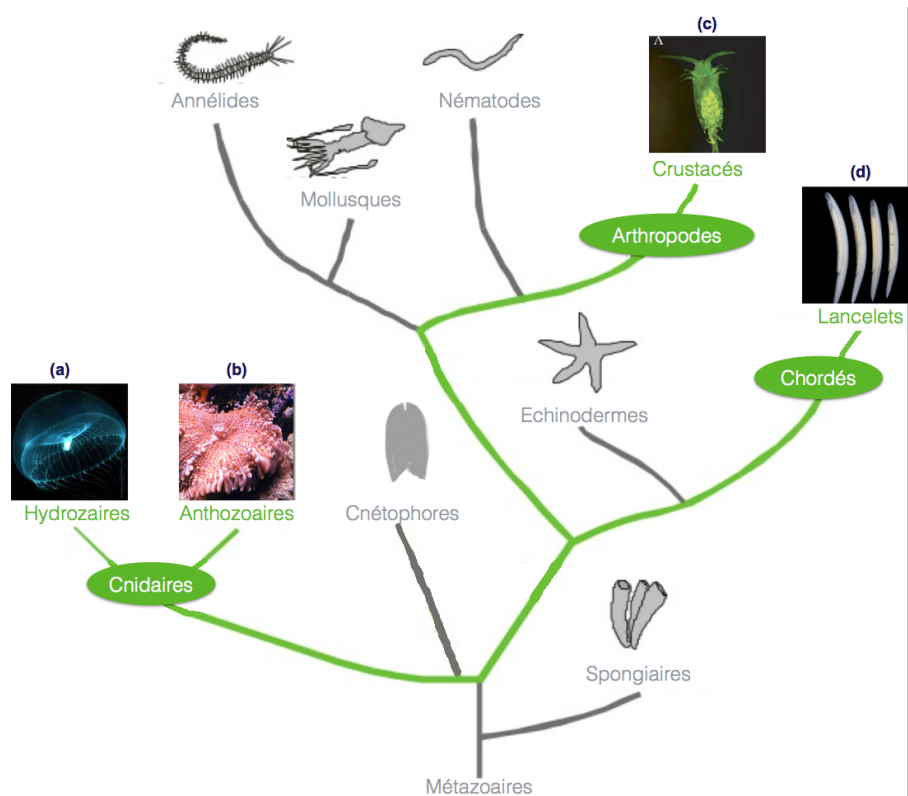


FIGURE II.5 : Arbre phylogénétique schématisé indiquant la provenance de protéines fluorescentes, (a) *Aequorea victoria* [Haddock, 2016], (b) *Discosoma striata* [Matz et al., 1999], (c) *Chiriduis poppei* [Masuda et al., 2006] et (d) *Branchiostoma lanceolatum* et *Branchiostoma floridae* [Shaner et al., 2013, Bomati et al., 2014] (adapté de [Lukyanov, 2013])

### B.2.a La méduse *Aequorea victoria*

L'espèce *A. victoria* est référencée sous la lignée complète, Eucaryote / Métazoaire / Cnidaire / Hydrozoaire / Aequoreidae / *Aequorea victoria* décrite pour la première fois par Murbach et Shearer en 1902 [Murbach and Shearer, 1902]. Le diamètre d'*A. victoria* oscille entre 8 et 20 cm et son aspect est gélatineux et transparent. Son habitat pélagique se situe sur la côte ouest des Etats-Unis d'Amérique entre le détroit de Bering et la pointe de la Californie. Son identification est facilitée par la présence, sur le pourtour de son ombrelle, de photo-organes fluorescents sous lumière bleue.



FIGURE II.6 : Méduse *A. victoria* [Haddock, 2016]

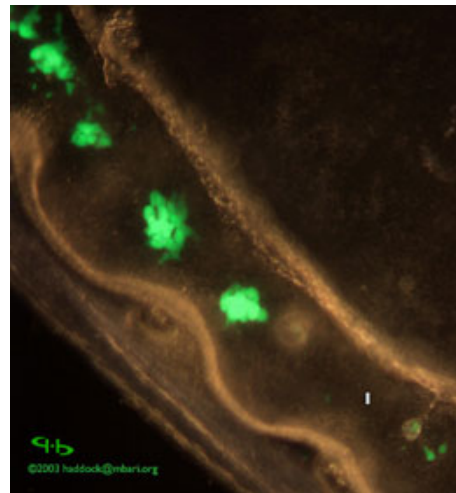


FIGURE II.7 : Les photo-organes d'*A. victoria* situés sur le pourtour de son ombrelle sont stimulés par de la lumière bleue [Zimmer, 2015]

### B.2.b Les vers de mer du genre *Branchiostoma*

Les travaux de Deheyn et coll. et Bomati et coll. ont montrés que des PF sont présentes dans des organismes du genre *Branchiostoma*, des vers marins, avec une large diversité spectrale avec les couleurs rouge, jaune, bleue et des chromoprotéines [Deheyn et al., 2007, Bomati et al., 2014]. sur les organismes de genre *Branchiostoma* [Deheyn et al., 2007, Bomati et al., 2014]. Les espèces *Branchiostoma lanceolatum* et *Branchiostoma floridae* ont été respectivement décrits par Pallas et Hubbs [Pallas, 1774, Hubbs, 1922]. Ces vers de mer ont une taille d'environ 4 à 5 cm. Ce sont des rostrés fouisseurs vivant dans les fonds sablonneux du littoral. Leur distribution sur le globe est dépendant de leur espèce puisque *B. lanceolatum* colonise la Méditerranée, la mer Noire, la Manche, la mer du Nord et la côte est des Etats-Unis d'Amérique quand *B. floridae* se situe uniquement sur dans le golfe du Mexique. La lignée complète du genre *Branchiostoma* est la suivante, Eucaryote / Métazoaire / Chordé / Branchiostomidae / *Branchiostoma*.

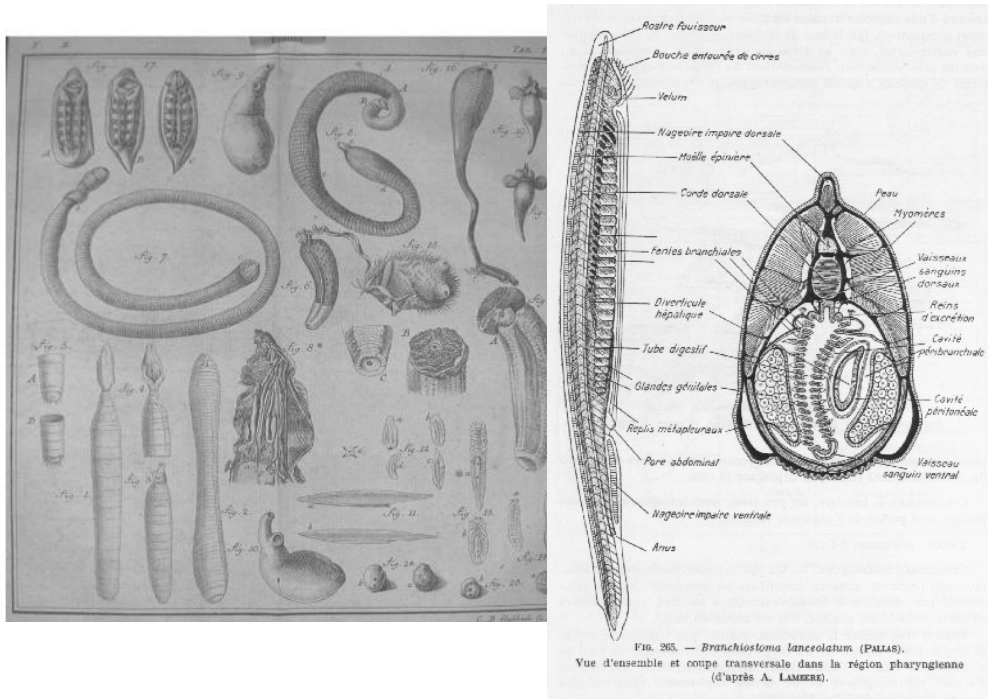


FIGURE II.8 : planche d'anatomie de Pallas classant *B. lanceolatum* dans l'embranchement des mollusques et de Poll [Pallas, 1774, Poll, 1947]

## C Structure tridimensionnelle des protéines fluorescentes

### C.1 Structure tertiaire

La structure tridimensionnelle de l'AvGFP forme un tonneau de  $\sim 42 \text{ \AA}$  de hauteur pour  $\sim 25 \text{ \AA}$  de largeur et qui est constituée de 220 à 240 résidus ( $\sim 27 \text{ kDa}$ ). Cette structure tertiaire est composée de onze brins  $\beta$  et d'une hélice  $\alpha$  coaxiale supportant un chromophore. L'arrangement des brins  $\beta$  dans le tonneau est de type feuillet antiparallèle avec une fermeture sur lui-même constitué de deux brins parallèles.

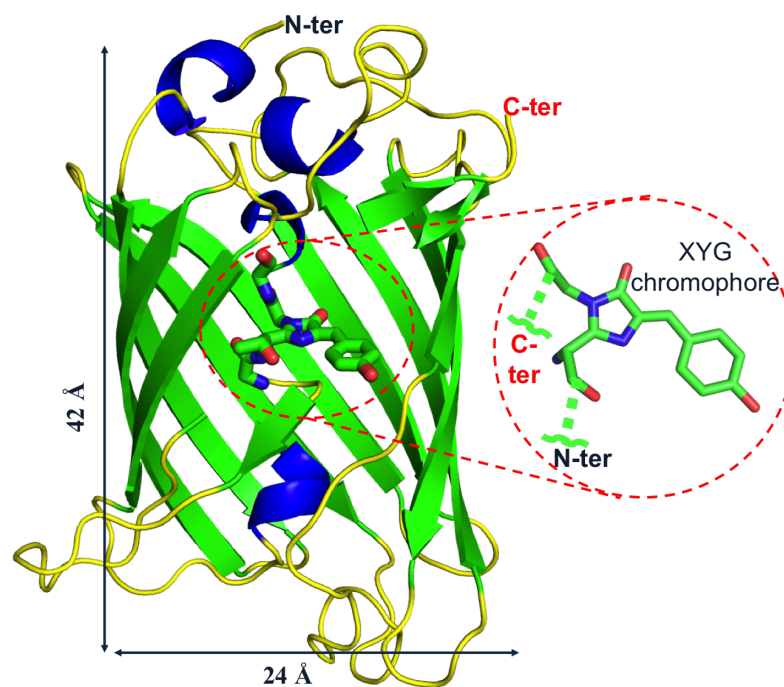


FIGURE II.9 : Structure tridimensionnelle de AvGFP en forme de tonneau de brins  $\beta$  (vert), avec son hélice  $\alpha$  coaxiale (bleu) et ses boucles (jaune) [Ormö et al., 1996, Yang et al., 1996]

Le chromophore fait partie intégrante de la chaîne polypeptidique chez tous les homologues de AvGFP connus. Malgré une disparité de séquence au sein de ces protéines homologues (figure II.10 et tableau II.1), il est intéressant de relever une conservation remarquable de deux acides aminés, notés Arg96 et Glu222 chez AvGFP [Kremers et al., 2011]. Ces acides aminés sont essentiels pour la catalyse de la formation du chromophore [Sniegowski et al., 2005]. Le tonneau de brins  $\beta$  n'est que partiellement fermé, une ouverture est présente due à la perturbation du réseau de liaisons hydrogène entre les brins 7 et 8 [Pitman et al., 2015].



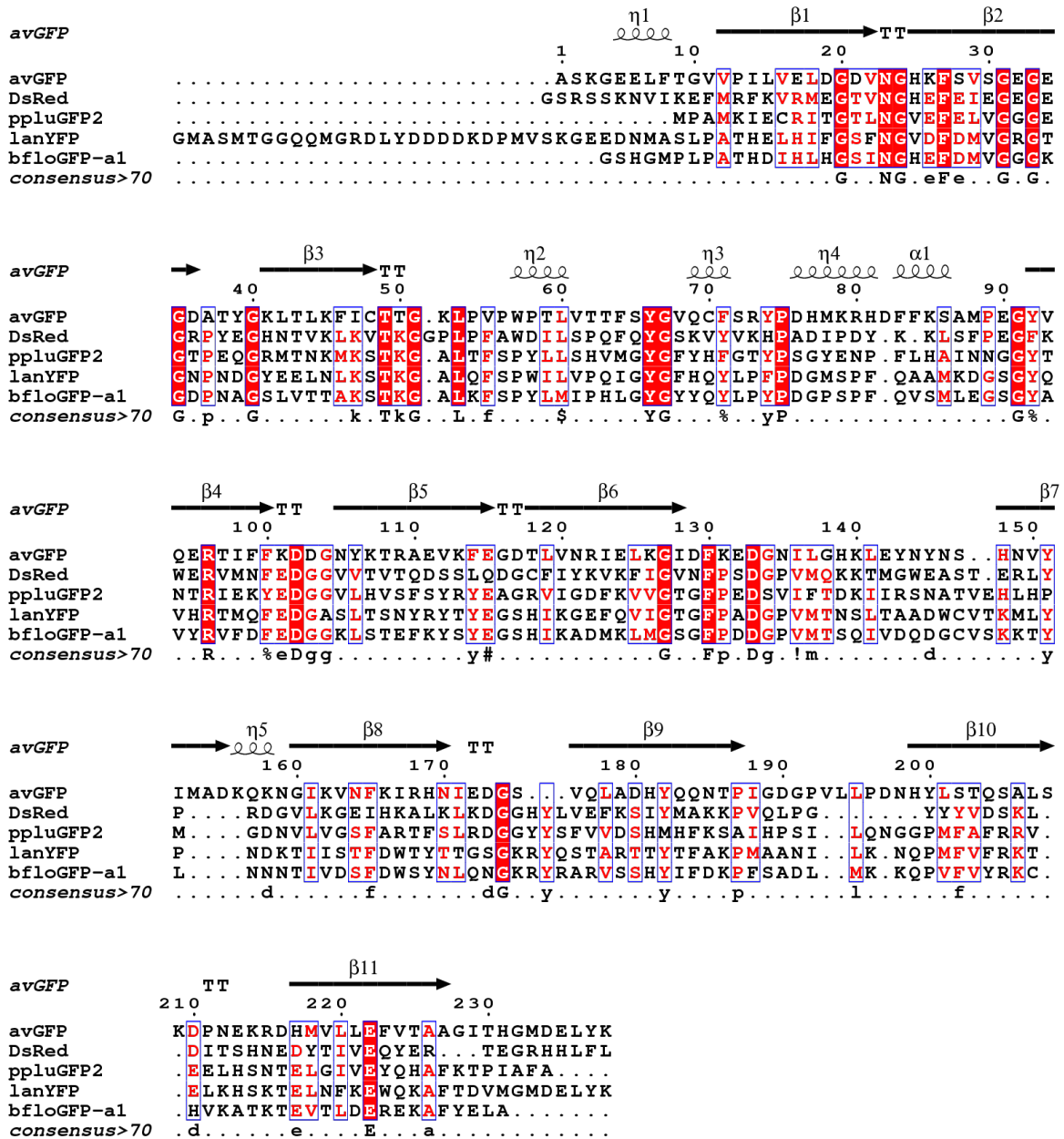


FIGURE II.10 : Alignement de séquence de protéines fluorescentes représentatives des embranchements cnidaire, arthropode et chordé réalisé avec le serveur Clustal Omega [Sievers and Higgins, 2014] et sa représentation avec le serveur ESPript 3.0 [Robert and Gouet, 2014].

	AvGFP	DsRed	ppluGFP2	lan YFP	bfloGFP-a1	
AvGFP	–	40	36	32	34	% similarité
DsRed	25	–	45	42	42	
ppluGFP2	21	30	–	52	55	
lan YFP	19	26	35	–	72	
bfloGFP-a1	21	25	34	52	–	

% identité

Tableau II.1 : Matrice d'identité et de similarité de séquence pour des protéines fluorescentes représentatives des embranchements cnidaire, arthropode et chordé (ESPrpt 3.0 [Robert and Gouet, 2014]).

## C.2 Le chromophore

La formation du chromophore est une réaction auto-catalytique conduisant à la formation d'un composé hétérocyclique nitré, 4-(p-hydroxybenzylidène)imidazolidin-5-one en trois étapes (HBDI). Elle est enclenchée par l'attaque nucléophile de l'azote du Gly67 sur le carbonyle de la chaîne carbonée de Ser65. Une réaction de déshydratation suit le réarrangement du proton de l'amine secondaire de Tyr66 sur le groupe hydroxyle formé par la cyclisation de l'imidazolinone. La dernière étape de la synthèse du chromophore de AvGFP consiste en l'oxydation de la liaison simple  $C\alpha-C\beta$  de Tyr66 qui conduit à la création d'un système électronique  $\pi$  conjugué sur l'ensemble du chromophore.

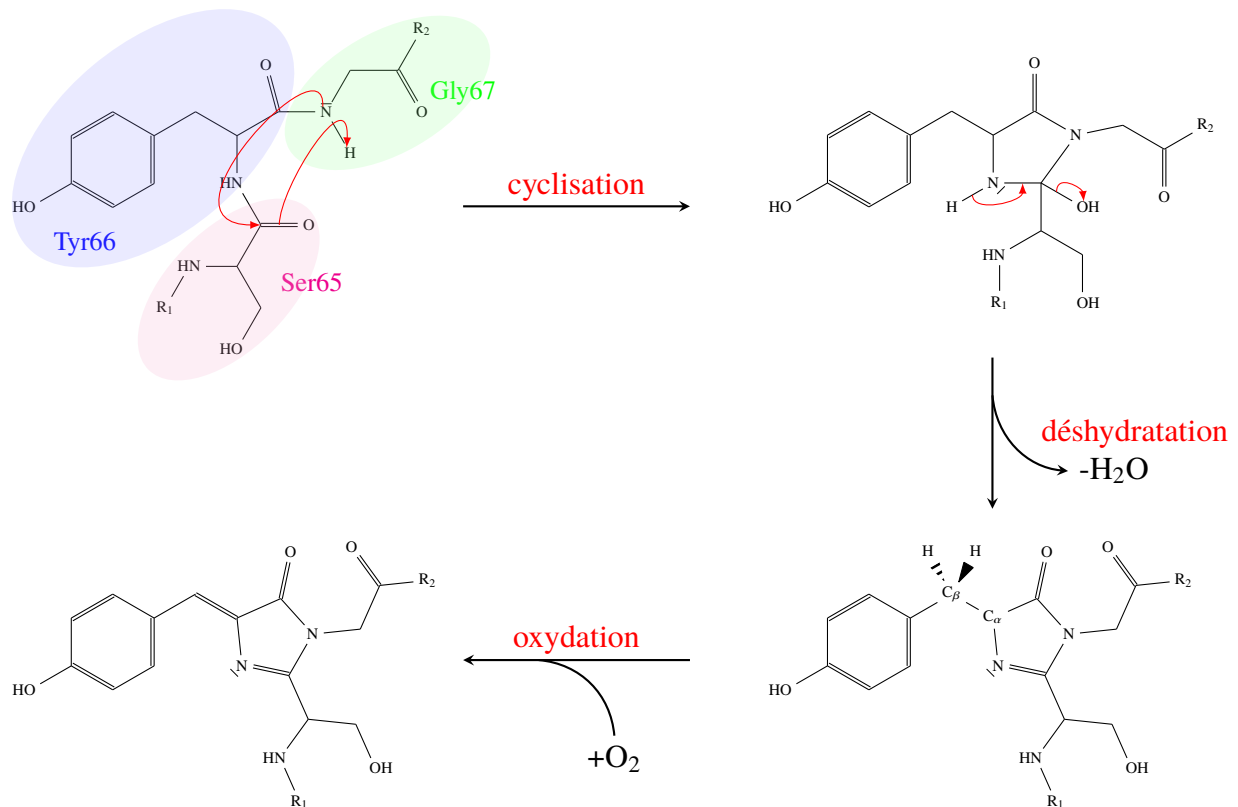


FIGURE II.11 : Mécanisme de formation du chromophore de AvGFP [Cubitt et al., 1995, Miyawaki et al., 2003]

La réaction chimique décrite dans la figure II.12 forme de manière prédominante un stéréoisomère *cis*-plan de chromophore, c'est à dire que les groupes phénol et imidazolinone sont du même côté que la double liaison  $C\alpha-C\beta$  et que leurs orbitales atomiques  $p_z$  sont alignées selon un axe perpendiculaire au plan du phénol. Les angles diédraux  $\tau$  et  $\varphi$  sont respectivement définis

par les atomes  $N_2/C_{A2}/C_{B2}/C_{G2}$  et  $C_{D1}/C_{G2}/C_{B2}/C_{A2}$  qui permettent le suivi de la planarité du chromophore au cours de simulation par dynamique moléculaire.

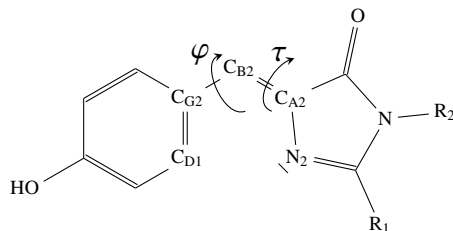


FIGURE II.12 : Représentation du chromophore protoné cis-plan et des angles diédraux  $\tau$  et  $\varphi$

La délocalisation des électrons sur les cycles phénol et imidazolinone permet de diminuer l'énergie d'excitation d'une tyrosine depuis l'UV lointain jusque dans le visible. On peut remarquer que dans le cas de systèmes  $\pi$  conjugués, la longueur d'onde de la transition d'absorption est d'autant plus grande que l'étendue du système conjugué est importante. Néanmoins, pour une même structure du chromophore, des déplacements spectraux peuvent être constatés jusqu'à 65 nm dans des protéines fluorescentes portant un chromophore de même structure. Ceci démontre l'impact important de l'environnement du chromophore induisant des variations du champ électrostatique autour de ce dernier [Drobizhev et al., 2015, Shaner et al., 2011].

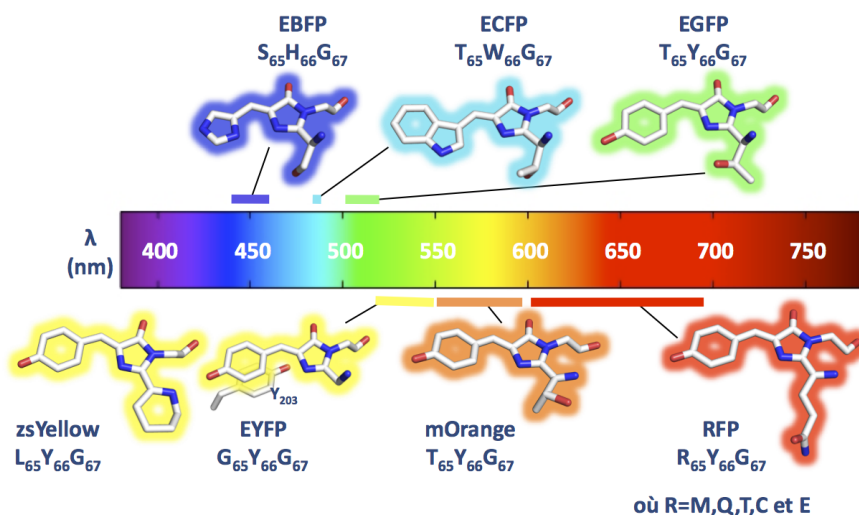


FIGURE II.13 : Typologie des chromophores chez les PF homologues de AvGFP en fonction de leur longueur d'onde d'émission.

## D La GFP sauvage (*Green Fluorescent Protein*)

Cette section traite de AvGFP découverte par Shimomura dans l'organisme *A. victoria*. Le chromophore de cette protéine est constitué par le motif TYG et sa structure a été résolue en 1996 [Ormö et al., 1996, Yang et al., 1996]. En 2010, une structure à très haute résolution (Res. = 0,90 Å) a été obtenue [Shinobu et al., 2010].

### D.1 Photophysique de AvGFP

Le spectre d'absorption de AvGFP à température ambiante possède deux bandes dont les maxima sont situés à 395 et 482 nm. La présence de deux bandes distinctes sur le spectre d'absorption de AvGFP démontre la coexistence de deux espèces chimiques. Ces bandes ont été respectivement attribuées aux formes neutre (forme A), prédominante à pH 7,4, et anionique (forme B) du chromophore SYG [Brejc et al., 1997] (fig. II.14). L'excitation de ces deux espèces chimiques génère des bandes d'émission de fluorescence distinctes mais voisines avec des maxima à 508 nm (neutre) et 510 nm (anionique). [Tsien, 1998, Merola et al., 2010, Day and Davidson, 2014]. Le fort déplacement de Stokes entre l'absorption de la forme neutre et l'émission de fluorescence a été interprété comme la conséquence d'un transfert de proton à l'état excité (ESPT).

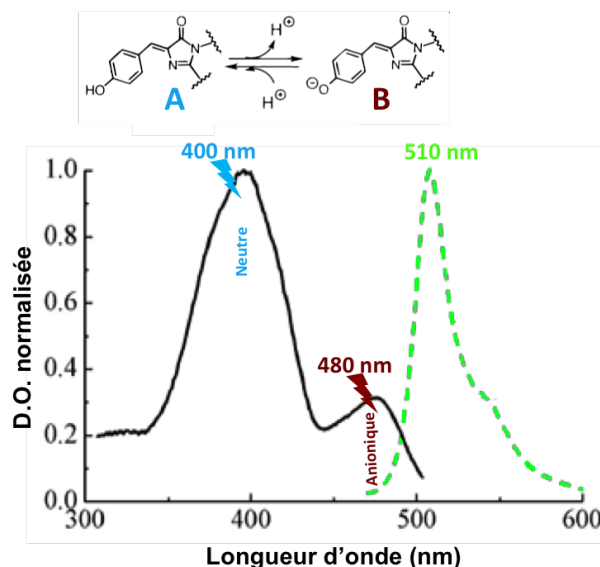


FIGURE II.14 : Spectre d'absorption et d'émission de fluorescence de AvGFP. Les deux bandes du spectre d'absorption dénotent l'équilibre acido-basique entre la forme A, neutre, et B, anionique, du chromophore [Tsien, 1998].

Le rendement quantique de l'émission de fluorescence de AvGFP a été mesuré à  $\phi_{fl} = 0,8$  et la durée de vie de fluorescence à  $\tau_{fl} = 3,3$  ns [Striker et al., 1999]. Les coefficients d'extinction molaire des espèces chimiques A et B ont été mesurés égales à  $\epsilon_A = 27,5 \pm 2,5$  mM<sup>-1</sup>cm<sup>-1</sup> et  $\epsilon_B = 11,8 \pm 2,3$  mM<sup>-1</sup>cm<sup>-1</sup>.

Forme AvGFP	Max. d'absorption (nm)	Max. d'émission (nm)	$\phi_{fl}$	$\epsilon$ (mM <sup>-1</sup> .cm <sup>-1</sup> )	Brillance (mM <sup>-1</sup> .cm <sup>-1</sup> )
A	397	510	0,8	27,5	22,0
B	495	510	0,8	11,8	9,5

Tableau II.2 : Propriétés photophysiques essentielles des formes A (neutre) et B (anionique) du chromophore de AvGFP

Le chromophore d'une protéine fluorescente dépliée par dénaturation basique à température ambiante absorbe à 440 nm avec un coefficient d'extinction molaire de 44 mM<sup>-1</sup>.cm<sup>-1</sup> [Ward et al., 1981]. Le chromophore seul, sans environnement protéique, ne fluoresce pas dans ces conditions ( $\phi_{fl} < 10^{-3}$ ). La fluorescence du chromophore modèle HBDI est restaurée par abaissement de la température à 77 K [Niwa et al., 1996]. De la même façon, une PF dénaturée ne fluoresce pas. Ces observations suggèrent que les contraintes stériques imposées par l'environnement du chromophore sont importantes pour l'émission de fluorescence de AvGFP. Elles restreignent les degrés de liberté des angles  $\tau$  et  $\varphi$  sans pour autant empêcher totalement le chromophore de se mouvoir dans sa cavité. En effet, des calculs quantiques et des simulations par dynamique moléculaire à l'état excité ont permis de montrer que ces torsions qui conduisent à une désactivation non-radiative du chromophore s'effectuent à l'échelle de la nanoseconde dans la protéine alors qu'elles se font à l'échelle de la picoseconde en solvant [Jonasson et al., 2011].

## D.2 Le transfert de proton à l'état excité chez AvGFP

Les différents processus photophysiques en jeu pour la AvGFP sont schématisés dans la figure II.15 [Brejc et al., 1997, Bell et al., 2000]. Les formes A (neutre) et B (anionique) du chromophore se distinguent par leur état de protonation (voir paragraphe II.D.1) et leur signature spectrale. La conversion dans le noir entre les formes A et B est peu fréquente et s'effectue après protonation de la forme A en une forme intermédiaire I et le réarrangement de Thr203 menant à la forme B. Cette conversion de la forme A vers B est accélérée avec une excitation lumineuse. La photo-excitation de la forme A de AvGFP induit un transfert de proton à l'état excité (ESPT) en direction du résidu Glu222. L'ESPT résulte de la déprotonation de la forme A\* en forme I\* à l'échelle de temps de la picoseconde. La relaxation radiative de l'état I\* génère un pic d'émission de fluorescence à 508 nm qui est similaire à celle de l'état B\* (510 nm). Le retour de l'état I à l'état fondamental A s'effectue par étape avec les états intermédiaires I et I<sub>2</sub> en quelque picosecondes [Chattoraj et al., 1996, Kennis et al., 2004]. De plus, un processus de décarboxylation du résidu Glu 222 est observé chez AvGFP et certains de ses mutants portant la mutation T203V-S205A [van Thor et al., 2002]. Ce processus est dû à l'excitation de la forme A\* en A\*\* avec une excitation à 800 nm [Langhojer et al., 2009].

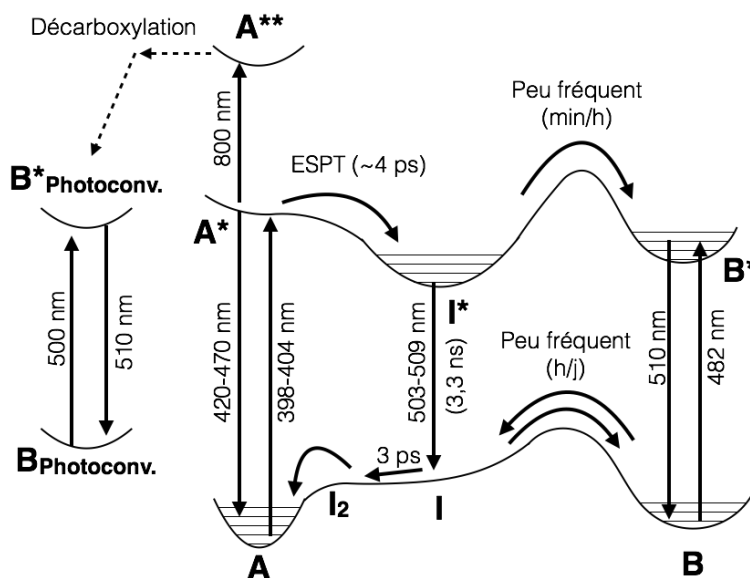


FIGURE II.15 : Processus mis en jeu lors de la photo-excitation de AvGFP (adapté de [Day and Davidson, 2014])

La présence d'un ESPT entre les formes A\* et I\* est renforcé par l'observation d'un chemin de translocation de proton sur la base de structures cristallographiques de AvGFP et de certains

de ces mutants. Ce chemin relie le phénol du chromophore au groupe carboxyle de Glu222 en passant successivement par une molécule d'eau et le groupe hydroxyle de Ser205 [Brejc et al., 1997, Palm et al., 1997]. Le déclenchement de l'ESPT par la photo-excitation est induit par le transfert de charge  $n \rightarrow \pi^*$  (non-liant à anti-liant) de l'oxygène du phénol vers les groupes phénol et imidazolinone du chromophore [Tolbert and Solntsev, 2002, Shu et al., 2007]. Par ailleurs, des données de spectroscopie vibrationnelle ultrarapide conduite sur AvGFP suggèrent que Glu222 est un accepteur final de proton [Stoner-Ma et al., 2005].

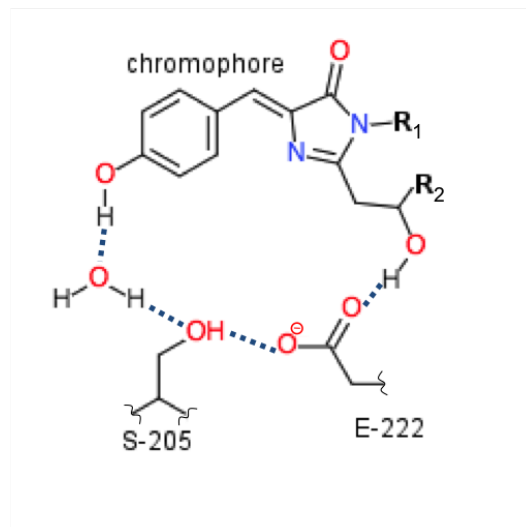


FIGURE II.16 : Chemin de translocation de proton supportant l'ESPT chez AvGFP et ses mutants [Shu et al., 2007]

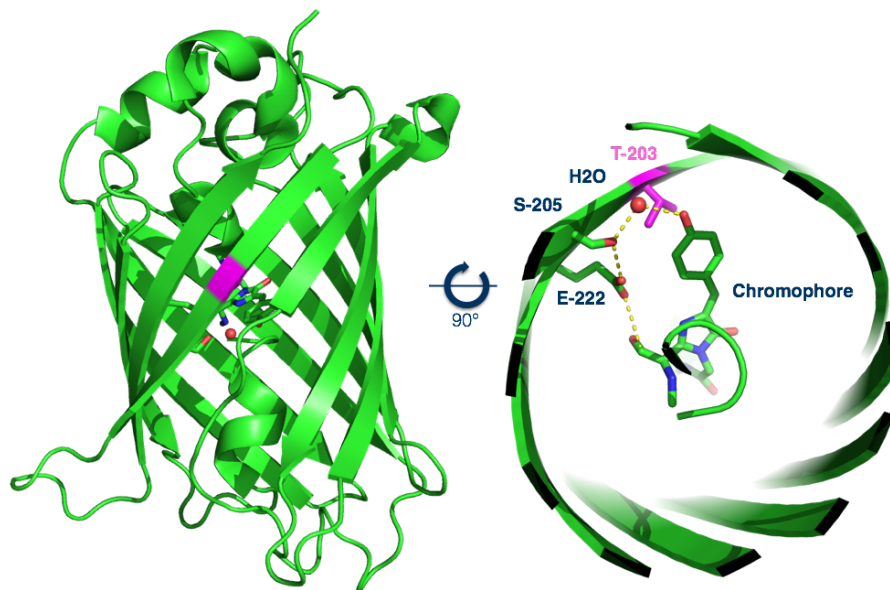


FIGURE II.17 : Représentation tridimensionnelle du chemin de translocation de proton supportant l'ESPT chez AvGFP et ses mutants



### D.3 Le glutamate 222 : accepteur final de proton

Les données de spectroscopie IR vibrationnelle ultrarapide (TIR) conduite sur AvGFP, où les processus de décarboxylation du résidu Glu222 induits par photo-excitation sont limités, atteste que Glu222 est l'accepteur final de proton [Stoner-Ma et al., 2005]. En effet, les résultats de ces expériences montrent qu'un signal à  $1565\text{ cm}^{-1}$  diminue de manière concomitante avec l'apparition d'un signal à  $1706\text{ cm}^{-1}$ . Les bandes de  $1565\text{ cm}^{-1}$  et  $1706\text{ cm}^{-1}$  ont été respectivement assignées à l'étirement antisymétrique de l'acide carboxylique et à la forme protonée du groupe carboxyle de Glu222 [Stoner-Ma et al., 2005]. Les données spectroscopiques TIR ont été corroborées par des simulations de transfert de proton entre le groupe phénol du chromophore et Glu222 [Lill and Helms, 2002].

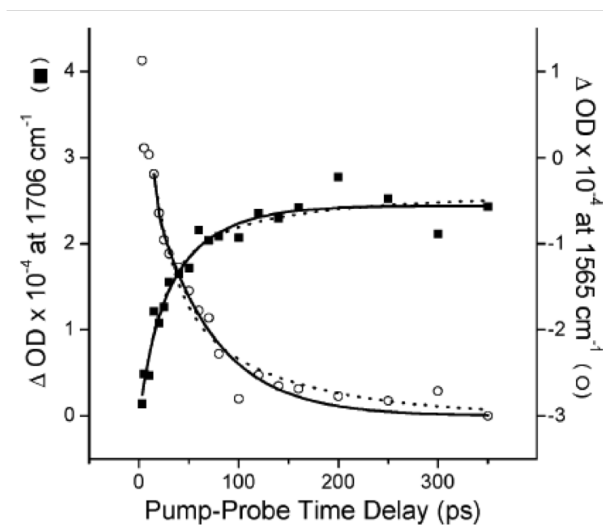


FIGURE II.18 : Absorption et blanchiment des pics IR qui sont observés par spectroscopie TIR et qui ont été associées à l'étirement antisymétrique du carboxylate ( $1565\text{ cm}^{-1}$  carrés pleins) et à la forme protonée du groupe carboxyle du résidu Glu 222 [Stoner-Ma et al., 2005]

### D.4 Modification du transfert de proton à l'état excité par mutagenèse

La mutagenèse dirigée permet de mettre en évidence différents chemins de translocation de proton mis en jeu dans l'étude de l'ESPT. Des mutations ponctuelles le long du chemin de translocation de proton exposé précédemment ou dans son environnement très proche a permis l'élaboration de mutants de AvGFP ayant des vitesses de transfert de proton,  $k_{PT}$ , modulées. Grâce à l'étude couplée des données spectroscopiques et des structures cristallographiques de ces mutants, des chemins de protons alternatifs ont été identifiés. La vitesse de transfert du proton ( $k_{PT}$ ) peut être estimée à partir du rapport des intensités des bandes d'émission de

fluorescence (I) et d'absorption (A) [Wineman-Fisher et al., 2014] :

$$k_{PT} = k_r \times \frac{I_{exROH}^{RO^-}}{I_{ROH}} \times \frac{A^{RO^-}}{A_{ROH}} \quad (\text{II.1})$$

Dans le cas où la durée de vie des espèces neutre (ROH) et anionique (RO<sup>-</sup>) du chromophore sont équivalentes [Wineman-Fisher et al., 2014] :

$$k_{PT} = \frac{I^{RO^-}}{I_{ROH}} \times \frac{1}{\tau_{fl}} \quad , \text{ où } \tau_{fl}^{O^-} \equiv \tau_{fl}^{OH} \quad (\text{II.2})$$

Un travail bibliographique a été mené pour rechercher des mutants dont la vitesse  $k_{PT}$  est modifiée et mesurée expérimentalement. Ces mutants portent des mutations au niveau du chemin de translocation de proton de la littérature (S-205) ou dans l'environnement proche du chromophore (S-65 et T-203). Ils sont listés dans la table II.3. Ces mutations induisent une modification du chemin de translocation de proton entre le phénol du chromophore et le glutamate en position 222. L'introduction d'une mutation en position 205 conduit à une réduction de la vitesse  $k_{PT}$  d'un facteur 20 à 30. Sur la base de l'analyse détaillée de la structure cristallographique de AvGFP-S205V (code PDB : 2QLE), un chemin de translocation de proton alternatif a été proposé. Il implique toujours une molécule d'eau à proximité du phénol du chromophore et est dévié via le résidu thréonine en position 203 [Shu et al., 2007, Erez et al., 2011]. Cette hypothèse a été étayée dans le cas du mutant AvGFP-S205V grâce à une étude par dynamique moléculaire permettant de suivre l'existence de liaisons hydrogène (LH) individuelles. Pour ce faire, la distribution des distances entre atomes lourds du chemin de translocation a été établie. Pour chacune de ces distances, plusieurs sous-populations ont été mises en évidence dont une peu peuplée au cours de la dynamique moléculaire compatible avec les critères de distances des LH [Simkovitch et al., 2013]. La mutation de la thréonine en position 203 précédemment relais de l'ESPT en supplément des mutations en position 205 réduit très fortement la vitesse  $k_{PT}$  (AvGFP-T203V-S205A) voire annihile l'ESPT (AvGFP-T203V-S205V) [Wineman-Fisher et al., 2014]. Afin d'expliquer la persistance de l'ESPT dans le mutant AvGFP-T203V-S205A, une étude par dynamique moléculaire met en évidence le rôle de molécules d'eau relais, permettant un ESPT du phénol du chromophore au solvant en passant par une molécule d'eau présente en permanence à proximité du chromophore. Ces molécules d'eau relais pénètrent de façon transitoire dans la structure protéique. La présence dans ces mutants d'une valine au lieu d'une alanine en position 205 permettrait d'augmenter la

compacité de la cavité du chromophore et de réduire l'accessibilité de ce dernier aux molécules d'eau du solvant. Ceci expliquerait l'absence d'un ESPT dans le mutant AvGFP-T203V-S205V.

Mutants	$k_{PT}$ (s <sup>-1</sup> )	$k_{PT}$ relatif à AvGFP	Trajet de l'ESPT
AvGFP <sup>a,b</sup>	$1,25 \times 10^{11}$	1	E-222/COO <sup>-</sup> → S-205/OH → H <sub>2</sub> O → Chrom./OH
AvGFP-S205A <sup>b</sup>	$6,60 \times 10^9$	20	E-222/COO <sup>-</sup> → T-203/OH → H <sub>2</sub> O → Chrom./OH
AvGFP-S205V <sup>b,c</sup>	$4,10 \times 10^9$	30	E-222/COO <sup>-</sup> → T-203/OH → H <sub>2</sub> O → Chrom./OH
AvGFP-T203V-S205A <sup>d</sup>	$3,60 \times 10^8$	350	E-222/COO <sup>-</sup> → H <sub>2</sub> O → solvant

Tableau II.3 : Vitesse de transfert de proton à l'état excité  $k_{PT}$  rapportée dans la littérature pour différents mutants de AvGFP. Le ratio  $k_{PT}$  est calculé comme  $k_{PT}^{AvGFP}/k_{PT}^{mut.}$ . [van Thor et al., 2006]<sup>a</sup>, [Erez et al., 2011]<sup>b</sup>, [Shu et al., 2007]<sup>c</sup>, [Wineman-Fisher et al., 2014]<sup>d</sup>,

## E Les protéines fluorescentes jaunes

### E.1 Protéines fluorescentes jaunes dérivées de AvGFP

Les YFP (*Yellow Fluorescent Protein*) sont des protéines fluorescentes jaunes dérivées de la AvGFP comportant une mutation en position 203 conduisant à la formation d'une interaction de type  $\pi$ - $\pi$  entre le cycle phénolate du chromophore et un cycle aromatique au niveau du résidu 203. Suite à la mutation T203Y portée par les YFP dérivées de AvGFP, cette interaction est responsable du déplacement bathochrome d'environ 20 nm de leur propriétés spectrales par rapport à AvGFP [Ormö et al., 1996, Tsien, 1998]. Le chromophore de ces protéines est de type GYG. La présence d'une tyrosine en position 66 induit des propriétés spectrales avec des spectres d'absorption constitués de deux bandes. Le maximum des ces bandes est situé au alentour de 390 nm et de 510 nm. Ces deux bandes ont été respectivement attribuées aux formes neutre et anionique du chromophore [Tsien, 1998]. Le spectre d'émission est quand à lui constitué d'une seule bande avec un maximum vers 530 nm (table II.5). Le  $\text{pH}_{1/2}$  des protéines fluorescentes qui représente le pH pour lequel l'absorption ou l'émission de fluorescence décroît de 50% varie entre 5,6 et 7,0.

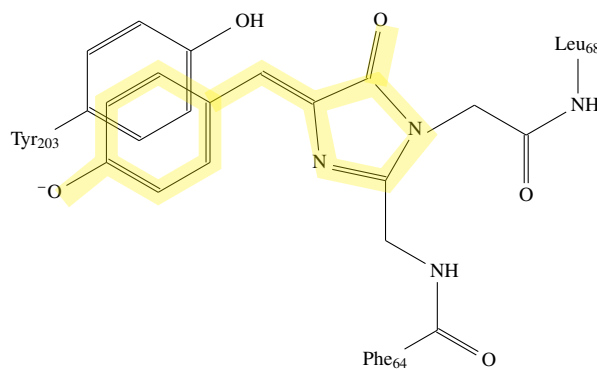


FIGURE II.19 : Chromophore de type GYG présent chez les protéines fluorescentes jaunes dérivées de AvGFP montrant l'interaction de type  $\pi$ - $\pi$  avec la tyrosine 203

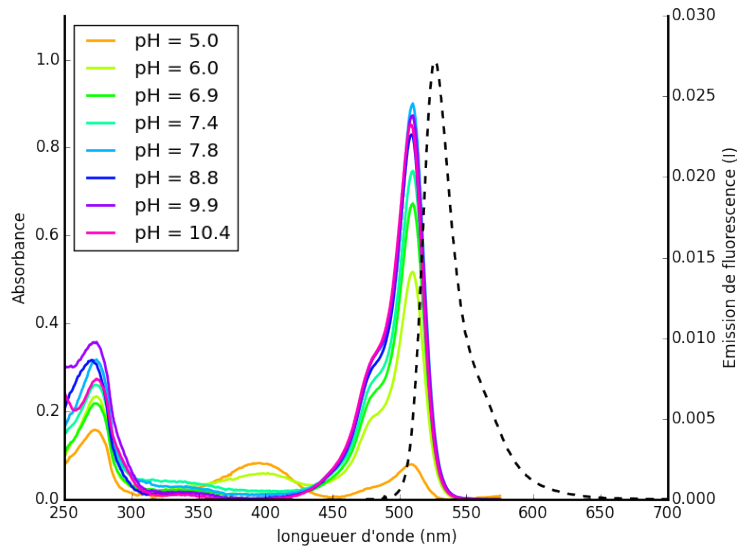


FIGURE II.20 : Spectre d'absorption à différents pH (trait pleins colorés) et d'émission de fluorescence à pH 7,4 (pointillés noirs, excitation à 514 nm) pour la protéine fluorescente jaune YFP .

Ces protéines présentent une sensibilité environnementale importante qui se définit comme étant la susceptibilité d'un spectre d'être influencé par des paramètres physico-chimiques (température, pH, pression) et la présence d'espèce chimiques (halogénures, ROS, métaux). La proportion de formes neutre et anionique du chromophore dépend du pH, mais également de la concentration en ions chlorures dans l'échantillon. Ces propriétés peuvent être un avantage lorsque ces protéines sont utilisées comme biosenseur de pH ou d'ions chlorures. En revanche, la faible stabilité de ces protéines face aux changements de pH ou de concentration en ions chlorures peut être un inconvénient majeur lors de l'analyse quantitative des signaux d'imagerie de fluorescence. En effet, le pH intra-cellulaire varie de 4,7 à 8,0 suivant le compartiment intra-cellulaire considéré [Casey et al., 2010]. Par exemple, l'incorporation de ces protéines dans des compartiments intra-cellulaires de pH plus faible que leur  $pH_{1/2}$  va favoriser la forme neutre non fluorescente de leur chromophore, les rendant inutilisables pour l'imagerie de fluorescence (table II.4). Quant aux ions chlorures, ils sont omniprésents dans les systèmes biologiques, en particulier dans les cellules de mammifères avec des concentrations inférieures entre 3 et 80 mM [Bregestovski et al., 2009]. Des efforts ont été entrepris afin de réduire la sensibilité environnementale de ces protéines. Cet objectif a été partiellement atteint avec la génération des mutants Citrine et Vénus de la YFP qui présentent des constantes de dissociation vis à vis des ions chlorures supérieures à  $10^4$  mM et des  $pH_{1/2}$  de 5,7 et 5,6 respectivement (table II.5) [Griesbeck et al., 2001, Nagai et al., 2002].

Compartiment cellulaire	pH <sup>a</sup> / pH <sub>1/2</sub> <sup>b</sup>	Protéines					
		YFP 7,00	EYFP 6,90	sEYFP 6,00	Venus 6,00	mCitrine 5,70	YPet 5,63
Mitochondrie	8,0	✓	✓	✓	✓	✓	✓
Cytosol	7,2	✗	✗	✓	✓	✓	✓
Noyau	7,2	✗	✗	✓	✓	✓	✓
Réticulum endoplasmique	7,2	✗	✗	✓	✓	✓	✓
Peroxisome	7,0	✗	✗	✗	✗	✓	✓
Appareil de Golgi	6,7 à 6,0	✗	✗	✗	✗	✗	✗
Endosome de recyclage	6,5	✗	✗	✗	✗	✗	✗
Endosome précoce	6,3	✗	✗	✗	✗	✗	✗
Endosome tardif	5,5	✗	✗	✗	✗	✗	✗
Lyosome	4,7	✗	✗	✗	✗	✗	✗

Tableau II.4 : Comportement des propriétés photophysiques des YFP suivant le compartiment intracellulaire, ✗ : inutilisable ( $pH < pH_{1/2} + 1$ ), ✓ : invariabilité ( $pH > pH_{1/2} + 1$ ). (a) [Casey et al., 2010], (b) table II.5

Protéine	Mutations de AvGFP	$\lambda_{abs}$ (nm)	$\lambda_{em}$ (nm)	$\epsilon$	$\phi_{fl}$	B	pH <sub>1/2</sub>
AvGFP <sup>a</sup>	–	495	510	11,8	0,80	9,44	4,5
YFP <sup>b,c</sup>	S65G-V68L-S72A-T203Y	514	528	83,0	0,61	50,6	7,0
EYFP <sup>c</sup>	S65G-V68L-T203Y-H231L	513	527	83,4	0,61	50,9	6,9
mCitrine <sup>f</sup>	YFP-Q69M	516	529	77,0	0,76	58,6	5,7
Venus <sup>d</sup>	EYFP-F46L-F64L-M153T -V163A-S175G	515	528	92,2	0,57	52,6	6,0
sEYFP <sup>d</sup>	EYFP-F64L-M153T-V163A-S175G	515	528	101	0,56	56,6	6,0
SYFP2 <sup>e</sup>	F46L-F64L-S65G-S72A-M153T-V163A S175G-T203Y-A206K	515	527	101	0,68	68,7	6,0
YPet <sup>f</sup>	S-EYFP-F46L-I47L-S72A-R79K-S208F -V224L-H231E-D234N	514	530	104	0,77	80,1	5,6

Tableau II.5 : Principales propriétés de AvGFP et de ses principaux dérivés émettant dans la partie jaune du spectre du visible.  $\epsilon$  désigne le coefficient d'absorption molaire ( $mM^{-1}.cm^{-1}$ ),  $\phi_{fl}$  le rendement quantique d'émission de fluorescence et B la brillance ( $mM^{-1}.cm^{-1}$ ), avec  $B = \phi_{fl} \cdot \epsilon$ . Les mutations F64L, M153T, V163A ou S175G ont été introduites afin d'améliorer les propriétés de repliement et de maturation de ces protéines. La protéine dont l'acronyme est sEYFP est appelé super-EYFP. [van Thor et al., 1998]<sup>a</sup>, [Tsien, 1998, Wachter et al., 1998]<sup>b</sup>, [Ormö et al., 1996]<sup>c</sup>, [Nagai et al., 2002]<sup>d</sup>, [Kremers et al., 2006]<sup>e</sup>, [Griesbeck et al., 2001]<sup>f</sup>

## E.2 Protéines fluorescentes jaunes d'autres organismes

Des organismes marins tels que le corail *Zoanthus sp.* [Matz et al., 1999], le vers marin *Branchiostoma lanceolatum* [Deheyn et al., 2007, Israelsson, 2013] et la méduse *Phialidium sp. SL-2003* [Pletneva et al., 2013], provenant de différents phyla, possèdent des gènes codant pour des protéines fluorescentes naturellement jaunes. Ces dernières ont une identité de séquence avec AvGFP inférieure à 50% (table II.6). Parmi ces PF naturelles, certaines ont des propriétés remarquables avec une brillance 3 fois supérieure à celle de la YFP et une insensibilité environnementale en pH ( $\text{pH}_{1/2} = 3,5$ ). En revanche, elles forment des oligomères ce qui présente des inconvénients en imagerie cellulaire.

Protéine (oligomérisation)	Identité de séquence avec AvGFP (%)	$\lambda_{ex}$ (nm)	$\lambda_{em}$ (nm)	$\epsilon^*$	$\phi_{fl.}$	B.*	$\text{pH}_{1/2}$
zFP538 <sup>a</sup> (tétramérique)	22,0	528	538	20	0,42	8,5	4,8
lanYFP <sup>b</sup> (tétramérique)	19,6	513	524	150	0,90	135,0	3,5
phiYFPv <sup>c</sup> (dimérique)	49,2	524	537	101	0,59	59,8	N.D.

Tableau II.6 : PF jaunes naturelles et leurs propriétés photophysiques, \* le coefficient d'extinction molaire  $\epsilon$  et la brillance B. sont exprimés en  $\text{mM}^{-1}.\text{cm}^{-1}$ . [Matz et al., 1999]<sup>a</sup>, [Shaner et al., 2013]<sup>b</sup>, [Pletneva et al., 2013]<sup>c</sup>.

## F Protéines fluorescentes du genre *Branchiostoma*

### F.1 *lanYFP* de *Branchiostoma lanceolatum*

L'embranchement des chordés a récemment fourni de nouvelles protéines fluorescentes naturelles particulièrement intéressantes pour leur forte brillance (*bfloGFP-a1*, *lanYFP*) ou leur maximum d'absorption dans le rouge (*lanRFP*) [Shaner et al., 2013, Pletnev et al., 2013, Bomati et al., 2014]. Le gène codant naturel *bIFP-Y3* découvert chez *B. lanceolatum* a été ingénieré en deux mutants brillants nommées *dlanYFP* et *mNeonGreen* [Shaner et al., 2013].

Les premiers travaux sur *bIFP-Y3* ont montré que les mutations I118K et N174T permettaient respectivement une dimérisation et un meilleur repliement [Shaner et al., 2004, Israelsson, 2013]. La monomérisation de *lanYFP* a utilisé une méthode couplant la prédiction de structure et le *molecular docking* (amarrage moléculaire) avec les serveurs *I-TASSER* et *ROSETTA-DOCK* [Chaudhury and Gray, 2008, Yang et al., 2015] grâce à la structure cristallographique de *ppluGFP2*, PF provenant d'un copépode [Wilmann et al., 2006]. Cette étape a permis de sélectionner les résidus cruciaux pour rompre les interfaces A/B et A/C qui sont définies selon le formalisme utilisé pour décrire la structure de DsRed [Yarbrough et al., 2001] (figure II.21).

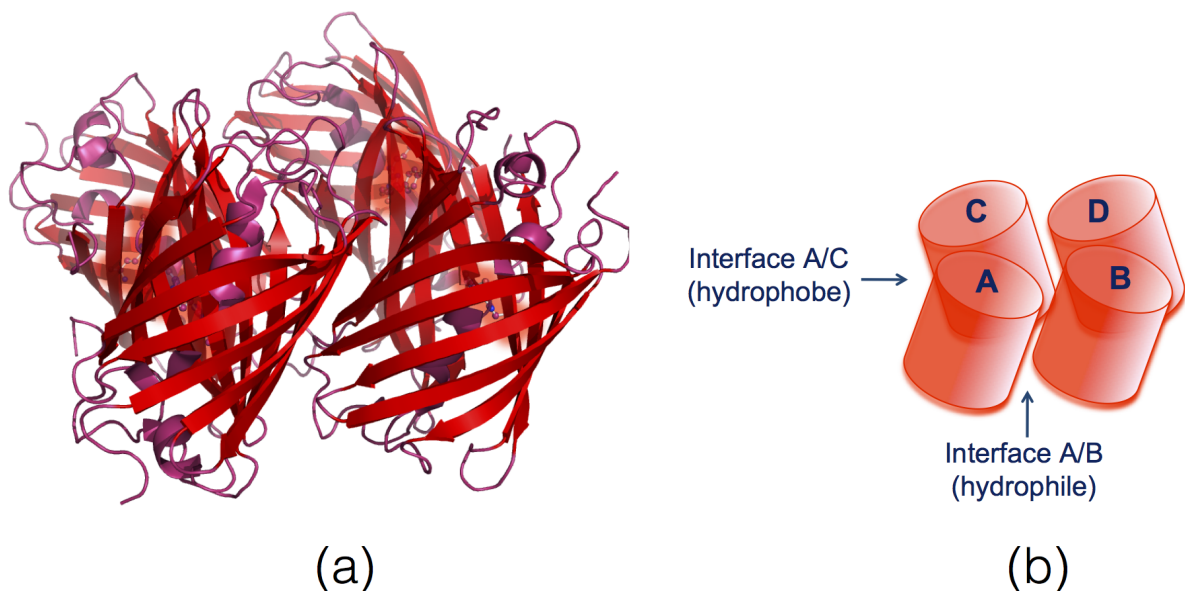


FIGURE II.21 : Formalisme décrivant les interfaces pour les tétramères naturelles de protéines fluorescentes, (a) protéines naturelles rouges DsRed (code PDB 1G7K), (b) représentation schématique des interactions entre monomères (adapté de [Shen et al., 2015])



L'introduction d'acides aminés avec des charges positives aux interfaces est une approche classique. Les étapes de destruction des interfaces ont été invariablement suivies par un cycle de restauration de l'émission de fluorescence par évolution dirigée. Les interfaces A/B et A/C présentes dans la structure prédite par *I-TASSER* ont été détruites successivement par l'introduction de deux et cinq mutations pour *dlanYFP* (I118K, N174T) et mNeonGreen (V140R, L144T, D156K, T158S, Q168R). La destruction des interfaces A/B et A/C entraîne invariablement des pertes aux niveaux des propriétés d'émission. La restauration de la fluorescence est réalisée avec trois mutations pour *dlanYFP* (A45D, S163N, V171A) et neuf mutations pour mNeonGreen (F15I, Q56H, F67Y, K79V, S100V, F115A, T141S, M143K, I185Y). L'intermédiaire dimérique obtenu au cours de ce travail a servi de modèle de base pour l'étape finale de monomérisation. L'évolution dirigée de *lanYFP* en mNeonGreen a nécessité au final une série de 21 mutations.

Protéine (oligomérisation)	$\lambda_{ex.}$ (nm)	$\lambda_{em.}$ (nm)	$\epsilon$ $M^{-1}.cm^{-1}$	$\phi_{fl}$	B. relative EGFP (%)	pH <sub>1/2</sub>
<i>lanYFP</i> (tétramérique)	513	524	150	0,95	311	3,5
<i>dlanYFP</i> (dimérique)	513	524	125	0,90	245	ND
mNeonGreen (monomérique)	506	517	116	0,80	202	5,7

Tableau II.7 : Propriétés photophysiques de *lanYFP* et de ses mutants [Shaner et al., 2013]

La caractérisation de mNeonGreen révèle un fort potentiel pour l'imagerie de fluorescence puisque la brillance de ce mutant est 2 fois supérieure à EGFP et 1,5 fois supérieure à mCitrine. La photostabilité de mNeonGreen est similaire à celle sfGFP (sf, super folder) et EGFP mais supérieure à celle de EYFP. La constante de temps de maturation du chromophore GYG est inférieur à 10 min et la sensibilité aux halogénures comme les ions chlorures n'a pas été rapportée. En revanche, le pH<sub>1/2</sub> de mNeonGreen est dégradé par rapport à *lanYFP* et sa valeur de pH<sub>1/2</sub> égale à 5,6 est similaire à mCitrine et Venus [Shaner et al., 2013].

Les paramètres exceptionnels de mNeonGreen montrent que la stratégie de monomérisation appliquée à la protéine fluorescente sauvage *bFP-Y3* pour obtenir un mutant monomérique est adaptée à l'élaboration de nouvelles PF performantes.

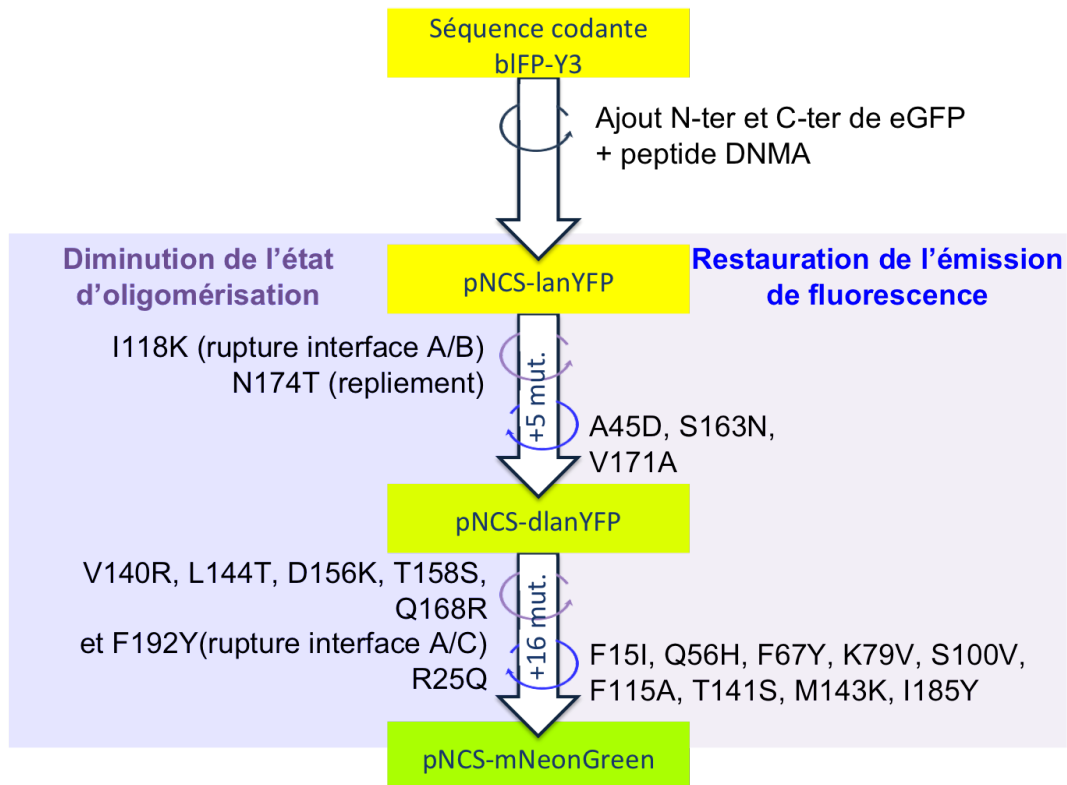


FIGURE II.22 : Schéma de principe de la monomérisation de lanYFP en mNeonGreen avec la protéine fluorescente jaune dimérique dlanYFP.

## F.2 Autres protéines fluorescentes du genre *Branchiostoma*

Une recherche de séquence codante d'homologue de AvGFP provenant du genre *Branchiostoma* dans la base de données du Centre Américain pour les Informations Biotechnologiques (NCBI) [Geer et al., 2009] permet d'identifier 71 séquences polypeptidiques dont plus de la moitié sont protégées par des brevets [Ledford, 2013].



Protéine (oligomérisation)	$\lambda_{ex.}$ (nm)	$\lambda_{em.}$ (nm)	$\epsilon$ $\text{mM}^{-1}.\text{cm}^{-1}$	$\phi_{fl}$	B. relative EGFP (%)	pH <sub>1/2</sub>
<i>bflo</i> GFP-a1 (dimérique)	497	512	121	1,04	381	3,0
<i>lan</i> RFP (dimérique)	521	592	71	0,10	22	4,7
<i>bflo</i> GFP-c1 (dimérique)	493	521	99	0,15	< 1	N.D.
<i>lan</i> GFP (tétramérique)	502	511	ND	N.D.	N.D.	N.D.

Tableau II.8 : PF dérivées d'organisme du genre *Branchiostoma* et leurs principales caractéristiques photophysiques

## **G Applications méthodologiques des protéines fluorescentes en biologie cellulaire**

### **G.1 Applications des protéines fluorescentes en microscope de fluorescence**

Les propriétés de fluorescence des PF sont exploitées pour étudier une gamme étendue de propriétés et de comportement en biologie cellulaire et médicale [Zimmer, 2002, Zimmer, 2015]. Les principales raisons de ce succès sont dues à la cyclisation post-traductionnelle de son chromophore sans ajout de co-facteur ou de substrat, la conservation de l'activité et de la mobilité des protéines fusionnées avec AvGFP et leur cytotoxicité modérée. De part leur structure compacte abritant le chromophore du solvant, les PF résistent dans une certaine mesure, à la chaleur, aux détergents, au photo-blanchiment, aux chaotropes et à la plupart des protéases, leur conférant des propriétés de résistance essentielles aux études en biologie cellulaire. La combinaison de la microscopie de fluorescence avec les PF permet la visualisation de phénomènes dynamiques sur des cellules ou des tissus vivants en particulier avec les techniques de microscopie confocale et de super-résolution. Une liste non exhaustive d'applications des PF en microscopie de fluorescence permet d'entrevoir leur large gamme d'utilisation.

Les propriétés de fluorescence des PF leur donne un rôle de marqueur de colocalisation en biologie cellulaire avec une première application comme rapporteur d'expression génique [Chalfie et al., 1994]. Les PF sont utilisées comme étiquette de protéine de fusion dans de nombreux organismes tel que les virus [Ward and Moss, 2001], les oeufs de xénopes [Polenzani et al., 1991] et les cellules de mammifères [Kanda et al., 1998].

Les PF peuvent servir à sonder le compartiment cellulaire grâce à leur sensibilité environnementale, aux ions calcium, halogénures et au pH [Miyawaki et al., 1997, Miyawaki et al., 1999, Wachter and Remington, 1999]. La mobilité cellulaire est étudiée grâce au PF via les techniques de redistribution de fluorescence après photo-blanchiment (FRAP) et de perte de fluorescence induite par photo-blanchiment (FLIP) [Phair and Misteli, 2000]. Le photo-blanchiment transitoire ou permanent des PF est utilisé afin d'étudier la vitesse de déplacement de molécules marquées par des PF (FRAP) ou de mettre en évidence des continuités dans les structures cellulaires (FLIP)

Avec le concours des PF, les interactions protéine / protéine sont déterminées par la technique

de transfert d'énergie résonant de Förster (FRET) [Valeur and Berberan-Santos, 2012, Lakowicz, 2013] et de bioluminescence (BRET) [Xu et al., 1999]. La technique de FRET nécessite une source extérieure de lumière pour exciter le donneur alors que le BRET utilise une protéine qui émet spontanément des photons par bioluminescence. La colocalisation détermine si deux protéines sont associées dans une même structure, tel que l'endosome, les mitochondries, les microtubules [Dunn et al., 2011].

## G.2 Principe du FRET

Le transfert d'énergie par résonance de type Förster (FRET) est répandu dans de nombreuses applications utilisant la fluorescence comme certains diagnostics médicaux [Kobayashi et al., 2009, Conway et al., 2014], l'analyse d'ADN [Didenko, 2001] ou l'imagerie de fluorescence [Ellenbroek and van Rheenen, 2014]. Dans ce domaine, la technique de FRET est une méthode largement utilisée pour estimer la distance entre deux sites d'une protéine, pour étudier l'interaction entre protéines. La technique de FRET repose sur le marquage de deux partenaires moléculaires d'intérêt par deux fluorophores, ici deux PF, dont une va jouer le rôle de donneur et l'autre d'accepteur [Lakowicz, 2013].

Un transfert résonnant d'énergie se produit entre le donneur ( $D_{FRET}$ ) et l'accepteur ( $A_{FRET}$ ) dont la vitesse  $k_{\tau}(r)$  dépend de la distance,  $r$ , entre  $D_{FRET}$  et  $A_{FRET}$ , du facteur d'orientation entre les moments de transition dipolaires du chromophore de  $D_{FRET}$  et de  $A_{FRET}$  ( paramètre  $\kappa^2$ ) et de l'intégrale de recouvrement  $J_{DA}$  entre le spectre d'émission du donneur et le spectre d'absorption de l'accepteur, selon :

$$k_{\tau}(r) = \left( \frac{90000(\ln 10)}{128\pi^5 N_A n^4} \right) \frac{\phi_D \kappa^2}{\tau_D r^6} \underbrace{\int_0^{\infty} F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda}_{\text{intégrale de recouvrement } J_{DA}(\lambda)} \quad (\text{II.3})$$

$$\text{En posant } R_0^6 = 8.79 \times 10^{23} \phi_D \kappa^2 n^{-4} J_{DA}(\lambda) \quad (\text{II.4})$$

$$k_{\tau}(r) = \frac{1}{\tau_D} \left( \frac{R_0}{r} \right)^6 \quad (\text{II.5})$$

Avec  $N_A$  le nombre d'Avogadro,  $n$  l'indice de refraction des biomolécules,  $\phi_D$  le rendement quantique du donneur  $D_{FRET}$ ,  $\kappa$  le facteur d'orientation des moments dipolaires de transition

des deux fluorophores,  $F_D(\lambda)$  (u.a.) l'intensité de l'émission de fluorescence du donneur en fonction de la longueur d'onde et  $\epsilon_A(\lambda)$  ( $M^{-1}.cm^{-1}$ ) le coefficient d'extinction molaire de l'accepteur  $A_{FRET}$ . La valeur de  $R_0$  est fortement dépendante de la longueur d'onde à laquelle le recouvrement  $J_{DA}(\lambda)$  s'opère ( $M^{-1}.cm^{-1}.mm^4$ ). L'efficacité de FRET,  $E$ , qui représente le nombre de photons absorbés par le donneur et transmis par transfert d'énergie résonant à l'accepteur est défini selon les équations II.6 :

$$E = \frac{k_{\tau}(r)}{\tau_D^{-1} + k_{\tau}(r)} \quad \text{ou} \quad E = \frac{R_0^6}{R_0^6 + r^6} \quad (\text{II.6})$$

La distance critique de Forster,  $R_0$ , correspond à la distance  $D_{FRET}/A_{FRET}$  pour laquelle l'efficacité de FRET est de 50%. La mesure de l'efficacité FRET n'a de sens que pour des distances  $D_{FRET}/A_{FRET}$  appartenant à l'intervalle  $0,5 R_0 < r < 2 R_0$ . Dans ces conditions, on se trouve dans une zone de variation importante de l'efficacité  $E$ .

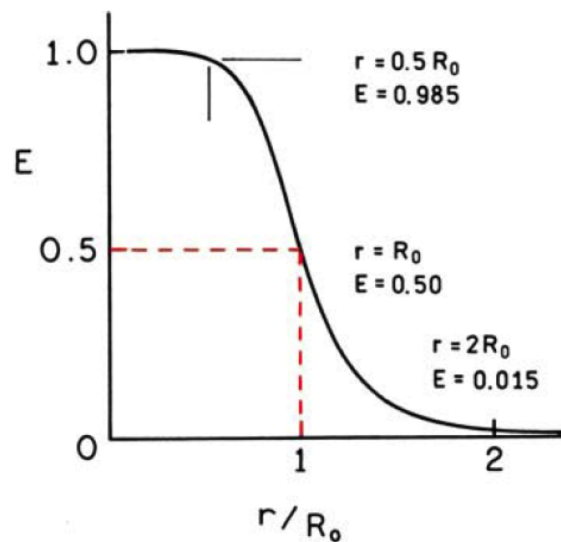


FIGURE II.24 : Evolution de l'efficacité de FRET ( $E$ ) en fonction du rapport  $r/R_0$  [Lakowicz, 2006]

Plusieurs méthodes permettent la visualisation voire la quantification du phénomène de FRET [Jares-Erijman and Jovin, 2003, Jares-Erijman and Jovin, 2006]. Certaines sont basées sur la mesure de la réduction de l'intensité ou de la durée de vie d'émission de fluorescence de  $D_{FRET}$ , de l'augmentation de l'intensité d'émission de  $A_{FRET}$ . Dans ces conditions, l'efficacité est estimée selon :

$$E = 1 + \frac{F_{DA}}{F_D} \quad \text{ou} \quad E = 1 + \frac{\tau_{DA}}{\tau_D} \quad (\text{II.7})$$

Toutes modifications des propriétés intrinsèques du donneur (rendement quantique, durée de vie et spectre d'émission) ou de l'accepteur (spectre d'absorption) indépendamment des changements de distance et d'orientation entre les fluorophores intrinsèques au processus de FRET peuvent être à l'origine d'artéfacts de mesure biaisant la quantification de l'efficacité de FRET.

### **G.3 Donneur et accepteur de FRET : un cahier des charges**

En supplément des critères classiques de distance, d'orientation et de recouvrement spectral présentés ci-dessus, d'autres conditions sont nécessaires pour permettre une analyse quantitative des signaux de FRET. Elles concernent les propriétés intrinsèques des protéines fluorescentes donneur et accepteur. En particulier, le donneur comme l'accepteur doivent être caractérisés par une forte brillance et des propriétés photophysiques simples avec des propriétés spectrales et cinétiques (déclins d'émission de fluorescence) homogènes. Ils doivent se comporter comme des monomères et leur chromophore doit maturer efficacement. Leur photostabilité et leur capacité à subir des réactions photo-induites doivent être adaptées à la méthode utilisée pour l'analyse des signaux de FRET. Enfin, ni le donneur, ni l'accepteur ne doivent présenter de sensibilité environnementale.

### **G.4 Donneurs et accepteurs de FRET**

#### **G.4.a Donneurs de FRET à la fluorescence cyan**

Les CFP sont des protéines fluorescentes ayant une émission de fluorescence dans le cyan (~475 nm), due à un chromophore présentant un tryptophane en position 66, la première mise au point s'appelant l'ECFP [Heim et al., 1994, Heim and Tsien, 1996]. De nombreux travaux ont été réalisés afin d'optimiser ECFP [Lelimousin et al., 2009, Fredj, 2012, Goedhart et al., 2012, Erard et al., 2013]. En effet, cette protéine présente des propriétés photophysiques complexes avec un rendement quantique et un coefficient d'absorption molaire faibles, un déclin d'émission de fluorescence multi exponentiel [Rizzo et al., 2004] et une forte sensibilité environnementale [Villoing et al., 2008]. Sur la base d'études structurales, photophysiques et théoriques de calculs quantiques et de dynamiques moléculaires [Lelimousin et al., 2009, Vallverdu et al., 2010, Jonasson et al., 2011], il a été montré que la photophysique de la CFP est gouvernée par la dynamique de la protéine et du chromophore dans sa cavité. Ces travaux ont permis



d'identifier différents sites de la protéine jouant un rôle clé dans sa dynamique et ont débouché sur l'élaboration de mutants aux propriétés quasi-idéales (Table II.9). En particulier, les deux mutants les plus performants Aquamarine et mTurquoise2 présentent un rendement quantique de l'ordre ou supérieur à 0,9 soit une amélioration d'un facteur 3 par rapport à ECFP [Goedhart et al., 2012, Erard et al., 2013]. En revanche, le coefficient d'extinction molaire de ces CFP conserve une valeur de  $\sim 30 \text{ mM}^{-1} \cdot \text{cm}^{-1}$  qui semble intrinsèque au chromophore de type XWG. Leur brillance atteint 80% de celle de EGFP ce qui semble être la limite pour ce type de chromophore. Enfin, l'amélioration de la photophysique de ces protéines a été de pair avec la réduction de leur sensibilité environnementale avec des  $\text{pH}_{1/2}$  dorénavant inférieurs à 3,5, à l'amélioration de leur photostabilité et à l'élimination des processus de photoconversion réversible. Leur utilisation dans tous les compartiments cellulaires est désormais possible.

Protéine	Mutation relative à ECFP (nombre)	$\lambda_{ex.}$ (nm)	$\lambda_{em.}$ (nm)	$\phi$	$\epsilon^a$	$B_{rel.}^b$ (%)	$\text{pH}_{1/2}$	$\langle \tau \rangle$ (ns)
ECFP	–	433	476	0,40	30	35,7	5,6	2,50
SCFP3A	S72A-H148D- S175G -A206K (4)	433	474	0,56	30	50,0	4,5	–
Cerulean	S72A-Y145A-H148D (3)	433	475	0,62	43	79,4	5,2	3,06
mTurquoise	T65S-S72A-H148D -S175G-A206K (5)	434	474	0,84	30	75,0	3,4	4,06
Aquamarine	T65S-H148G (2)	433	474	0,89	26	68,9	3,3	4,12
mTurquoise2	T65S-S72A-H148D -S175G-A206K-I146F (6)	434	474	0,93	30	83,0	3,1	3,80

Tableau II.9 : Propriétés photophysiques des donneurs de FRET dérivés de la ECFP.  $\epsilon$  (a) et  $B_{rel.}$  (b) sont respectivement exprimés en  $\text{mM}^{-1} \cdot \text{cm}^{-1}$  et en pourcentage brillance relative à EGFP.

Enfin, malgré des performances photophysiques très différentes, toutes les CFP présentent des spectres d'absorption et d'émission de fluorescence de formes semblables avec une forme caractéristique de double bosses due au noyau indole du chromophore. Le maximum d'absorption est situé vers 434 nm et celui d'émission de fluorescence vers 474 nm. L'étalement du spectre d'émission de fluorescence s'effectue sur plus de 200 nm ce qui peut être un inconvénient dans les études de FRET. En effet, suite à l'excitation du donneur, cette émission peut recouvrir celle provenant de l'accepteur et être collectée dans le canal spectral de l'accepteur en plus de l'intensité émise par ce dernier suite au FRET.

#### **G.4.b Accepteurs de FRET pour dérivés de la *Cyan Fluorescent Protein***

Les accepteurs de FRET historiques des dérivés de la CFP sont la YFP et ses mutants. Néanmoins, ces dernières présentent une forte sensibilité environnementale limitant les analyses quantitatives des signaux de FRET (voir paragraphe II.G.2). La découverte de *lanYFP* ouvre de nouvelles perspectives puisque celle-ci présente des propriétés performantes avec une forte brillance et une insensibilité au pH. Néanmoins son caractère tétramérique est un obstacle pour l'imagerie de FRET ce qui a motivé l'élaboration de mNeonGreen [Shaner et al., 2013]. La monomérisation de *lanYFP* a permis de préserver des propriétés photophysiques performantes malgré une dégradation de la sensibilité en pH avec un  $pH_{1/2}$  de 5,6 qui est similaire à celui de la citrine.

## Etude par dynamique moléculaire des réseaux de liaisons hydrogène de AvGFP

### A Les liaisons hydrogène dans les systèmes biologiques

#### A.1 Rôle et définition

Compte tenu de leur caractère fortement directionnel et de leur grand nombre dans les systèmes biologiques, les liaisons hydrogène (LH) jouent un rôle de premier plan dans la formation et la stabilisation de la structure secondaire, tertiaire et quaternaire des systèmes biologiques [Hubbard and Kamran Haider, 2010, Pace et al., 2014]. Elles jouent également un rôle crucial dans la fonction de ces systèmes. Leur énergie intermédiaire entre les interactions de van der Waals et les liaisons covalentes leur permet de se former et de se rompre rapidement à température ambiante, conférant à ces systèmes biologiques, une dynamique conformationnelle nécessaire à leur fonction.

La liaison hydrogène est une interaction attractive qui se forme entre un atome d'hydrogène H lié de façon covalente à un atome électronégatif ,D ,et un atome électronégatif voisin, A, porteur d'un doublet non liant. D et A sont respectivement donneur et accepteur de proton. Cette interaction présente une énergie qui varie entre 2 et 20 kJ/mol. Elle est plus forte que les liaisons de van der Waals classiques (entre 1 à 4 kJ/mol) tout en restant plus faible que les liaisons covalentes qui pour leur part ont des énergies entre 200 et 800 kJ/mol [Gerschel, 2012]. Dans les systèmes biologiques, les donneurs classiques sont N-H, O-H et S-H alors que les accepteurs sont les atomes d'azote, d'oxygène et de soufre portant au moins un doublet non liant. Baker et Hubbard ont mené une étude très approfondie afin d'identifier les partenaires préférentiels impliqués dans les liaisons hydrogène observées dans les protéines globulaires

mettant en jeu le squelette carboné et/ou les chaînes latérales des résidus polaires [Baker and Hubbard, 1984]. Ippolito et coll. se sont quant à eux intéressés à la stéréochimie préférentielle des liaisons hydrogène impliquant les chaînes latérales des résidus polaires grâce à l'analyse de la structure cristallographique de 50 protéines [Ippolito et al., 1990]. Dans les deux études, des configurations récurrentes des liaisons hydrogènes ont été observés, indépendamment de la structure de la protéine. En particulier, Ippolito et al ont montré que la stéréochimie des liaisons hydrogène est gouvernée par la configuration électronique de l'accepteur, l'accessibilité du donneur de proton D et enfin par la conformation des chaînes latérales des résidus. Pour chacun des 13 résidus (parmi les 20) portant au niveau de leur chaîne latérale un atome d'oxygène, d'azote ou de soufre susceptible d'être impliqué dans une liaison hydrogène, la distance moyenne D-A constatée dans ces 50 protéines a été calculée. Ces distances ne dépassent pas 3 Å dans le cas des résidus portant un oxygène ou un azote et montent à 3,5 Å en présence de soufre.

La liaison hydrogène présente un caractère directionnel imposé par le caractère directionnel des doublets non liants de l'accepteur. La liaison hydrogène implique un recouvrement de l'orbitale 1s de l'hydrogène avec celui du doublet non liant. En dynamique moléculaire classique, l'orientation du nuage électronique n'est pas pris en compte. Ce dernier est implicitement représenté par une charge effective. Dans ce contexte des critères géométriques de distances et d'angles doivent être utilisés pour décrire les liaisons hydrogène. Nous avons utilisé les critères d'existence des liaisons hydrogène par défaut de la suite AMBER qui sont une distance  $d$  inférieure à 3,2 Å et un angle  $\Omega$  supérieur à 120° (figure III.1). Dans la suite nous avons choisi de représenter les LH avec une orientation. Celle-ci est schématisée par une flèche de A vers D correspondant à la direction de déplacement des électrons lors d'un potentiel transfert de proton. En d'autres termes le sens de cette flèche est l'inverse de celui d'une translocation de proton (figure III.1).

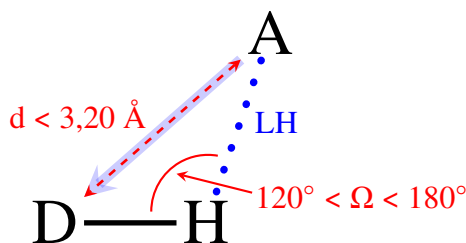


FIGURE III.1 : Critères qui définissent une liaison hydrogène (LH) avec une distance  $d$  entre D et A et l'angle  $\Omega$  ( $\widehat{DHA}$ ). Les valeurs les plus souvent utilisées sont telles que  $d$  est inférieure à 3,2 Å et que  $\Omega$  est compris entre 120° et 180°.

## A.2 Réseaux de liaisons hydrogène et chemin de translocation de protons

Les liaisons hydrogène peuvent s'interconnecter si elles partagent un atome qui dans ce cas sera à la fois donneur et accepteur. Les liaisons hydrogène qui sont connexes forment de proche en proche un réseau de liaisons hydrogène. Ces réseaux peuvent être le support d'une translocation de protons qui joue un rôle essentiel dans la réactivité et la fonction de nombreux systèmes biologiques. La translocation de protons et la recherche des réseaux de LH qui les soutiennent ont été étudiées dans de nombreux systèmes tels que la bactériorhodopsine, une pompe à proton transmembranaire [Kandori, 2000, Bondar et al., 2004], le complexe IV de la chaîne respiratoire [Taraphder and Hummer, 2003, Peng and Voth, 2012], le centre réactionnelle bactérien [Paddock et al., 2003] ou encore les protéines fluorescentes [Agmon, 2005, Shinobu et al., 2010]. Par exemple, certaines protéines fluorescentes sont le lieu d'un transfert de proton à l'état excité (ESPT) qui gouverne leur photophysique.

L'objectif de notre travail est de mettre au point un algorithme permettant la recherche systématique des réseaux de LH dans les protéines et d'apporter des informations sur la dynamique de ces réseaux par le biais de simulations par dynamique moléculaire (voir partie VI.D). Pour plus de simplicité nous utiliserons le vocabulaire suivant :

- **Réseau de LH** : ensemble des LH qui sont connexes à un instant  $t$  (structure cristallographique ou un pas de dynamique)
- **Chemin de LH orientées** : réseau de LH où toutes les liaisons hydrogène à un instant  $t$  ont une orientation telle qu'il peut y avoir un transfert de proton de l'atome de départ, donneur de proton initial, vers le dernier atome du chemin, l'accepteur final.
- **Réseau global de LH pour un atome donné** : tous les réseaux de LH passant par un atome donné (par exemple, l'oxygène du phénol du chromophore de AvGFP) sont cumulés sur l'ensemble de la dynamique. Ils forment le réseau global de LH passant par cet atome. Au sein du réseau global de LH, la connectivité de toutes les LH n'est pas obligatoirement effective à chaque pas de la dynamique moléculaire. Chacune de ces LH présente une densité de probabilité d'existence qui est corrélée au nombre d'instantanés où elles sont observées au cours de la dynamique moléculaire. Cette densité de probabilité ( $dP_{HB_i}$ ) est calculée selon l'équation III.1 où  $N$  est égal au nombre total d'instantanés considérés au cours d'une dynamique moléculaire,  $n$  le numéro d'un instantané et  $P_{HB_i}(n)$  désigne la probabilité d'existence de cette LH à cet instantané  $n$ .

$$dP_{HB_i} = \frac{\sum_{n=1}^n P_{HB_i}(n) > 0}{N} \quad (\text{III.1})$$

- **Chemin de translocation** : ensemble des chemins orientés faisant partie d'un réseau globale susceptible de sous tendre un transfert de proton ou translocation de proton d'un donneur de proton initial,  $D_i$ , à un accepteur final,  $A_f$ , sur l'ensemble de la dynamique. Par cette définition nous éliminons des sous-ensembles du réseau global de LH entre  $D_i$  et  $A_f$  et pour lesquelles les orientations des différentes LH ne sont pas compatibles avec un translocation de proton. En d'autres termes, parmi ces LH, au moins une possède une orientation distincte. De façon générale, la vitesse de transfert de proton ( $k_{PT}$ ) est une fonction complexe qui dépend du chemin physique emprunté par le proton (nombre d'atomes relais, longueur du chemin) et de la probabilité de transfert du proton entre chacun des atomes D et A qui composent le chemin de translocation :

$$k_{PT} = f(\text{chemin de translocation}, P(\text{transfert du proton})) \quad (\text{III.2})$$

## B Recherche des réseaux de liaisons hydrogène dans les protéines

### B.1 Contexte

La recherche des réseaux de LH s'effectue de façon usuelle sur la base de l'analyse de la structure cristallographique des protéines qui donne une représentation statique des liaisons hydrogène et des réseaux de LH intra-protéine. Cependant, l'amélioration de la résolution des structures cristallographiques offre la possibilité d'apporter des informations additionnelles sur l'existence de LH statiques sans pour autant apporter d'informations sur leur dynamique. Par exemple, l'obtention de structures à haute résolution ( $<1 \text{ \AA}$ ) permet uniquement d'affiner les connaissances sur la position relative des protons dans les structures cristallographiques [Shinobu et al., 2010]. Récemment, une nouvelle approche a été développée. Elle consiste à exploiter des dynamiques moléculaires afin d'en extraire la trajectoire des distances inter-atomiques entre les atomes relais qui composent un chemin de translocation [Simkovitch et al., 2013, Wineman-Fisher et al., 2014]. Cette approche offre la possibilité de confirmer ou de proposer des chemins de translocation sans pour autant pouvoir en évaluer la dynamique, à savoir d'estimer avec quelle probabilité l'ensemble des LH de ce chemin sont connexes.

Notre objectif est de pouvoir rechercher (i) l'ensemble des réseaux globaux de LH dans une protéine, (ii) celui qui est connecté à un atome d'intérêt et (iii) d'étudier leur dynamique. L'approche suivie au cours de ces travaux est dans une certaine mesure assez proche de celle proposée dans la littérature puisqu'elle repose sur l'analyse de dynamiques moléculaires. En revanche, elle implique la mise au point d'un algorithme permettant la recherche puis l'agrégation des LH connexes pour former des réseaux et ce pour chaque instantané de la simulation. Cet algorithme a également comme objectif d'extraire le réseau global de LH contenant un atome d'intérêt. Ces données sont analysées afin d'estimer la densité de probabilité d'existence des LH constituant ce réseau global de LH et notamment celles intervenant dans les chemins de translocation. Cet algorithme offre de plus la possibilité d'identifier l'ensemble des chemins de translocation entre le donneur de proton initial,  $D_i$ , et l'accepteur final  $A_f$ , d'estimer leur densité de probabilité et de pointer les maillons faibles au sein de ces chemins. Enfin, la fréquence du passage des réseaux LH dans le solvant peut être évaluée.

Pour valider cet algorithme, nous avons choisi de l'appliquer à l'étude du chemin de translocation susceptible de sous-tendre le transfert de proton à l'état excité observé chez la AvGFP et certains de ses dérivés. Ce processus a été en effet très étudié dans la littérature et de nombreuses données expérimentales concernant notamment la vitesse de transfert de proton ( $k_{PT}$ ) sont disponibles. De plus, sur la base de l'analyse de la structure cristallographique, et de l'étude de la dynamique de LH spécifiques, obtenue par dynamique moléculaire, les auteurs ont proposé l'existence de divers chemins de translocation alternatifs au chemin usuel pour certains dérivés de la AvGFP [Wineman-Fisher et al., 2014, Shinobu and Agmon, 2015]. Ainsi, l'application de notre algorithme à ces protéines nous permettra de confronter les résultats issus de notre algorithme à ceux de la littérature.

Dans cet optique, nous avons sélectionné en plus de AvGFP, 4 de ses mutants dont la vitesse  $k_{PT}$  est modulée et pour lesquelles une structure cristallographique est disponible. A partir de ces structures, nous avons estimé la longueur des chemins de translocation proposés dans la littérature. Il en ressort que, indépendamment du chemin suivi, cette longueur est assez similaire dans les différentes protéines malgré une grande variation de  $k_{PT}$ . Ceci montre que le critère de longueur n'est pas le seul paramètre gouvernant  $k_{PT}$ . La dynamique et la concomitance des LH composant ces chemins sont d'autres critères non intégrés dans la structure, susceptibles de moduler  $k_{PT}$ , ce qui a motivé notre approche.

Mutants	Code PDB (nb. de molécules dans l'unité asymétrique)	pH	$k_{PT}$ ( $s^{-1}$ )	longueur (Å)
AvGFP	1W7S (4)	7,80	$1,25 \times 10^{11}$	$7,91 \pm 0,08$
AvGFP-S205V	2QLE (4)	8,00	$4,10 \times 10^9$	$8,61 \pm 0,15$
AvGFP-T203V-S205A	4OGS (2)	7,80	$3,60 \times 10^8$	–
AvGFP-S65T	1C4F (1)	4,60	–	9,71
AvGFP-T203V-S205V	–	–	–	–

Tableau III.1 : Mutants dérivés de AvGFP étudiés. Sont indiqués le code PDB de la structure cristallographique choisie, le pH de cristallisation (pH), la vitesse de transfert de proton à l'état excité  $k_{PT}$  et la longueur du chemin de translocation de protons proposé dans la littérature. Cette longueur a été estimée à partir des structures cristallographiques en mesurant la distance entre les atomes lourds du chemin (moyenne établie à partir de tous les monomères des structures cristallographiques).

Dans ce qui suit, nous présenterons, dans un premier temps, l'algorithme mis au point pour la recherche des réseaux de LH dans les protéines. En second lieu, le choix des structures, des coordonnées de départ et des états de protonation pour les dynamiques moléculaires sera



discuté. Une première analyse des dynamiques en portant une attention particulière sur la stabilité des structures secondaires et aux mouvement locaux de résidus clefs sera ensuite effectuée. Finalement, nous présenterons les résultats obtenus par l'algorithme pour chacune des cinq protéines étudiées en évaluant ses performances et en discutant les différents chemins de translocation identifiés par l'algorithme.

## **B.2 Algorithme de recherche des réseaux de liaisons hydrogène au sein d'une protéine**

### **B.2.a Structure générale de l'algorithme**

L'algorithme mis au point permet la recherche et l'agrégation de LH au cours d'une dynamique moléculaire ainsi que la caractérisation des réseaux de LH ainsi formés. La figure III.2 donne une représentation graphique de l'ensemble des actions séquentielles accomplies par l'algorithme. Il est structuré en trois phases :

- **Entrée des données** : cette étape vise à spécifier les paramètres à appliquer lors de la recherche des réseaux de LH ainsi que les trajectoires de dynamique moléculaire à exploiter.
- **Recherche des réseaux de LH** : l'identification et l'agrégation des liaisons hydrogène en réseaux de LH sont effectuées au cours de cette phase.
- **Analyse des données** : la caractérisation des réseaux de LH, leur visualisation et l'extraction des données essentielles à la compréhension de la dynamique de ces réseaux et de leur physico-chimie sont réalisées à ce stade.

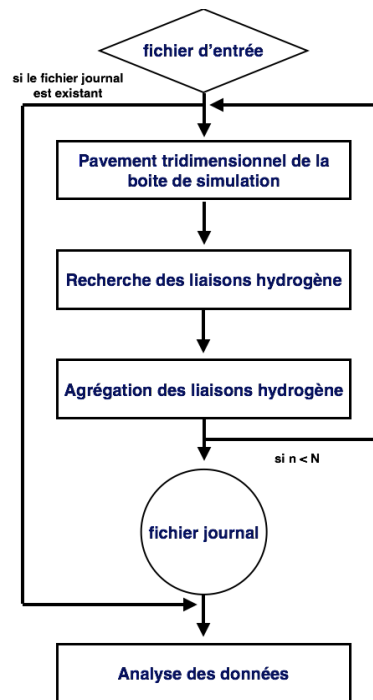


FIGURE III.2 : Structure de l'algorithme de recherche des liaisons hydrogène, de leur agrégation et de leur analyse.  $n$  correspond au numéro séquentiel de l'instant analysé et  $N$  nombre total d'instantanées constituant la trajectoire de dynamique moléculaire à analyser.

### B.2.b Entrée des données

L'utilisateur contrôle les paramètres de l'algorithme via un fichier source, unique. C'est un fichier texte dont les paramètres peuvent être modifiés selon le souhait de l'utilisateur et en accord avec les conseils dispensés dans le fichier source. Certains paramètres peuvent être commentés avec l'aide d'un point d'exclamation. La lecture combinée des fichiers de topologie et d'une trajectoire de dynamique moléculaire permet de connaître la nature, la charge, les coordonnées et la vitesse de chaque atome qui constituent une dynamique moléculaire. Le fichier de topologie est commun à l'ensemble des trajectoires d'une dynamique moléculaire. Par conséquent, il est lu en début d'analyse par l'algorithme qui garde en mémoire les informations sur les atomes telles que leur nom, leur numéro et le nom du résidu dont ils font partie. A ce niveau, les atomes de la protéine, les contre-ions ainsi que les atomes de solvant sont traités séparément. Un numéro séquentiel est attribué à chacun et selon l'ordre protéine / contre-ion / solvant. Cette numérotation est conservée tout au long des calculs de l'algorithme.

### **B.2.c Typologie des atomes : nature donneur/accepteur de proton**

La définition de donneur et accepteur de proton est gérée par un fichier source (do\_acc\_AA.in). Tous les résidus de la protéine génèrent des liaisons hydrogène via au minimum les atomes O et N(H), respectivement accepteur et donneur de proton, des groupes carbonyle et N-H formant la liaison peptidique. De plus, parmi les 20 résidus couramment rencontrés dans les protéines, 13 possèdent une chaîne latérale dont certains atomes peuvent jouer le rôle de donneur ou d'accepteur de proton (table III.2). La nature de ces atomes (donneur ou accepteur) est également dépendante de l'état de protonation du résidu considéré comme dans le cas des 3 résidus acide aspartique (ASP, ASH), acide glutamique (GLU, GLH) et histidine (HIE, HID, HIP) ou encore de leur état d'oxydoréduction comme la cystéine (CYS, cystéine oxydée).

La liste des D/A de proton a été complétée par 6 atomes dits spéciaux qui permettent de décrire les LH intramoléculaires pour les résidus dits protonables. L'ajout de ces atomes dans la liste des D/A de proton permet à l'algorithme d'inclure la résonance des 3 résidus acide aspartique, acide glutamique et histidine. La liste D/A contient tous les résidus décrits de façon homogène avec le format AMBER. Cette liste décrit 61 atomes donneur de proton et 52 atomes accepteur de proton. La contribution en atomes D/A des résidus non standards tels que le chromophore et les molécules d'eau est 5 atomes donneur et 6 atomes accepteur.

résidus	Code AMBER	donneur (proton)	accepteur
Ser	SER	OG(HG)	OG
Cys	CYS	SG(HG)	SG
”	CYX	–	SG
Thr	THR	OG1(HG1)	OG1
Asp	ASP	OD2(CG)	OD1, OD2
”	ASH	OD2(HD2), <b>OD2(CG)</b>	OD1, OD2
Asn	ASN	ND2(HD21), ND2(HD22)	OD1
Glu	GLU	<b>OE1(CD)</b>	OE1, OE2
”	GLH	<b>OE1(CD)</b> , OE2(HE2)	OE1, OE2
Gln	GLN	NE2(HE21), NE2(HE22)	OE1
Met	MET	–	SD
Lys	LYS	NZ(HZ2), NZ(HZ3)	–
His	HID	ND1(HD1), <b>ND1(CE1)</b>	NE2
”	HIE	NE21(HE2), <b>NE2(CE1)</b>	ND1
”	HIP	NE2(HE2), ND2(HD1), <b>ND1(CE1)</b>	–
Tyr	TYR	OH(HH)	OH
Arg	ARG	NE(HE), NH1(HH11), NH1(HH12), NH2(HH21), NH2(HH22)	–
Chromophore	CRZ	N1(H1), OG1(HG1), OH(HH)	OH, OG1,O2,O3,N2
H <sub>2</sub> O	WAT	O(H1),O(H2)	O

Tableau III.2 : Définition des atomes donneur et accepteur de proton pour les chaînes latérales de résidu. La nomenclature adoptée spécifie le nom de l'atome ainsi que le nom du proton entre parenthèse. Les atomes en rouge sont dit spéciaux puisqu'ils établissent les LH intra-moléculaires, par conséquent l'atome entre parenthèses n'est pas un proton mais un atome qui mime la résonance d'un groupe chimique.

## B.2.d Recherche des réseaux de liaisons hydrogène

### B.2.d i Pavage tridimensionnel de la boîte de simulation

La boîte de simulation contient l'ensemble des atomes qui décrivent, d'une part, la protéine et, d'autre part, le solvant. Aussi, le nombre d'atomes contenus dans cette boîte est très important : il est, à titre d'exemple, proche de  $30 \times 10^3$  dans le cas d'une protéine globulaire d'environ 26 kDa. Dans ces conditions, il n'est pas raisonnable de tester les critères de liaison hydrogène sur l'ensemble des couples D/A que contient un instantané. Par conséquent, le test sur les LH doit s'effectuer sur un ensemble restreint de couples D/A afin de diminuer significativement le temps de calcul nécessaire à la cartographie de l'ensemble des couples D/A. Par ailleurs, le solvant composé de molécules d'eau renferme intrinsèquement une multitude de liaisons hydrogène qui n'ont pas d'intérêt dans le cadre de la recherche des réseaux de LH internes à la protéine. De plus, cette multitude de liaisons hydrogène dans le solvant risquerait de connecter

ensemble tous les réseaux de LH internes à la protéine et d'empêcher leur discrétisation. La recherche des LH doit donc être limitée aux atomes de la protéine et à son environnement proche correspondant à la première ou, à la rigueur, à la seconde couche d'hydratation.

La première difficulté consiste donc à identifier les premières couches d'hydratation de la protéine. Dans ce cadre, la méthodologie dite de détermination de l'aire superficielle accessible (ASA) permet de définir les contours d'une protéine à l'intérieur d'une boîte de simulation [Lee and Richards, 1971, Ausaf Ali et al., 2014]. Cependant, cette méthode est gourmande en temps de calculs.

Nous avons développé une nouvelle stratégie pour pallier cet inconvénient, mais également pour pouvoir distinguer les atomes participant aux réseaux de LH se trouvant à l'intérieur de la protéine de ceux qui se situent sur le bord ou encore pour pouvoir distinguer les molécules d'eau faisant partie de la première couche d'hydratation de la protéine de celles se trouvant dans la structure protéique. Cette stratégie repose sur le pavage tridimensionnel de la boîte de simulation. Celle-ci a été scindée en un nombre entier de sous-boîtes de telle sorte que la largeur de ces sous-boîtes soit au maximum égal au critère de distance maximum d'une liaison hydrogène auquel a été ajouté 1 Å, soit au maximum 4,2 Å. La boîte de simulation est ainsi subdivisée en 8000 sous-boîtes cubiques dont la largeur des côtés ne dépasse pas 3,6 Å ce qui limite le nombre d'atomes à 3 ou 4 par sous-boîte. Ceci réduit également le nombre de LH par sous-boîte. Ce pavage permet de distinguer 4 classes différentes de sous-boîtes définissant des zones distinctes de la protéine et de la boîte de simulation (figure III.3). Ainsi, afin de limiter la propagation des réseaux de LH dans le solvant, la recherche des LH sera limitée aux boîtes de la sphère de la protéine à savoir les sphères interne et de bord de la protéine.

- La **sphère interne de la protéine** (sous-boîtes 1). Les sous-boîtes de cette zone contiennent a minima un atome de la protéine et ne sont entourés que de sous-boîtes ne contenant, également a minima, qu'un atome de la protéine.
- La **sphère de bord de la protéine** (sous-boîtes 2) regroupe l'ensemble des sous-boîtes contenant a minima un atome de la protéine et qui sont adjacentes de sous-boîtes à la fois de la sphère interne et de la première sphère externe à la protéine.
- La **première sphère externe à la protéine** (sous-boîtes 3) correspond grossièrement à la première couche d'hydratation de la protéine. Les sous-boîtes qui la composent sont uniquement remplies de molécules de d'eau (solvant) et sont adjacentes aux sous-boîtes de la sphère de bord de la protéine.

- La **seconde sphère externe à la protéine** (sous-boîtes 4) est la zone la plus peuplée avec des sous-boîtes ne contenant que des molécules de d'eau (solvant), non adjacentes à des sous-boîtes de la sphère de bord de la protéine.

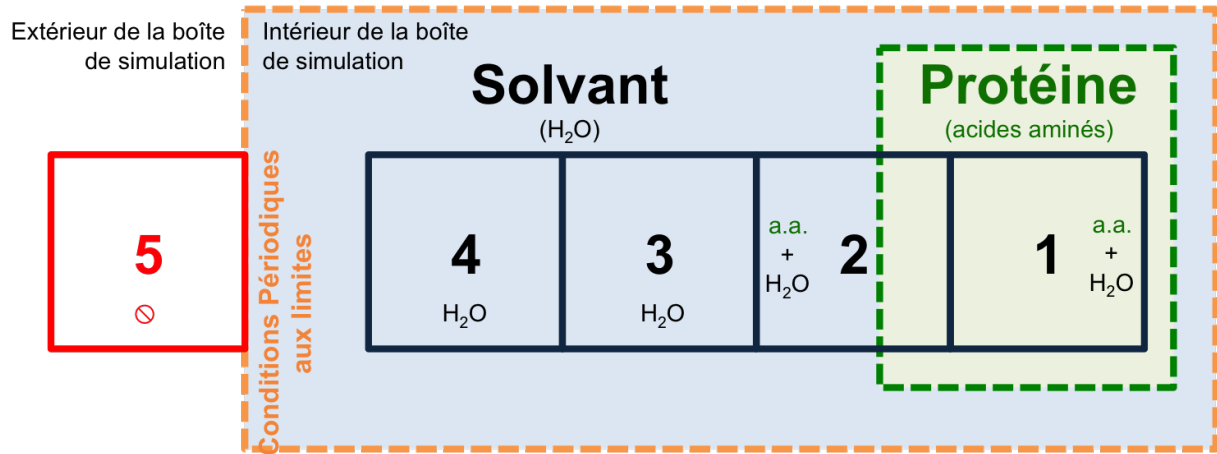


FIGURE III.3 : Représentation schématique de la typologie des sous boîtes qui composent le pavage tridimensionnel mis en place pour l’algorithme. Les atomes de la protéine sont présents dans les sous-boîtes de type 1 et 2. Les sous boîtes de type 3 et 4 sont uniquement constituées de molécules d’eau. La typologie des sous-boîtes est déterminée par leur composition et leur voisinage.

### B.2.d ii Recherche des liaisons hydrogène

La recherche des couples D/A formant des LH est réalisée de façon itérative le long des sous-boîtes et des trajectoires. La recherche des liaisons hydrogène s’effectue à partir de la première sous-boîte de plus petit numéro qui contient un atome ne faisant pas partie des sous-boîtes de la seconde sphère externe à la protéine puis s’étend aux sous-boîtes adjacentes. Le nombre des sous-boîtes adjacentes oscille entre 7 et 26, respectivement pour les sous-boîtes positionnées dans un angle et pour celles situées au cœur de la boîte de simulation. Une liaison hydrogène est identifiée si et seulement si les deux conditions suivantes sont simultanément vérifiées : (i) La liaison entre les atomes D et A est contenue dans une zone non exclue par le calcul à savoir elle est située dans les sphères internes et/ou de bord de la protéine et/ou dans la première sphère externe à la protéine. (ii) les critères géométriques de distance et d’angle sont respectés entre atomes D, H et A. Enfin, pour chaque instantané, l’algorithme vérifie que les liaisons hydrogène entre atomes lourds appartenant à des sous-boîtes adjacentes ne sont pas comptées en double. Les paramètres des liaisons hydrogène (numéro séquentiel des atomes D, H et A, distance  $d$ , angle  $\Omega$ ) et leur numéro séquentiel de détection sont mémorisés en fonction du numéro d’instantané.

### **B.2.d iii Agrégation des liaisons hydrogène en réseau de liaisons hydrogène**

L'agrégation des LH en réseau de LH est effectuée par instantané. La procédure d'agrégation consiste à chercher les LH qui partagent un atome lourd au sein de la liste de LH d'un instantané. L'agrégation des LH débute en partant de la première LH détectée pour le premier instantané. Si deux liaisons hydrogène partagent un même atome lourd, alors elles font partie du même réseau. On considère désormais que toutes liaisons hydrogène connexes aux deux premières LH de l'agrégat font partie du même agrégat. Cette opération est répétée jusqu'à épuisement des connexions entre liaisons hydrogène. Dans le cas où plus aucune LH n'est en connexion avec l'agrégat, alors la procédure d'agrégation reprend en partant d'une liaison hydrogène indépendante des réseaux de LH précédemment trouvés et ayant le plus petit numéro séquentiel. Afin de limiter la diffusion des réseaux de LH dans le solvant, le processus d'agrégation est stoppé à ce niveau dès que trois molécules d'eau consécutives sont en liaisons hydrogène.

### **B.2.d iv Réseau global de liaisons hydrogène connecté à un atome d'intérêt**

Le flux d'informations généré par l'algorithme est important. Par conséquent, l'algorithme offre la possibilité de rechercher les réseaux de liaisons hydrogène connectés à un atome d'intérêt. Un second fichier journal est alors généré regroupant uniquement ces réseaux. La recherche des réseaux de LH étant effectuée de telle sorte qu'il ne peut y avoir au plus qu'un réseau de LH par instantané connectés à un atome d'intérêt, le nombre maximum de réseaux que contient ce fichier ne dépasse pas le nombre total d'instantanés de la simulation.

### **B.2.d v Fichiers résultats**

Le flux d'informations généré par l'algorithme est important. Par conséquent, l'algorithme a été conçu de telle sorte qu'il est possible de restreindre l'analyse à la recherche des réseaux de liaisons hydrogène à un atome d'intérêt. Celui-ci constituera l'atome de départ de la recherche des réseaux. Le fichier journal est réduit pour ne contenir que les réseaux de liaisons hydrogène qui incluent l'atome de départ. Finalement, deux types de fichiers de sortie sont générés : des fichiers PDB contenant les coordonnées (x,y,z) des atomes qui composent les réseaux de liaisons hydrogène et des fichiers de données brutes contenant la topologie des réseaux de liaisons hydrogène. Ce fichier contient pour chaque liaison hydrogène, les numéros séquentiels des atomes qui la composent, la valeur de la distance  $d$  et de l'angle  $\Omega$  ( $\widehat{DHA}$ ) de cette liaison hydrogène ainsi que les numéros d'instantané pour lesquelles elle est détectée.

## B.3 Représentations graphiques des réseaux de liaisons hydrogène

### B.3.a Représentation sagittale (2D)

Le réseau global de LH connecté à un atome d'intérêt est schématisé par une matrice d'adjacence permettant la représentation sagittale de celui-ci. Il s'agit d'une représentation graphique de l'ensemble des LH constituant ce réseau mettant en évidence les différents atomes donneurs D et accepteurs A de protons et leur interconnexion par liaison hydrogène. Ces graphes sont orientés, à savoir qu'ils indiquent, par le biais de flèches, la direction des LH (figure III.5). Pour réduire la complexité de tels graphes, il est possible de ne représenter que les LH dont la densité de probabilité,  $dP_{HB_i}$ , est supérieure à une valeur seuil,  $T$ , définie par l'utilisateur. Ceci permet en particulier de repérer les atomes prépondérants dans le réseau ainsi que les liaisons hydrogènes les plus probables. Cette dernière information est également donnée par l'épaisseur de chaque flèche, proportionnelle à la densité de probabilité d'existence de la liaison hydrogène considérée au cours de la dynamique moléculaire. La valeur de cette densité est, en supplément, annotée à côté de la flèche, ainsi que le nom du proton partagé et la distance moyenne de la liaison hydrogène au cours de la dynamique moléculaire.

Enfin, les molécules d'eau peuvent être des acteurs importants dans les réseaux de LH comme c'est en particulier le cas dans le chemin de translocation sous-tendant l'ESPT de la AvGFP. La possibilité d'un échange avec d'autres molécules d'eau, cristallographiques ou provenant de la MD, est une information intéressante à connaître, susceptible de nous renseigner sur la stabilité de ces molécules dans la structure protéique et sur la perméabilité de la protéine vis-à-vis des molécules de solvant. Dans ce contexte, une première approche consiste à distinguer, sur ces graphiques, les molécules d'eau cristallographique (encadrées d'un trait bleu) de celles venant de la dynamique moléculaire (encadrées d'un trait violet).



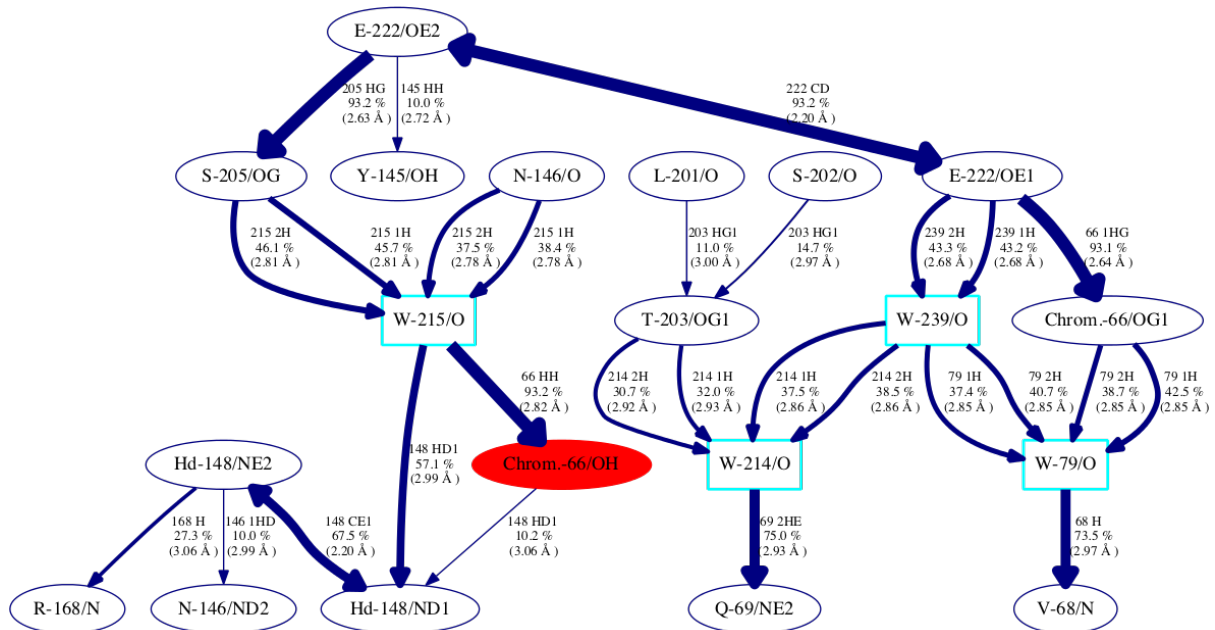


FIGURE III.4 : Représentation sagittale des réseaux de LH connectés à l'atome OH (donneur de proton) du phénol du chromophore dans le cas de la AvGFP. Ces réseaux ont été obtenus lors de l'analyse par l'algorithme d'une dynamique moléculaire de 30 ns (soit  $N = 15000$  instantanés) effectuée sur la structure protéique 1W7S. L'atome d'intérêt sélectionné, ici Chrom-66/OH, est mis en exergue (entouré d'un ovale rouge) grâce à la connaissance de son numéro séquentiel. Seules les LH avec une densité de probabilité supérieure à  $T = 10\%$  sont représentées.

### B.3.b Représentation 3D

Une représentation tridimensionnelle des réseaux globaux de LH est disponible. Cette représentation permet de les visualiser directement sur la structure de la protéine étudiée et d'apprécier visuellement leur étendue dans la structure, voire jusqu'au solvant (figure III.5). Les LH sont représentées par des traits (de couleur jaune) entre atomes dont l'épaisseur est, là encore, corrélée à la densité de probabilité d'existence de la liaison hydrogène. La structure choisie pour cette représentation est celle correspondant au dernier instantané de la dynamique moléculaire.

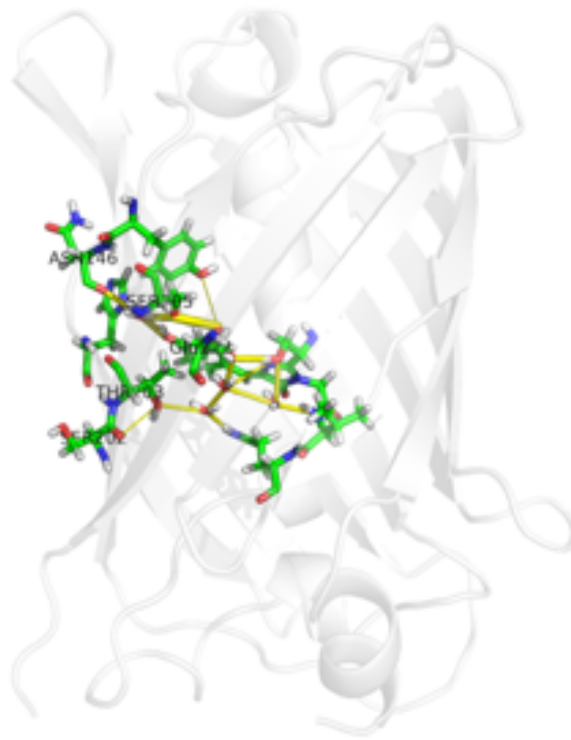
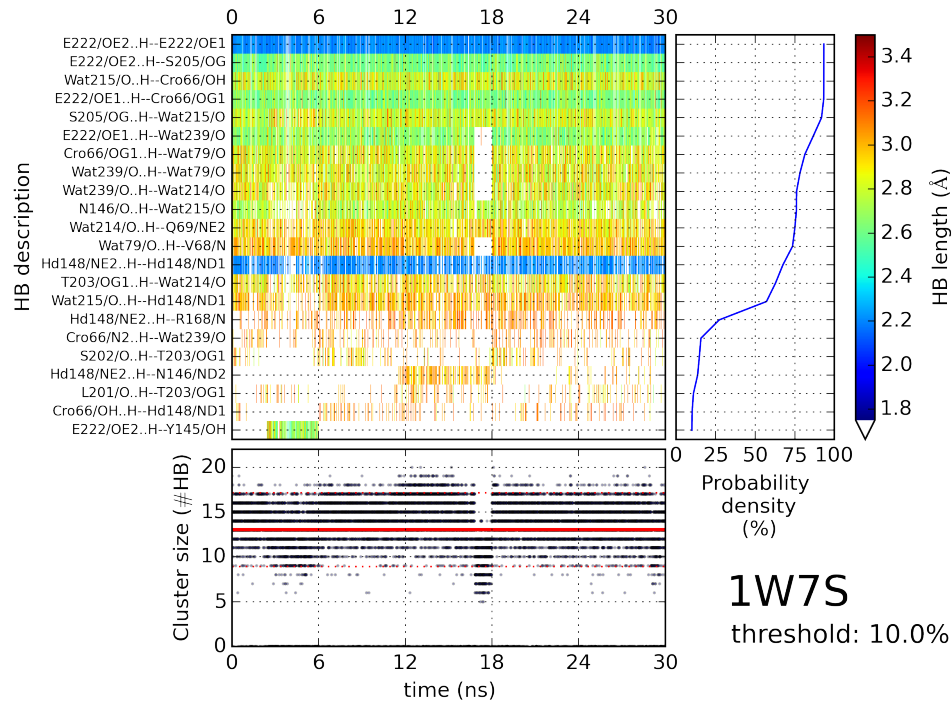


FIGURE III.5 : Représentation tridimensionnelle du réseau de LH connecté à l'atome OH (donneur de proton) du phénol du chromophore dans le cas de la AvGFP (structure 1W7S). Seuls les résidus dont au moins un atome est impliqué dans une LH sont représentés. Les LH sont représentées par des traits jaunes dont l'épaisseur est corrélée à leur densité de probabilité calculée sur une MD de 30 ns ( $N=15000$ ). Ici, seules les LH avec une densité de probabilité supérieure à  $T=10\%$  sont représentées.

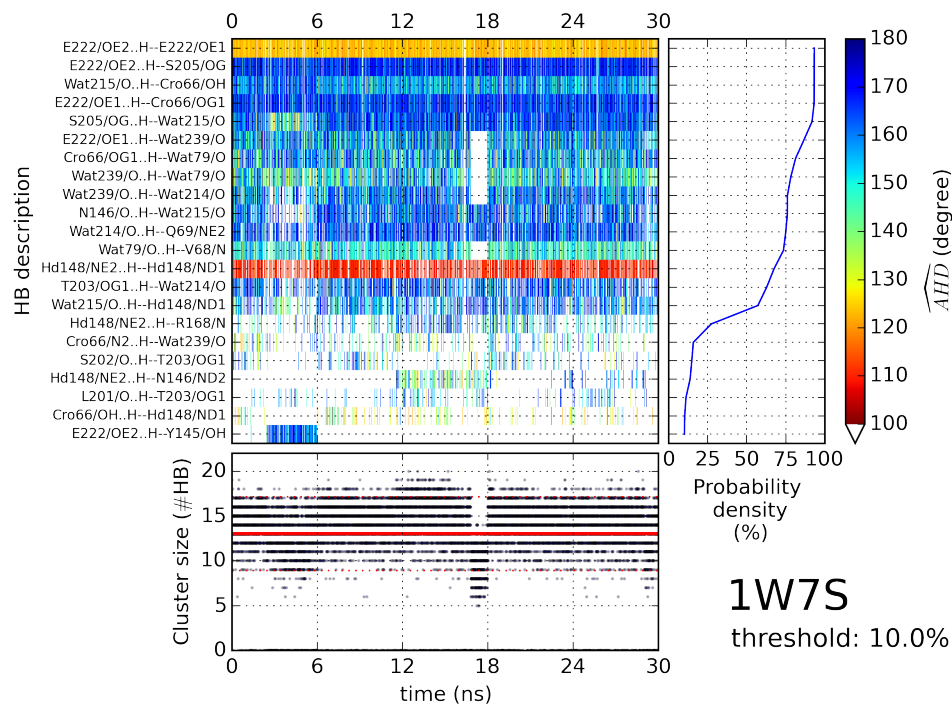
### B.3.c Représentation matricielle

La représentation simultanée de la persistance et de la concomitance des LH d'un réseau global, le long d'une dynamique moléculaire, ainsi que de la densité de probabilité respective de ses LH procure une vue d'ensemble de la stabilité des LH constituant ce réseau. Pour des réseaux de taille importante, comme pour les représentations précédentes, il est possible de se restreindre aux LH dont la densité de probabilité est supérieure à une valeur seuil  $T$  (indiqué en bas à droite de cette représentation). Plus précisément, dans le cadre central (figure III.6, l'existence de chaque liaison hydrogène du réseau global est spécifiée. Lorsqu'une liaison hydrogène apparaît à un instant  $t$  de la dynamique, un trait vertical est positionné pour un abscisse correspondant à  $t$ , avec une couleur corrélée à la distance  $d$  de cette liaison hydrogène ou à l'angle  $\Omega$ . Plus la couleur tend vers le bleu, plus la liaison hydrogène est forte (courte distance  $d$  et/ou grand angle  $\Omega$ ). La densité de probabilité de cette liaison hydrogène (proportion d'instantanés où elle vérifie les deux critères de distance et d'angle) est reportée dans le graphe de droite. La taille

du réseau global de LH, à savoir le nombre de LH concomitantes contenues dans ce réseau à un instant  $t$ , est reporté dans le graphique du bas. La valeur moyenne de la taille du réseau est signalée par un trait de couleur rouge.



(a)



(b)

FIGURE III.6 : Représentation matricielle du réseau global de liaisons hydrogène connecté à l'atome OH, donneur de proton, du phénol du chromophore de AvGFP, en fonction de (a) la distance  $d$ , (b) de l'angle  $\Omega$ . Les noms des atomes entre lesquels les LH s'établissent sont spécifiés sur l'axe des ordonnées du graphe central. Ces LH sont représentées par ordre décroissant de leur densité de probabilité. Les résultats sont issus de l'analyse d'une dynamique moléculaire d'une durée de 30 ns effectuée à partir de la structure protéique 1W7S ( $N = 15000$  instantanés,  $T = 10\%$ ). En moyenne, le réseau global connecté à OH du phénol comporte 12 liaisons hydrogène par instantané. De plus, la liaison hydrogène la plus probable présente une densité de probabilité de plus de 90% dans cette protéine.

Il est également possible d'effectuer une représentation similaire en considérant un chemin de translocation de proton, sous-partie du réseau global. Dans ces conditions, en supplément des paramètres précédemment exposés, la densité de probabilité de ce chemin est égale à la plus basse densité de probabilité de liaison hydrogène et la variation en fonction du temps du nombre total de LH concomitantes partant du donneur de proton (marqueur noire, courbe du bas).

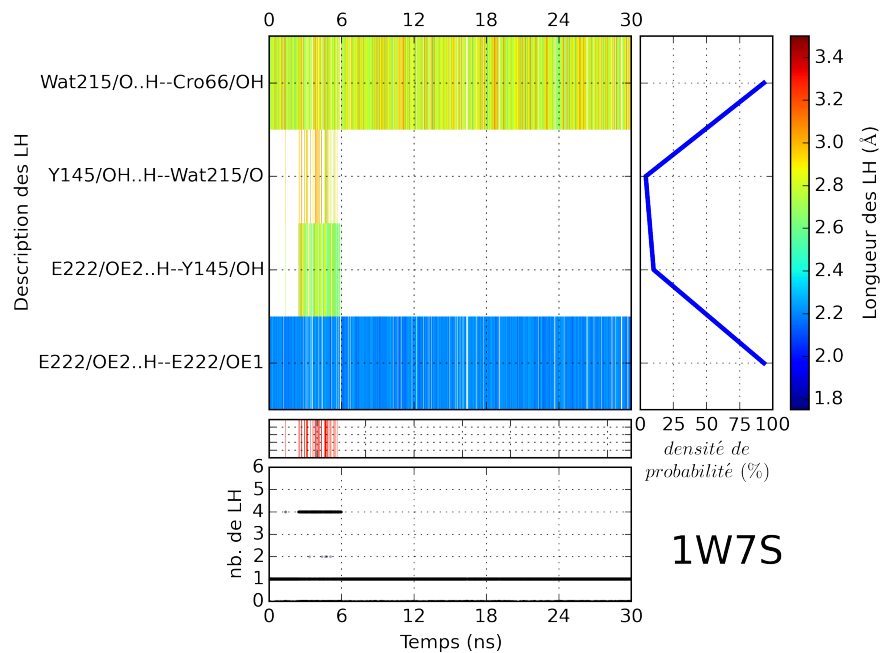


FIGURE III.7 : Représentation matricielle du chemin de translocation de proton partant de l'OH du phénol du chromophore de AvGFP et arrivant à l'atome d'oxygène OE2, de l'accepteur final Glu222. Ce chemin fait intervenir 3 LH, représentées dans le sens de la translocation, de  $D_i$  à  $A_f$ , impliquant le groupe OH du chromophore, l'oxygène de la molécule d'eau 215, l'hydroxyle de Tyr145 et enfin OE2 du Glu222. Il s'agit d'un chemin de translocation de proton alternatif au réseau usuel peu probable. Structure protéique : 1W7S ( $N = 15000$  instantanés,  $T = 10\%$ ).

## C Analyse des dynamiques moléculaires

Au cours de ces travaux, afin de valider l'algorithme, nous avons recherché et étudié les réseaux de LH présents dans AvGFP et certains de ses mutants dont la vitesse d'ESPT est modulée. Cela requière de mettre en œuvre des simulations de dynamique moléculaire pour l'ensemble de ces protéines. Ceci suppose dans un premier temps d'effectuer une sélection au niveau des structures cristallographiques disponibles pour ces protéines et de choisir une configuration de départ probable, en accord avec les conditions de pH de la structure cristallographique et de l'ESPT. Une fois la dynamique effectuée, il est nécessaire d'évaluer l'amplitude des déformations subies par la protéine au cours de la dynamique moléculaire, afin d'identifier des trajectoires non physiques reflétant un dépliement partiel de la structure. Ces deux étapes sont détaillées ci-après.

Nous avons étudiées quatre dynamiques moléculaires de 8 ns des quatre mutant de AvGFP étudiés, partant de points de départ distincts. Les deux premières nanosecondes ont été considérées comme faisant partie de la phase d'équilibration et ont été ôtées de l'analyse. Sur la base de l'hypothèse d'ergodicité, le temps de simulation de chacune des trajectoires est découpé en 5 tranches de 6 ns.

### C.1 Sélection des structures cristallographiques

#### C.1.a Structures cristallographiques

Nous souhaitons rechercher les réseaux de LH internes à la protéine. Aussi le système à étudier en dynamique moléculaire doit être une protéine monomérique, solvatée dans une boîte de simulation. Cependant, parmi les mutants étudiés, certains sont cristallisés sous forme de dimère ou de tétramère. De fait, la présentation des structures cristallines de chaque mutant s'impose permettant d'évaluer leurs paramètres cristallographiques et de justifier le choix du monomère sélectionné pour réaliser les simulations par dynamique moléculaire. Les critères de choix utilisés sont :

- un facteur de température (B) qui doit être le plus faible garant d'une structure peu agitée
- un nombre d'atomes modélisés le plus proche possible de celui défini par la séquence protéique
- un nombre de molécules d'eau ayant des conformations alternées le plus faible.

Les facteurs de température, le nombre de résidus et d'atomes de la structure et le nombre de molécules d'eau ayant des conformations alternées obtenus pour les 5 structures étudiés sont rassemblés dans le tableau III.3.

***AvGFP (1W7S)*** : Nous avons sélectionné la structure cristallographique 1W7S obtenue à pH 7,8 avec une résolution de 1,85 Å [[van Thor et al., 2006](#)]. Cette structure modélise 97% de la séquence protéique entre les résidus 2 à 231. Aucun résidu n'est déclaré aberrant selon les critères des graphiques de Ramachandran [[Ramachandran et al., 1963](#)]. Notre choix, sur la base des critères exposés ci-dessus, s'est porté sur le monomère A.

***AvGFP-S205V (2QLE)*** : La structure cristallographique de ce mutant, 2QLE, a été déposée avec une résolution de 1,59 Å [[Shu et al., 2007](#)]. Elle présente quatre monomères dans l'unité asymétrique et le pH de cristallisation est de 8. Le choix du monomère utilisé pour les simulations de dynamique moléculaire s'est porté sur le monomère A au regard du nombre d'atomes modélisés et de l'inexistence de conformations alternées pour les molécules d'eau malgré un facteur de température très légèrement supérieur à celui du monomère B.

***AvGFP-S65T (1C4F)*** : De nombreuses structures cristallographiques sont proposées pour ce mutant. Nous avons porté notre attention sur la structure 1C4F, monomérique, de résolution 2,25 Å, obtenue à pH acide (4,6) [[Elslinger et al., 1999](#)] et pour laquelle nous pouvons supposer que le chromophore se trouve dans sa forme protonée.

***AvGFP-T203V-S205A (4OGS)*** : La structure dimérique de ce mutant, 4OGS, a été obtenue, à pH 7,8, à 2,21 Å de résolution et nous avons sélectionné le monomère A [[Wineman-Fisher et al., 2014](#)].

***AvGFP-T203V-S205V (4OGS-S205V)*** : Ce mutant, bien qu'étudié spectroscopiquement [[Wineman-Fisher et al., 2014](#)], n'a pas de structure cristallographique. Un mutant *in silico* a été généré sur la base de la structure cristallographique de *AvGFP-T203V-S205A* en réalisant la mutation du résidu Ala205 en valine. Cette mutation permet d'accommoder deux groupements méthyles sur le carbone  $C_{\beta}$  du résidu alanine. L'hy-

bridation du carbone  $C_{\beta}$  de la valine étant de type  $sp^3$ , l'orientation des groupements méthyles est à prendre en compte. Le choix de l'orientation a été guidé par la position du groupement méthyle par rapport à l'environnement et aux statistiques fournies par la librairie de rotamères du logiciel *COOT* [Lovell et al., 2000]. Dans le cas du résidu Val205, le rotamère t est proposé avec un angle  $\chi_1 = 175^\circ$  et une fréquence d'apparition de 73%. L'orientation des groupements méthyle du rotamère t de Val205 est opposée à Tyr145 ce qui l'oriente en direction de Glu222 (figure III.8).

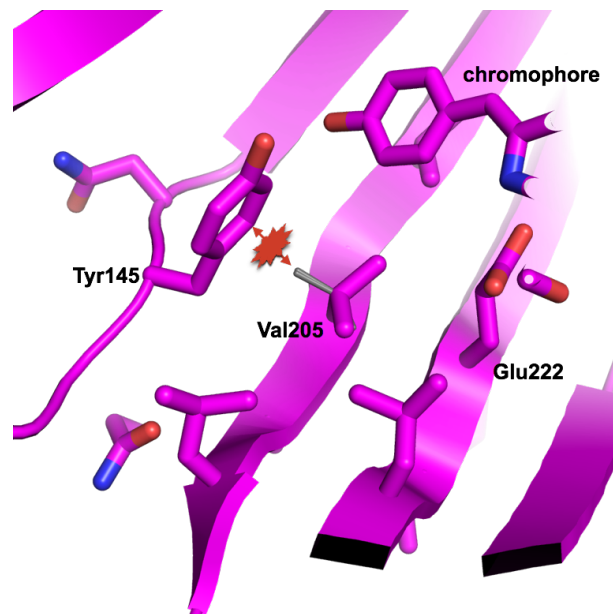


FIGURE III.8 : Rotamère t de la valine choisi pour la mutation *in silico* de la position 205 à partir de la structure cristallographique de 4OGS pour générer le mutant AvGFP-T203V-S205V.

Protéine (code PDB)	Chaîne	Facteur B moyen ( $\text{\AA}^2$ )	# Résidus (atomes)	# molécules H <sub>2</sub> O (conf. alternée)
AvGFP (1W7S)	A	16,2	2 à 231 (1830)	254
	B	18,9	3 à 231 (1824)	190
	C	16,9	2 à 231 (1830)	231
	D	18,4	2 à 231 (1830)	234
AvGFP-S205V (2QLE)	A	30,0	4 à 232 (1818)	238
	B	28,5	4 à 231 (1756)	240 (3)
	C	32,3	4 à 231 (1774)	213
	D	32,4	4 à 231 (1729)	173 (2)
AvGFP-S65T (1C4F)	A	43,6	3 à 229 (1777)	64
AvGFP-T203V-S205A (4OGS)	A	56,0	3 à 230 (1818)	89
	B	64,2	4 à 231 (1756)	66

Tableau III.3 : Principales caractéristiques des structures cristallographiques étudiées : facteurs de température moyens par chaîne (B), nombre de résidus et d'atomes et nombre de molécules d'eau pour chaque chaîne.



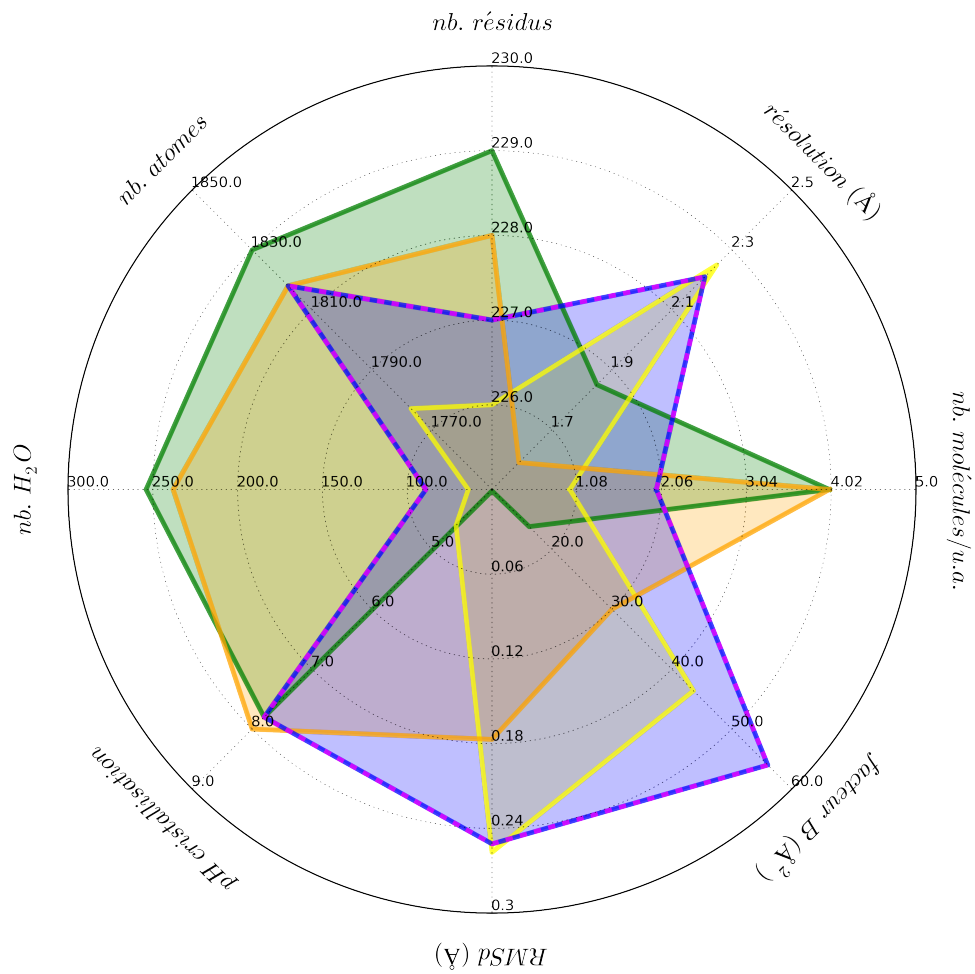


FIGURE III.9 : Différences structurales entre les structures choisies pour l'étude des réseaux de liaisons hydrogène AvGFP(vert), AvGFP-S205V (orange), AvGFP-S65T (jaune), AvGFP-T203V-S205V (bleu), AvGFP-T203V-S205A (mauve). Les valeurs de RMSD ont été calculées sur la base d'un alignement tridimensionnel avec la structure de 1W7S et des atomes de la chaîne principale (CA,C,N,O) et le terme u.a. signifie unité asymétrique.

### C.1.b Différences structurales majeures au sein des protéines étudiées

Le nombre de molécules d'eau observées dans les structures cristallographiques choisies oscille entre 64 et 254 soit une différence de 190 molécules d'eau. La gamme de résolution des protéines étudiées qui s'étalent entre 0,9 et 2,25 Å est responsable de la différence du nombre de molécules d'eau. En cristallographie, le nombre de molécules d'eau modélisées est inversement proportionnel à la résolution de la structure [Carugo and Bordo, 1999]. A haute résolution, les informations de diffraction aux rayons X contenues dans les réflexions permettent de modéliser plus de molécules d'eau (atomes d'oxygène) sans pour autant altérer la qualité du modèle construit principalement dans la couche d'hydratation des protéines. Les molécules d'eau définies avec précision dans les structures sont celles qui ont les facteurs de température B, les plus bas. En conséquence pour les structures dont la résolution est supérieure à 1,2 Å, nous avons cherché à identifier dans une coquille de 8 Å autour du chromophore la présence d'un volume susceptible d'accueillir une molécule d'eau n'ont résolue par cristallographie en utilisant le logiciel *CAVER* disponible sous *PyMOL* [Pavelka et al., 2016]. Dans notre cas, les protéines étudiées ne peuvent pas contenir plus de molécules d'eau dans l'environnement proche du chromophore qui nous intéresse pour la recherche des chemins de translocation partant de Chrom.66/OH.

Par ailleurs, sur l'ensemble des mutants, une différence notable au niveau structural se situe sur le résidu 65 participant à la structure du chromophore. Hormis dans le mutant AvGFP-S65T (1C4F) où une thréonine occupe cette position, toutes les autres protéines étudiées portent une sérine. Comme précédemment rapporté pour la AvGFP ou encore d'autres dérivés cyan de la AvGFP [Brejc et al., 1997, Fredj, 2012], l'orientation du groupement hydroxyle du résidu 65 est différente dans le cas de AvGFP-S65T par rapport aux autres structures. Dans ces dernières, il est orienté vers une couche de molécules d'eau présente à proximité du chromophore et est caractérisé par un angle dièdre de  $\chi_{65}(\text{OG1/C1/CA1/CB1}) = 65,2 \pm 6,95^\circ$ . Dans AvGFP-S65T, l'angle dièdre  $\chi_{65}$  est égale à  $-63,8^\circ$ . Le groupement hydroxyle dans ce mutant est orienté à l'opposé de la couche de molécules d'eau par rapport au chromophore. Dans cette configuration, le groupe méthyle de la thréonine se trouve logé dans une poche hydrophobe constituée majoritairement de leucines. Dans la suite, pour plus de simplicité et pour reprendre une terminologie proche de celle employée dans la littérature, nous nommerons ces conformations respectivement *65OH-down* et *65OH-up*.

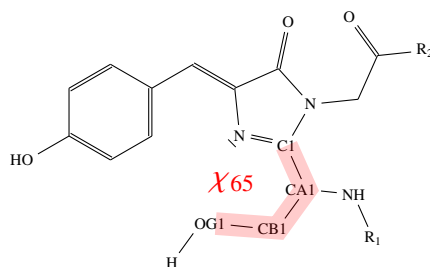


FIGURE III.10 : Angle dièdre de  $\chi_{65}(OG1/C1/CA1/CB1)$  permettant de décrire les conformations 65OH-down et 65OH-up

La position 203 est occupée par un acide aminé polaire (thréonine) ou apolaire (valine). Cette position se trouve dans l'environnement direct du chromophore. La distance moyenne entre le phénol du chromophore et l'hydroxyle de Thr203, pour les structures où elle est présente, est de  $3.74 \pm 1.03 \text{ \AA}$ . Dans ces structures, nous remarquons que l'hydroxyle de la thréonine présente deux orientations, dont une favorisant l'établissement d'une liaison hydrogène avec le phénol du chromophore ( $|\chi_{203}| < 60^\circ$ ). AvGFP-S205V (2QLE) présente une liaison hydrogène à la fin de la phase d'équilibration entre Thr203/OH et Chrom.66/OH, selon les critères de liaison hydrogène choisis pour cette étude.

Tout comme le résidu 203, la position 205 est occupée par des résidus polaires (sérine) ou apolaire (valine ou alanine). Cette position est également située dans l'environnement direct du chromophore puisque la distance moyenne entre le phénol du chromophore et l'hydroxyle de Ser205 est de  $\sim 4 \text{ \AA}$ . Dans toutes les structures portant une sérine, l'orientation de l'hydroxyle est invariable avec  $|\chi_{205}| \sim -170^\circ$ . D'après les structures, lors de l'introduction d'une valine, un des méthyles de celle-ci prend la place de l'hydroxyle de la sérine. Le remplacement en position 205 de la sérine par une valine, dont le volume de la chaîne latérale est plus important, ne conduit pas à une modification de l'orientation de la chaîne latérale. Enfin, en présence de la sérine à cette position, les structures suggèrent la présence d'une liaison hydrogène avec le résidu hautement conservé Glu222.

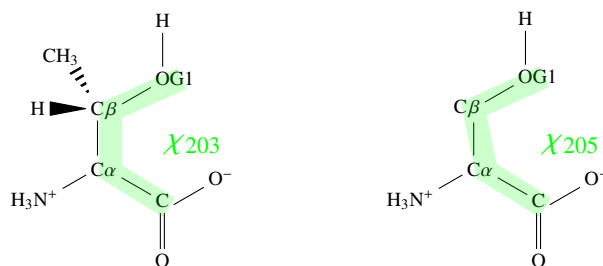


FIGURE III.11 : Angle dièdres  $\chi_{203}$  et  $\chi_{205}$  décrit par les atomes  $OG1/C\beta/C\alpha/C$  dans les résidus thréonine et sérine

## C.2 Choix des états de protonation et des conformations locales de départ

### C.2.a Etats de protonation

Le  $pK_{1/2}$  des 99 résidus protonables de AvGFP et de ses différents mutants étudiés ont été estimés par  $H^{++}$ . Bien que présentant un caractère acido-basique, il n'a pas été possible d'estimer le  $pK_{1/2}$  du chromophore, résidu non standard défini par un champ de force spécifique, lors des simulations par dynamique moléculaire.

Compte tenu des valeurs de  $pK_{1/2}$  obtenues et du pH de cristallisation des structures, toutes les tyrosines, lysines et arginines ont été protonées. A noter que la suite *AMBER* ne propose pas de patron pour les tyrosines et les arginines dans un état déprotoné (basique). A l'inverse, tous les acides aspartiques sont sous forme basique. Il y a deux cystéines dans la structure des GFP. Pour les deux, la forme neutre (protonée) a été choisie (CYS). Concernant les acides glutamiques, leur  $pK_{1/2}$  déterminé par le serveur  $H^{++}$  diffère fortement suivant la localisation du résidu dans la protéine et d'une structure à une autre. Sa valeur peut en effet varier d'environ 4 à près de 9 (figure III.12). Les résultats obtenus avec  $H^{++}$  soulignent le comportement particulier du résidu Glu222, constitutif chez les protéines fluorescentes homologues de AvGFP. Son  $pK_{1/2}$  est estimé à plus de 7 et peut avoisiner 9 suivant le mutant considéré. Nous avons observé un comportement similaire de ce résidu avec des valeurs de  $pK_{1/2}$  similaires, en utilisant *PROPKA* [Dolinsky et al., 2004], un autre serveur permettant le calcul des états de protonation d'une protéine à un pH donné. Ces résultats suggèrent qu'au pH auquel les structures ont été obtenues, entre 7 et 8, Glu222 peut, a priori, exister sous sa forme acide (protoné) ou basique (anionique). Cherchant à étudier les chemins de translocation de protons susceptibles de se produire à l'état excité à partir du phénol du chromophore, nous avons choisi de conserver, dans nos simulations, le Glu222 sous sa forme basique, anionique en accord avec les résultats expérimentaux. En effet, il a été montré que le phénol du chromophore et Glu222 présentent un couplage anticoopératif très fort à savoir que la déprotonation d'un des résidus empêche la déprotonation de l'autre [Bizzarri et al., 2007].

Pour ce qui est des histidines, un travail d'orientation de celles-ci a été mené en accord avec la structure hautement résolue 2WUR et de sa carte de densité  $2F_o - F_c$ . Dans une seconde phase, l'état de protonation choisi a été celui permettant de favoriser les interactions par liaisons hydrogène des histidines avec les résidus environnants. Parmi les neuf histidines invariablement présentes dans les structures étudiées (voir séquence consensus, Table III.4), deux ont été

protonées sur l'azote  $N_\epsilon$  (HIE), quatre sur l'azote  $N_\delta$  (HID). Enfin, pour les autres histidines, leur état de protonation diffère suivant le mutant. Ces histidines sont situées loin du chromophore.

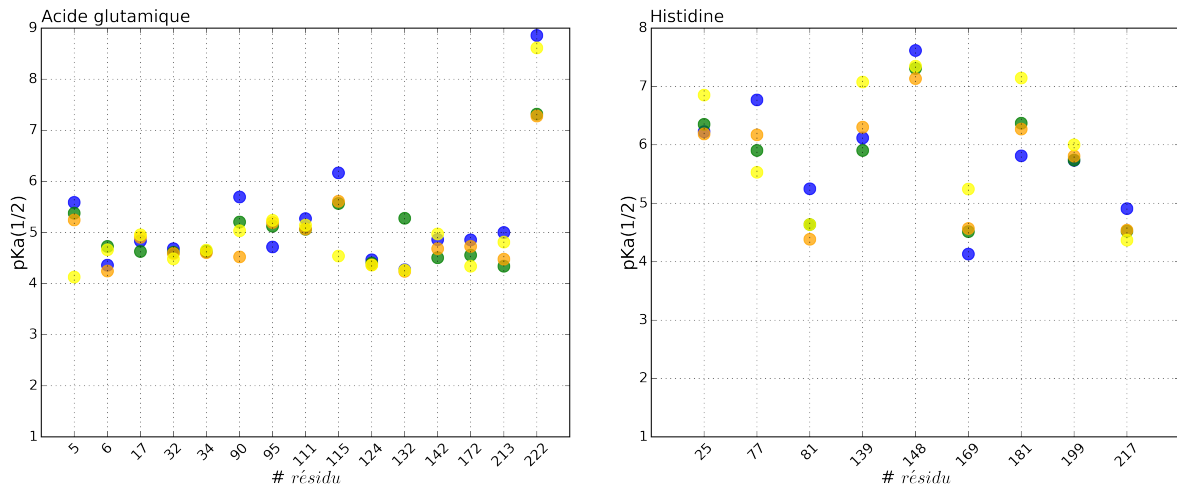


FIGURE III.12 : Valeurs du  $pK_{1/2}$  calculées par le logiciel  $H^{++}$  pour le groupe cible de protéines étudiées au cours de ces travaux avec AvGFP en vert, AvGFP-S65T en orange, AvGFP-S205V en jaune et AvGFP-T203V-S205A en bleu

1	Met	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu
16	Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	HIX	Lys	Phe	Ser	Val	Ser
31	Gly	Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys	Leu	Thr	Leu	Lys
46	Phe	Ile	Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp	Pro	Thr	Leu
61	Val	Thr	Thr	Phe	65	Tyr	Gly	Val	Gln	Cys	Phe	Ser	Arg	Tyr	Pro
76	Asp	HIE	Met	Lys	Arg	HIX	Asp	Phe	Phe	Lys	Ser	Ala	Met	Pro	Glu
91	Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly	Asn
106	Tyr	Lys	Thr	Arg	Ala	Glu	Val	Lys	Phe	Glu	Gly	Asp	Thr	Leu	Val
121	Asn	Arg	Ile	Glu	Leu	Lys	Gly	Ile	Asp	Phe	Lys	Glu	Asp	Gly	Asn
136	Ile	Leu	Gly	HID	Lys	Leu	Glu	Tyr	Asn	Tyr	Asn	Ser	HID	Asn	Val
151	Tyr	Ile	Met	Ala	Asp	Lys	Gln	Lys	Asn	Gly	Ile	Lys	Val	Asn	Phe
166	Lys	Ile	Arg	HIX	Asn	Ile	Glu	Asp	Gly	Ser	Val	Gln	Leu	Ala	Asp
181	HID	Tyr	Gln	Gln	Asn	Thr	Pro	Ile	Gly	Asp	Gly	Pro	Val	Leu	Leu
196	Pro	Asp	Asn	HID	Tyr	Leu	Ser	203	Gln	205	Ala	Leu	Ser	Lys	Asp
211	Pro	Asn	Glu	Lys	Arg	Asp	HIX	Met	Val	Leu	Leu	Glu	Phe	Val	Thr
226	Ala	Ala	Gly	Ile	Thr										

Tableau III.4 : Séquence consensus de AvGFP et ses mutants mettant en évidence les états de protonation des différentes histidines qui sont surlignées de jaune. Les sites dont les états de protonation sont différents dans la série de mutants de AvGFP étudiés sont désignés par le type HIX. Le chromophore et Glu222 et les positions mutées dans la série de mutants de AvGFP sont respectivement surlignés en vert, blue et rouge.

### C.2.b Choix des conformations locales

Une attention particulière a été portée sur les conformations locales de certains acides aminés en particulier ceux dont l'état de protonation est modulable. Une comparaison de toutes les structures cristallographiques de AvGFP et ses mutants a été effectuée en prenant pour référence la structure de AvGFP la plus résolue, 2WUR [Shinobu et al., 2010]. De plus, les conformations des histidines situées à l'intérieur de la protéine ont été vérifiées pour éviter des conformations locales non adaptés. En ce qui concerne, les conformations locales des résidus en position 65, 203 et 205, elles ont été conservées. Par ailleurs, l'orientation du proton de la fonction phénol du chromophore a été invariablement tournée en direction de la Ser205 soit pointant vers une molécule d'eau cristallographique omniprésente dans les structures de mutants de AvGFP.

## C.3 Stabilité de la structure protéique au cours des simulations

Deux grandeurs évaluent l'amplitude des déformations subies par la protéine au cours de la dynamique moléculaire. Il s'agit du RMSd qui permet de visualiser des mouvements structuraux de la protéine et le dépistage d'une perte de structuration secondaire de la protéine le long de la trajectoire. Le RMSf apporte une information plus locale et renseigne sur la mobilité de chaque résidu au cours de la simulation.

### C.3.a RMSd

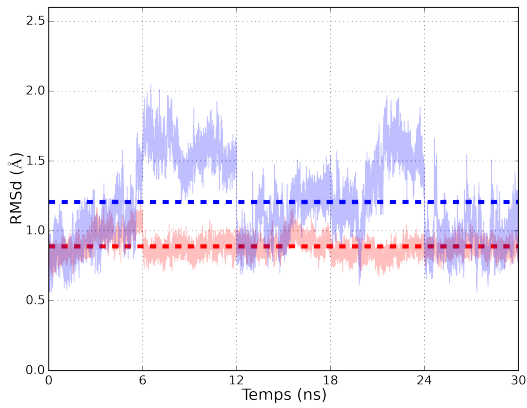
Le principe de cette mesure est de comparer les coordonnées des atomes du squelette (N, C $\alpha$ , C, O) de chaque résidu au cours des différents instantanés de la trajectoire avec celle de la structure de départ (structure cristalline). La différence des positions renseigne sur les mouvements subis par les structures secondaires.

La figure III.13 montre les variations des RMSd des différents mutants de AvGFP étudiées au cours de ces travaux. Pour l'ensemble de ces protéines, nous avons constatés une bonne stabilité des brins  $\beta$  avec une variation faible de leur RMSd au cours de la trajectoire suggérant que ces structures secondaires sont conservées au cours de la simulation (table III.5 et figure III.13). Pour autant, l'agitation des brins  $\beta$  est importante pour les mutants AvGFP-S65T et AvGFP-T203V-S205V. Dans le cas de AvGFP-S65T, les déplacements marqués du RMSd entre 12 et 18 ns ont pour origine des mouvements d'ampleur des brins  $\beta_9$  et  $\beta_{10}$ , bien que la valeur

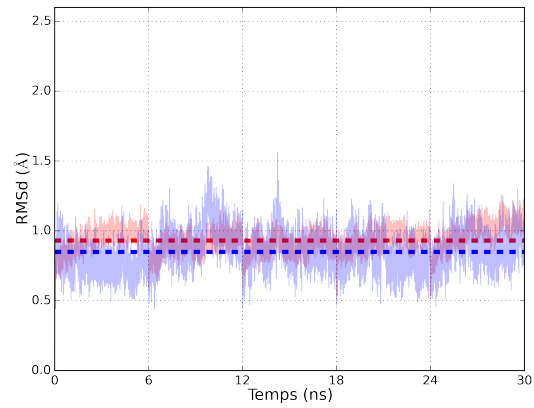
de leur RMSd reste inférieure à 1 Å. Dans AvGFP-T203V-S205V, on observe une déviation du RMSd moyen des brins  $\beta$  sur la fin de la dynamique à partir de 22 ns. Cette déviation semble due à l'agitation marquée des brins  $\beta_2$ ,  $\beta_4$  et  $\beta_5$  durant ce laps de temps. Cependant, le RMSd individuel de chacun des brins  $\beta$  cités n'excède pas 1 Å avec  $\text{RMSd}(\text{brin } \beta_2) = 0,66 \pm 0,09$  Å,  $\text{RMSd}(\text{brin } \beta_4) = 0,62 \pm 0,11$  Å et  $\text{RMSd}(\text{brin } \beta_5) = 0,93 \pm 0,16$  Å. Enfin, nos analyses montrent que le brin  $\beta_9$  est le brin le plus agité, même si les valeurs de RMSd reste modeste, de l'ordre de 1 Å voire inférieure. Par ailleurs, les hélices  $\alpha$  montrent majoritairement une agitation plus importante que les brins  $\beta$ . L'hélice  $\alpha_1$ , qui se situe à proximité du N-ter, sur l'extérieur de la protéine, contribue de façon significative à cette agitation. Sa perte de structure secondaire au cours de la trajectoire peut être à l'origine des variations repérées sur les hélices  $\alpha$ , comme c'est par exemple le cas de AvGFP-T203V-S205A (figure III.13.f). Enfin, nous avons porté une attention particulière à l'agitation des brins portant les résidus mutés dans les différents mutants étudiés. Il s'agit du brin  $\beta_{10}$  qui porte les résidus en position 203 et 205. De façon générale, ce brin est stable avec un RMSd inférieur à 1 Å. La position 65 et donc le chromophore sont portés par l'hélice  $\alpha_2$  centrale. Nous avons suivi son agitation dans le cas du mutant AvGFP-S65T. Cette hélice présente une bonne stabilité avec un RMSd de  $0,34 \pm 0,09$  Å.

RMSd (Å)	AvGFP	AvGFP-S205V	AvGFP-S65T	AvGFP-T203V-S205A	AvGFP-T203V-S205V
Hélices $\alpha$	1,21±0,30	0,85±0,14	1,43±0,36	1,05±0,38	1,15±0,27
Hélice $\alpha_1$	1,14±0,61	0,39±0,12	0,63±0,27	0,58±0,25	0,50±0,22
Brins $\beta$	0,89±0,07	0,93±0,09	1,34±0,23	0,95 ±0,09	1,39±0,18
Brin $\beta_9$	0,97±0,16	1,01±0,15	0,92±0,18	0,98±0,15	1,02±0,13
Brin $\beta_{10}$	0,63±0,10	0,56±0,11	0,71±0,18	0,76±0,13	0,65±0,12

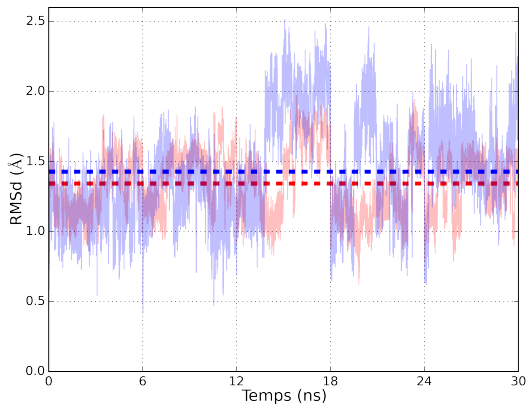
Tableau III.5 : RMSd des atomes du squelette (N, C $\alpha$ , C, O) des résidus impliqués dans les brins  $\beta$  ou les hélices  $\alpha$  pour les différentes protéines étudiées



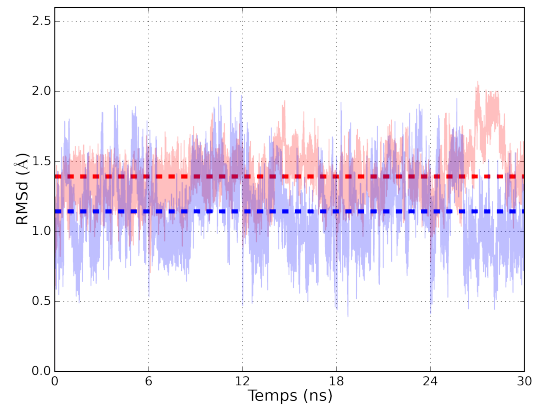
(a)



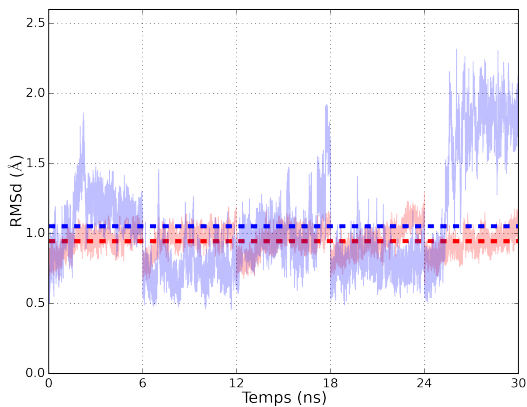
(b)



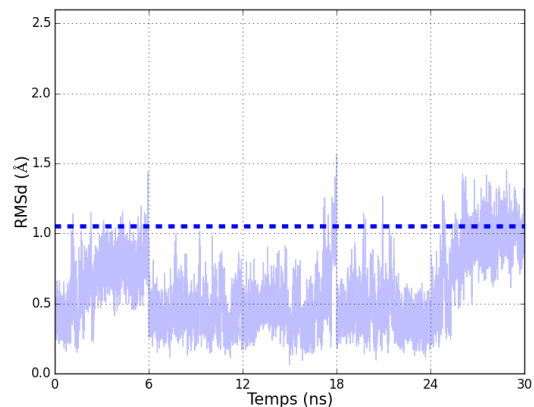
(c)



(d)



(e)



(f)

FIGURE III.13 : RMSd des structures secondaires feuillet  $\beta$  (en rouge) et hélices  $\alpha$  (en bleu) avec leur moyenne respective représentée en pointillés, (a) AvGFP, (b) AvGFP-S205V, (c) AvGFP-S65T, (d) AvGFP-T203V-S205V, (e) AvGFP-T203V-S205A, (f) contribution de l'hélice  $\alpha_1$  à la valeur du RMSd des hélices de AvGFP-T203V-S205A



### C.3.b RMSf

Le calcul de RMSf pour chacune des dynamiques permet d'accéder au facteur de température moyen pour chacun des résidus (atomes N, C $_{\alpha}$ , C, O) et ainsi de mesurer leur flexibilité [Medvedev et al., 2014]. La superposition des facteurs B moyens entre les dynamiques montrent la dynamique des structures secondaires qui composent les protéines.

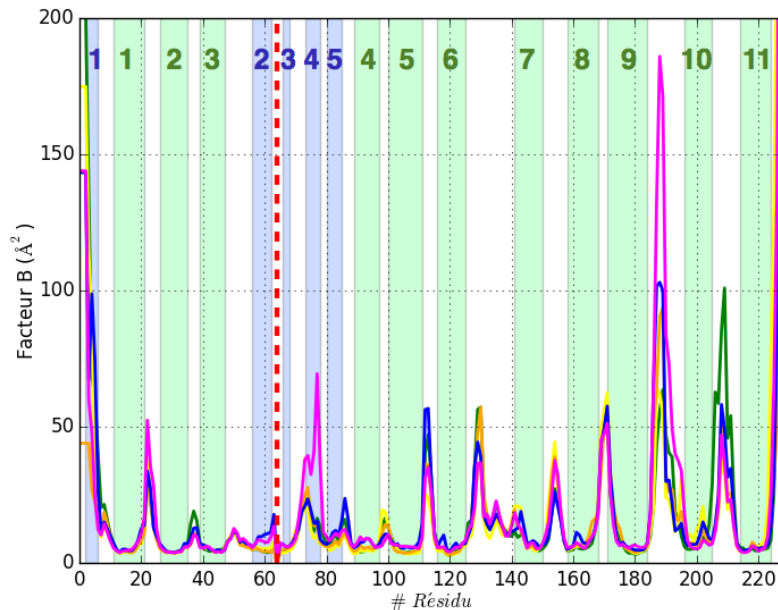


FIGURE III.14 : Facteur de température  $B$  en fonction du numéro de résidu au cours des dynamiques de AvGFP (trait vert), AvGFP-S65T (trait jaune), AvGFP-S205V (trait orange), AvGFP-T203V-S205A (trait bleu) et AvGFP-T203V-S205V (trait mauve). Les structures secondaires des mutants de AvGFP sont localisées avec un fond vert pour les brins  $\beta$  et un fond bleu pour les hélices  $\alpha$ . La position du chromophore est représentée par les pointillés rouges.

Sur la figure III.14, nous constatons que les résidus les plus agités se situent dans les zones blanches soit en dehors des structures secondaires en brins  $\beta$  et en hélices  $\alpha$  à l'exception de l'hélice  $\alpha_4$ . Comme attendu, les bords du graphique montrent une agitation importante des N et C terminaux avec des valeurs de facteur B particulièrement élevées. Les autres zones d'agitation (10 au total), caractérisées par des pics au niveau du facteur B, se situent entre les brins  $\beta$  et correspondent toutes à des coudes entre ces mêmes brins.

De façon générale, les coudes sont classifiés selon deux types, court avec des liaisons hydrogène (T) et long sans liaisons hydrogène (S) selon la nomenclature DSSP (algorithme *Define Secondary Structure of Proteins*) [Kabsch and Sander, 1983]. Dans les protéines fluorescentes dérivés de AvGFP, les structures secondaires de type brin  $\beta$  sont espacées par des coudes plus

ou moins longs. Cette différence de structuration influe le facteur de température. En effet, les coudes entre les brins  $\beta_{2/3}$  et  $\beta_{4/5}$ , de caractère T, sont caractérisés par une valeur de B inférieure à  $20 \text{ \AA}^2$ . En revanche, les coudes longs de caractère S ont tendance à avoir une valeur de facteur de température supérieure à  $60 \text{ \AA}^2$  comme le coude entre les brins  $\beta_{9/10}$ .

## C.4 Première analyse des dynamiques moléculaires : aspects structuraux

### C.4.a Phénol du chromophore

Le champ de force du chromophore comprend des contraintes permettant de limiter la rotation de la liaison OH du phénol par rapport au plan du chromophore. Nous avons vérifié, qu'au cours des dynamiques moléculaires, la liaison OH reste dans cette même configuration par l'intermédiaire du suivi de la trajectoire de l'angle dièdre  $\chi_{OH}(\text{HH}/\text{OH}/\text{CZ}/\text{CE1})$ . Ceci est le cas pour toutes les protéines étudiées hormis le mutant AvGFP-S65T. En effet, pour ce dernier, une rotation de la liaison de  $180^\circ$  a été observée conduisant à un réarrangement local des liaisons hydrogènes. Au cours de la dynamique de 30 ns, cet événement se produit deux fois et de façon rapprochée : cette conformation alternée reste stable sur une durée de 140 et 166 ps, soit 1% du temps de simulation, et est entrecoupée d'une phase avec une orientation normale de 218 ps.

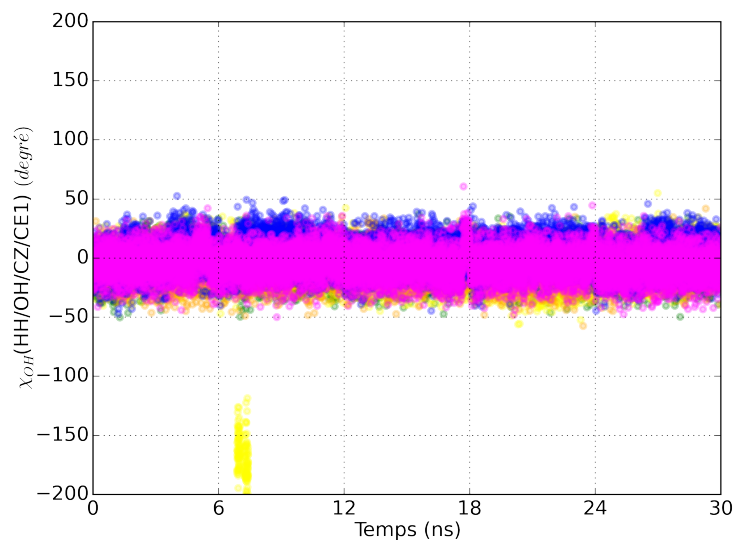


FIGURE III.15 : Trajectoire de l'angle dièdre  $\chi(\text{HH}/\text{OH}/\text{CZ}/\text{CE1})$  du phénol du chromophore au cours des dynamiques moléculaires de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve).

#### C.4.b Résidu 65

L'orientation dans la structure de départ de l'hydroxyle en position 65 a été celle observée dans les structures cristallographiques utilisées (voir paragraphe III.C.1.b). L'analyse des dynamiques moléculaires des différentes protéines étudiées fait apparaître trois conformations distinctes de cet hydroxyle. Outre les deux conformations précédemment explicitées (65OH-*down* ou 65OH-*up*), une troisième conformation est explorée au cours des dynamiques moléculaires de certains mutants (figure III.16 et III.17). Dans cette conformation, l'hydroxyle est inséré dans la poche hydrophobe de leucines, précédemment décrite (voir paragraphe III.C.1.b). Dans la suite, nous nommerons cette conformation 65OH-*opposée*.

Seule la dynamique moléculaire de AvGFP montre, tout au long du temps de simulation, une constance de la conformation de l'hydroxyle. Celui-ci adopte la conformation 65OH-*down*, orienté vers les molécules d'eau et le Glu222, avec lequel il développe une liaison hydrogène persistante (100%). L'introduction de mutations au niveau du chemin de translocation ou dans son environnement a une incidence sur la conformation de ce résidu, puisque dans les autres mutants étudiés, l'hydroxyle explore plusieurs conformations au cours de la dynamique. Dans le cas de AvGFP-S205V et AvGFP-S65T, une conformation est majoritairement présente avec un temps de présence de plus de 90%. Dans AvGFP-S205V, il s'agit de la conformation 65OH-*down*, conservée 99,0% du temps alors que dans AvGFP-S65T, la conformation majoritaire (93,4% du temps de simulation) reste la conformation observée sur la structure cristallographique, à savoir 65OH-*up*. Pour ces deux protéines, la conformation alternative est 65OH-*opposée*. Dans AvGFP-S205V, l'hydroxyle subit de façon très transitoire des rotations vers cette conformation qui lorsqu'elle est atteinte, est explorée moins de 1% du temps. Dans AvGFP-S65T, cette conformation est plus favorisée avec un temps de présence total de 6,6% soit ~2ns.

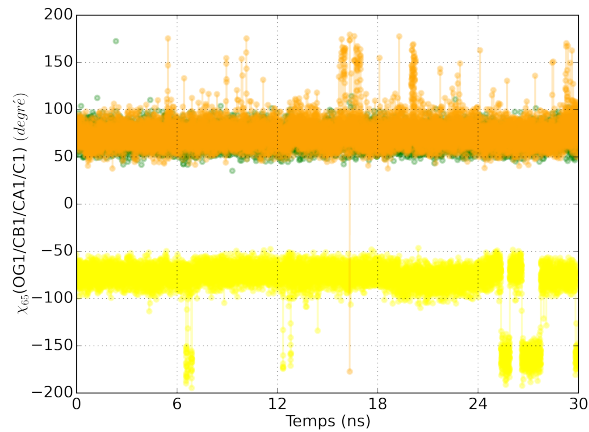


FIGURE III.16 : Trajectoires de l'angle dièdre  $\chi_{65}$  caractérisant la conformation de l'hydroxyle en position 65 dans AvGFP (vert), AvGFP-S65T (jaune) et AvGFP-S205V (orange). Les conformations 65OH-down, 65OH-up, et 65OH-opposée de l'hydroxyle sont respectivement caractérisées par  $\chi_{65} \sim +65^\circ$ ,  $\chi_{65} \sim -65^\circ$  et  $\chi_{65} \sim -180^\circ$  ou  $\chi_{65} \sim +180^\circ$ .

Les dérivés AvGFP-T203V-S205A et AvGFP-T203V-S205V montrent un comportement légèrement différent, puisque dans ces deux protéines, les trois conformations sont explorées au cours de la simulation. La conformation 65OH-down reste majoritaire avec un temps de présence total de 22,9 ns (76,2%) pour AvGFP-T203V-S205A et de 25,1 ns (83,8 %) pour AvGFP-T203V-S205V. La conformation alternative la plus visitée est la conformation 65OH-opposée avec un temps de présence cumulé de 5,7 ns pour AvGFP-T203V-205A et de 4,1 ns pour AvGFP-T203V-S205V. Enfin, de façon transitoire, l'hydroxyle explore également la conformation 65OH-up pendant une durée cumulée de moins de 0,8 ns.

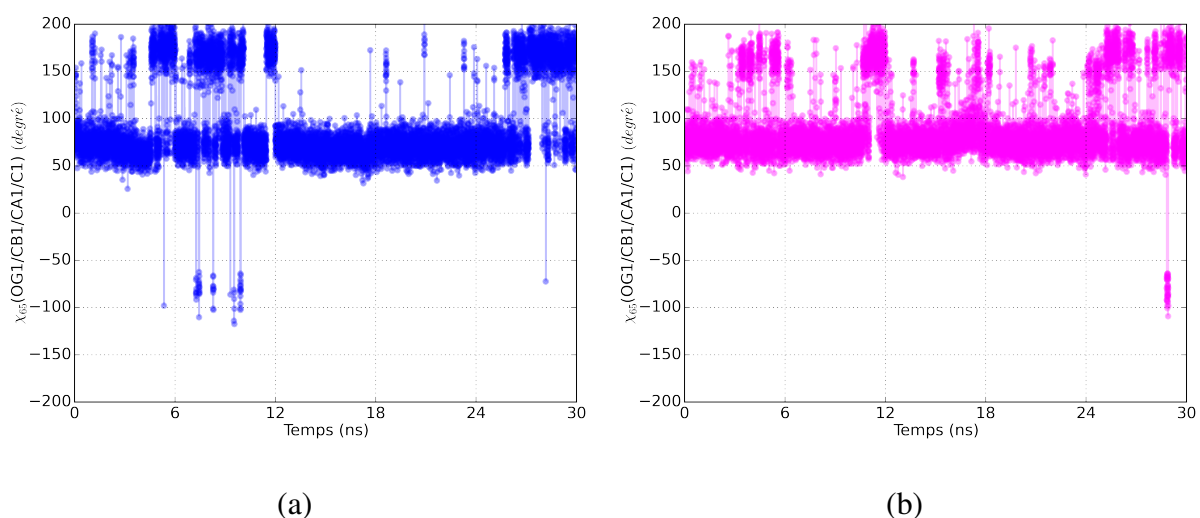


FIGURE III.17 : Trajectoires de l'angle dièdre  $\chi_{65}$  caractérisant la conformation de l'hydroxyle en position 65 pour les simulations de AvGFP-T203V-S205A (a) et AvGFP-T203V-S205V (b).

En résumé, en présence d'une thréonine, la conformation *65OH-up* est prépondérante et la conformation *65OH-down* n'est pas observée. Du point de vue structural, la conformation *65OH-up* correspond à une conformation où le méthyle, hydrophobe, de Thr65 est inséré dans une poche hydrophobe constituée de 3 leucines (42, 44 and 220), d'une valine (Val68) et d'une phénylalanine (Phe46) avec lesquels il est en interaction favorable. La conformation *65OH-down*, quant à elle, impliquerait une interaction défavorable entre le méthyle et le carbonyle hydrophile de la valine 61. Avec une thréonine en position 65, la conformation *65OH-up* apparait en effet la plus favorable.

En présence d'une sérine, cette conformation *65OH-up* est très peu explorée tout au long des simulations et la conformation prépondérante est dorénavant *65OH-down*. A partir de nos données, il est difficile de mettre en évidence un chemin préférentiel permettant de passer de *65OH-down* à *65OH-up* et inversement. Dans *AvGFP-T203V-S205A*, ce passage observé 8 fois implique systématiquement l'intermédiaire de la conformation *65OH-opposée*. Dans *AvGFP-T203V-S205V*, le changement de conformation de *65OH-down* à *65OH-up* et vice-versa s'effectue sans intermédiaire. Néanmoins, l'absence ou la très faible occurrence de la rotation de l'hydroxyle en position 65 de la conformation *65OH-up* à celle *65OH-down* et inversement, qu'une sérine ou une thréonine soit en position 65, suggère que cette rotation implique le passage d'une barrière de potentiel importante nécessitant la concomitance d'un ensemble de mouvements locaux. Lorsque cette condition est remplie, la rotation s'effectue sur une échelle de temps inférieure à 2 ps. Enfin, la conformation alternative la plus explorée au cours des simulations est *65OH-opposée*. Néanmoins, elle est explorée de façon transitoire ce qui peut s'expliquer par le passage du groupe hydroxyle, hydrophile, dans la poche hydrophobe de leucines. De plus, l'alternance entre cette conformation et l'une des deux autres conformations au cours des simulations suggère que, dans ces protéines, le réseau de liaisons hydrogène n'est plus suffisamment stable pour maintenir l'hydroxyle du résidu 65 dans sa conformation la plus stable thermodynamiquement.

### C.4.c Histidine 148

L'histidine 148 se situe en face du phénol du chromophore. Contrairement à l'azote NE2, l'azote ND1 est dans son état protoné dans toutes les protéines étudiées et est susceptible d'interagir par LH en tant que donneur de proton avec un doublet non liant du phénol du chromophore. Le suivi de la torsion de l'angle dièdre  $\chi_{148}(\text{CA/CB/CG/ND1})$  permet d'évaluer les mouvements de la chaîne latérale de l'histidine 148 au cours de la dynamique. La figure III.18 suggère que la chaîne latérale de l'histidine 148 ne subit pas de retournement ( $\pm 180^\circ$ ) pour l'ensemble des protéines qui possèdent une thréonine en position 203. En présence d'une valine à cette position, un retournement épisodique s'opère. Deux événements d'une durée de 50 ps et 586 ps soit 2,10% du temps de la dynamique ont été observés dans le cas du mutant AvGFP-T203V-S205V. Dans AvGFP-T203V-S205A, ce retournement n'est pas effectif.

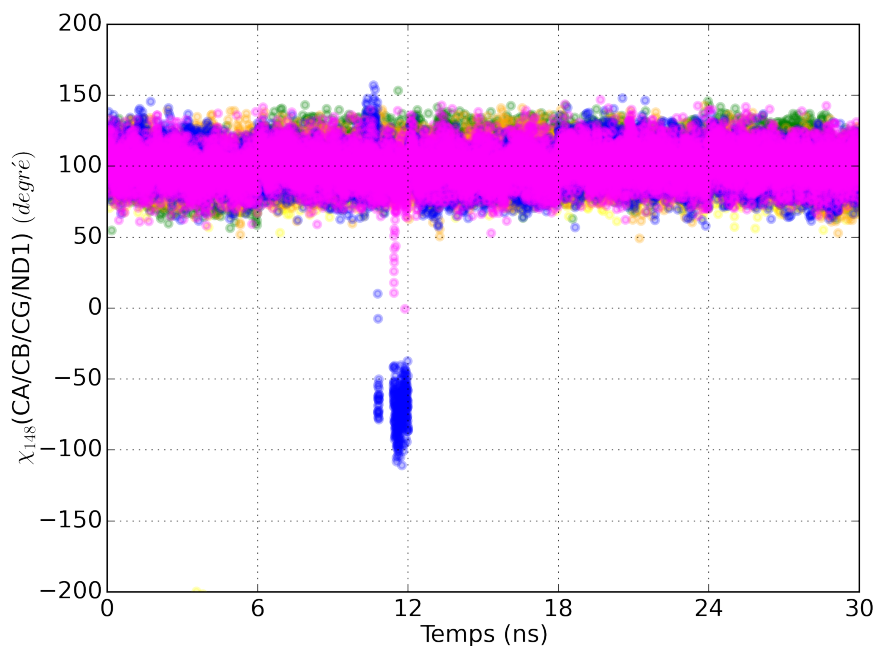


FIGURE III.18 : Trajectoires de l'angle dièdre  $\chi_{148}$  caractérisant la conformation de la chaîne latérale de l'histidine 148 au cours des dynamiques de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-S205A (bleu) et AvGFP-T203V-S205V (mauve).

#### C.4.d Résidu 203

Les résultats rapportés dans la littérature indiquent que la thréonine en position 203 joue un rôle crucial dans l'établissement d'un chemin de translocation alternatif lorsque le chemin principal passant par la sérine 205 est rompu. L'introduction de résidus non polaires en positions 205 et 203 conduit à une réduction drastique du transfert de proton à l'état excité voire à sa suppression. A ce titre, les mouvements du groupement hydroxyle de la thréonine ou encore d'un des deux méthyles de la valine peuvent être visualisée grâce au suivi des trajectoires des angles dièdres, respectivement  $\chi_{V203}(N/CA/CB/OG1)$  ou  $\chi_{T203}(N/CA/CB/CG1)$ .

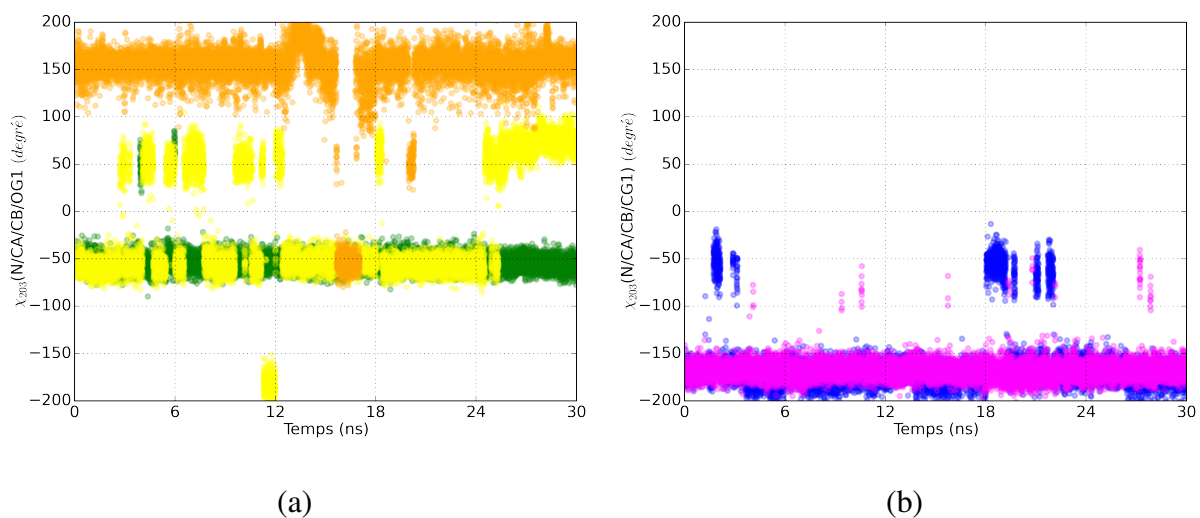


FIGURE III.19 : Trajectoires de l'angle dièdre  $\chi_{T203}$  (a) pour AvGFP (vert), AvGFP-S65T (jaune) et AvGFP-S205V (orange) ainsi que  $\chi_{V203}$  (b) pour AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve) en fonction du temps des dynamique.

L'espace conformationnel est plus largement visité avec une thréonine en position 203 à l'exception de la dynamique de AvGFP qui démontre une stabilité notable. Cette différence de comportement est à mettre en parallèle avec les contraintes stériques des acides aminés pour cette position, une valine est plus volumineuse qu'une thréonine avec un groupement méthyle supplémentaire. De plus, la fonction polaire hydroxyle de la thréonine est remplacée par un groupement méthyle, apolaire, ce qui réduit les possibilités de liaisons hydrogène.

### C.4.e Résidu 205

La position 205 est une des positions clés de l'ESPT dans AvGFP et ses dérivés. La rotation de l'angle dièdre  $\chi_{205}(\text{N/CA/CB/OG})$  renseigne sur la rotation de la chaîne latérale de ce résidu. Le résidu 205 est dans AvGFP une sérine et une valine pour les mutants AvGFP-S205V, AvGFP-T203V-S205V.

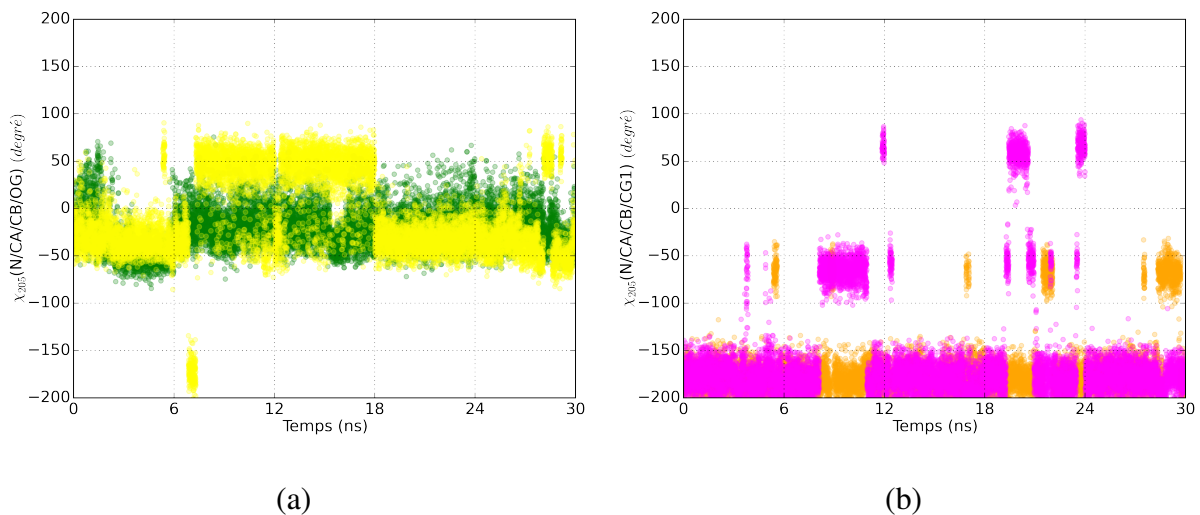


FIGURE III.20 : Trajectoires de l'angle dièdre  $\chi_{205}$  en fonction du temps et du résidu en position 205, (a) sérine avec les mutants AvGFP (vert) et AvGFP-S65T (jaune) et (b) valine avec AvGFP-S205V (orange) et AvGFP-T203V-S205V (mauve).

On observe que la rotation de  $\chi_{205}$  est effective quel que soit le résidu en position 205. L'encombrement stérique additionnel du résidu valine induit par les deux groupements méthyle contre un groupe hydroxyle dans le cas d'une sérine ne modifie la dynamique de ce résidu.



#### C.4.f Résidu 222

Le Glu222 est l'accepteur final de l'ESPT dans AvGFP et ses mutants. Dans toutes les dynamiques moléculaires faites au cours de ces travaux, nous l'avons choisi déprotonné. Compte tenu du caractère symétrique du carboxylate, l'échange entre OE1 et OE2 a, a priori, peu d'incidence sur la recherche des réseaux de liaisons hydrogène. Néanmoins, cet échange et sa fréquence donne une première idée de la force des liaisons hydrogène dans lesquels l'un et/ou l'autre de ces atomes sont impliqués. Pour cela, nous avons porté une attention particulière à l'angle dièdre  $\chi_{222}(\text{OE1/CD/CG/CB})$ , permettant le suivi de la rotation du carbonyle autour de la liaison C-C portant le groupe carboxylate. Dans toutes les dynamiques, cette inversion est observée. C'est un évènement rapide (échelle de temps inférieure à 2 ps) dont la fréquence dépend des mutants. A titre d'exemple, l'inversion du carbonyle dans AvGFP survient dix fois durant les 30 ns de la simulation et la durée des évènements oscillent entre 98 ps et 1,18 ns. Le mutant AvGFP-S65T montre une stabilité plus importante de ce résidu. En revanche, pour les protéines AvGFP-S205V, AvGFP-T203V-S205A et AvGFP-T203V-S205V, les dynamiques font apparaître des inversions très fréquentes. Ceci suggère un faible maintien du groupe par liaisons hydrogène.

L'inversion des atomes OE1 et OE2 nuit néanmoins à la recherche des liaisons hydrogène par notre algorithme puisque cette recherche s'effectue en considérant les numéros séquentiels des atomes. OE1 et OE2 ont des numéros séquentiels différents et même si tous les autres atomes du réseau sont identiques, l'algorithme ne considérera pas ces deux réseaux comme identiques. Ceci a motivé l'introduction d'une liaison hydrogène, permanente, interne au carboxylate entre OE1 et OE2 afin de permettre de propager la recherche du réseau de liaisons hydrogène automatiquement à l'autre atome du carboxyle.

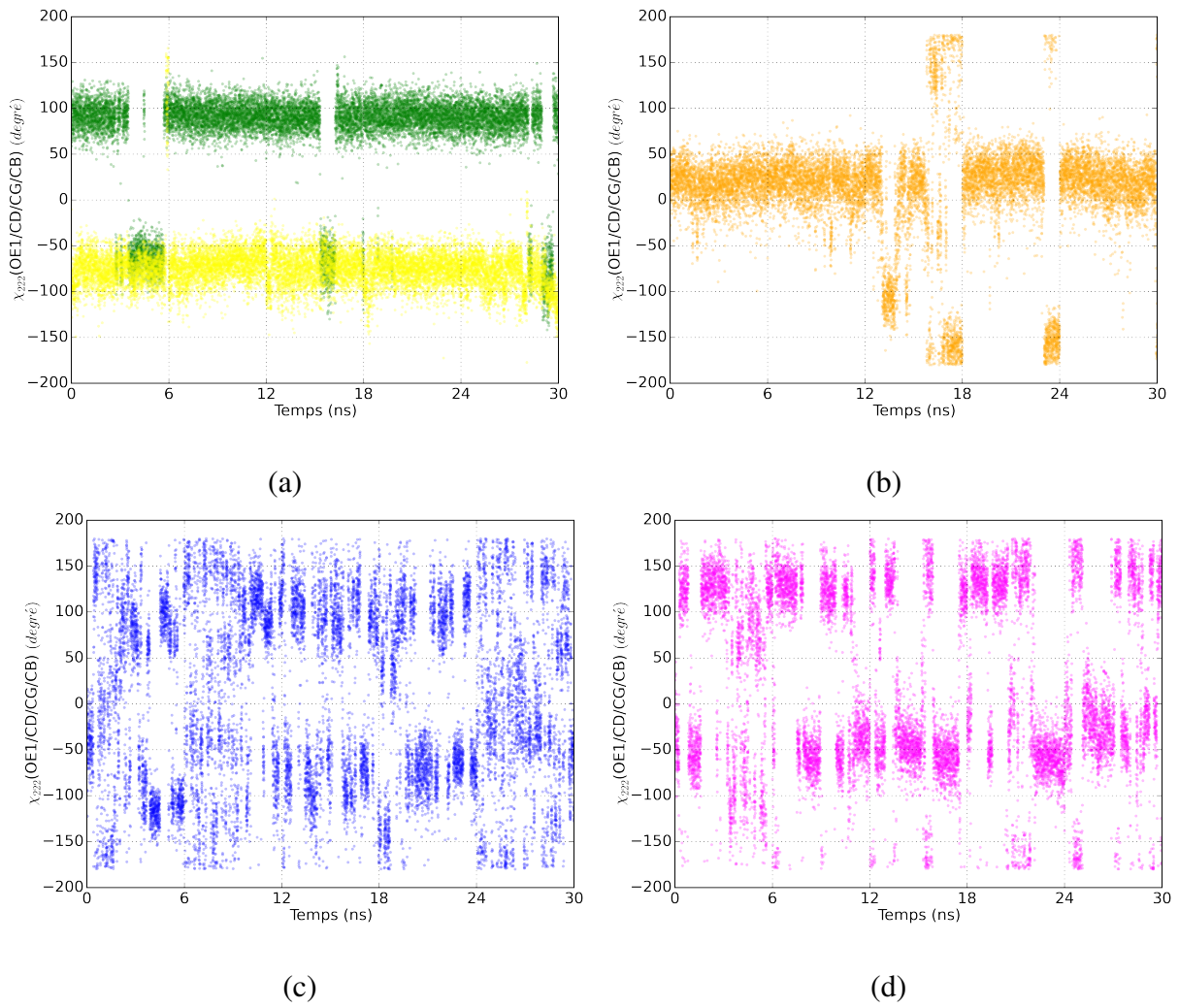


FIGURE III.21 : Trajectoires de l'angle dièdre  $\chi_{222}$  caractérisant l'échange des atomes OE1 et OE2 du Glu222 au cours des dynamiques, (a) AvGFP (vert) et AvGFP-S65T (jaune), (b) AvGFP-S205V (orange), (c) AvGFP-T203V-S205A (bleu) et (d) AvGFP-T203V-S205V (mauve).

## **D Dynamique des réseaux de liaisons hydrogène et chemins de translocation de proton**

Dans cette partie, sont présentées les données issues de l'algorithme, pour les différentes protéines étudiées. Les performances générales de l'algorithme sont évaluées dans une première phase. En second lieu, les résultats sur la recherche des réseaux globaux de LH, des chemins de translocation et de leur dynamique sont détaillés pour chacune de ces protéines.

### **D.1 Performances de l'algorithme**

#### **D.1.a Suivi du pavage tridimensionnel**

Notre stratégie de recherche des liaisons hydrogène repose sur le pavage de la boîte de simulation et l'identification des sous-boîtes susceptibles de contenir les atomes participant aux réseaux de liaisons hydrogène internes à la protéine.

Le pavage de la boîte de simulation a conduit à des sous-boîtes de volume d'environ  $47,4 \text{ \AA}^3$ , ce qui correspond à une longueur moyenne des côtés de ces sous-boîtes de l'ordre de  $3,6 \text{ \AA}$  (table III.6). L'écart type du volume des sous-boîtes pour une protéine ainsi que les petites variations du volume des sous-boîtes observées entre les différentes protéines résultent de l'ajustement du volume de la boîte de simulation afin de maintenir la pression du système à 1 atm.

Par ailleurs, parmi les quatre types de sous-boîtes définies, seules celles constituant les sphères interne et de bord de la protéine et la première sphère externe à la protéine (voir paragraphe II.) nous importent pour la recherche des LH. Compte tenu du mouvement de la protéine dans la boîte de simulation, l'affectation de chaque sous-boîte peut varier au cours de la simulation. Malgré tout, le nombre de sous-boîtes de chaque type de sphère ne devrait pas varier de façon significative au cours de la dynamique, puisque pour toutes les protéines étudiées, aucune déstructuration importante n'a été observée (voir paragraphe III.C.3). Nous avons vérifié cet aspect en suivant la variation au cours de la dynamique du nombre de sous-boîtes dans chacune des trois catégories citées ci-dessus (figure III.22). Les trois courbes obtenues ont une allure similaire et les valeurs du nombre de sous-boîtes restent à peu près constantes pour les trois types de sphères. Des comportements similaires ont été observés pour l'ensemble des protéines étudiées. De plus, le nombre de sous-boîtes de chaque catégorie est semblable d'une protéine à une autre (table III.6). Ce résultat n'est pas surprenant, dans la mesure où les

protéines ne se différencient que par une ou deux mutations, sans perturbation majeure de la structure. Elles présentent toutes une structure tridimensionnelle similaire. La sphère de bord de la protéine est constituée d'environ 685 sous-boîtes, celle de la sphère interne d'environ 185 alors que la première sphère externe, correspondant à la première couche d'hydratation, en compte environ 1140. le volume protéique constitué des sphères interne et de bord de la protéine est de l'ordre de  $41 \times 10^3 \text{ \AA}^3$ . Il est entouré de la première couche d'hydratation d'un volume d'environ  $54 \times 10^3 \text{ \AA}^3$ .

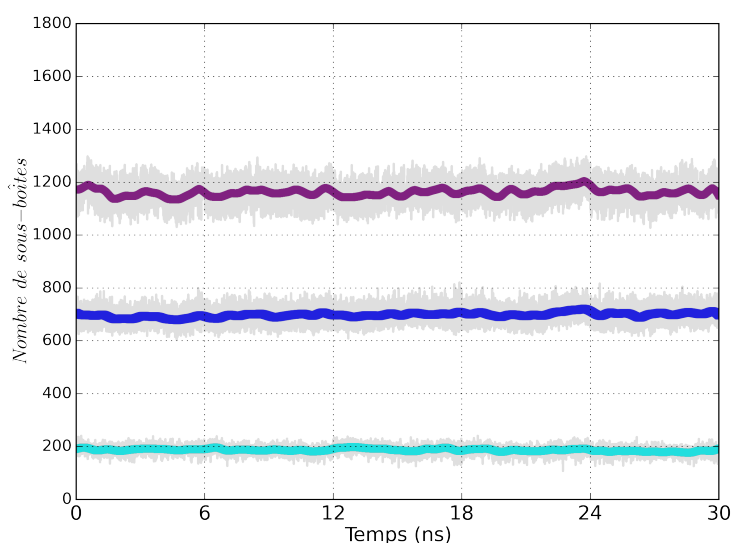


FIGURE III.22 : Variation du nombre de sous-boîtes des sphères internes et de bord de la protéine et de la première sphère externe à la protéine lors du pavage de la boîte de simulation au cours de la dynamique moléculaire de AvGFP (1W7S) (trait fin gris). La moyenne glissante des sous-boîtes (100 instantanés) est représentée en trait gras avec de couleurs différentes pour les sphères interne (cyan), de bord (bleu) et externe (mauve).

Protéine	Volume moyen des sous-boîtes ( $\text{\AA}^3$ )	Volume ( $\times 10^3 \text{ \AA}^3$ )		Nombre de sous-boîtes (%)		
		Protéine*	1 <sup>ière</sup> sphère externe	Sphère interne	Sphère bord	1 <sup>ière</sup> sphère externe
AvGFP	$47,06 \pm 2,61$	41,7	54,7	$187 \pm 15$	$698 \pm 34$	$1163 \pm 46$
AvGFP-S205V	$47,63 \pm 2,75$	41,5	54,5	$183 \pm 15$	$689 \pm 35$	$1144 \pm 49$
AvGFP-S65T	$46,99 \pm 2,55$	41,1	53,4	$189 \pm 16$	$686 \pm 33$	$1136 \pm 46$
AvGFP-T203V-S205A	$47,47 \pm 2,72$	41,2	53,6	$189 \pm 16$	$682 \pm 35$	$1129 \pm 48$
AvGFP-T203V-S205V	$47,36 \pm 2,53$	41,4	53,8	$187 \pm 16$	$688 \pm 32$	$1135 \pm 45$

Tableau III.6 : Principales caractéristiques issues du pavage de la boîte de simulation pour les 5 protéines étudiées. (\*) Le volume de la protéine est défini ici comme le volume des sous-boîtes constituant les sphères interne et de bord de la protéine.

### D.1.b Répartition des molécules d'eau par sphère

La structure cristallographique contient intrinsèquement des molécules d'eau modélisées à partir de la carte de densité électronique par le cristallographe. A ces molécules s'ajoutent celles adjointes dans la boîte de simulation pour représenter le solvant autour de la protéine. Lors des étapes d'équilibration puis lors de la dynamique moléculaire, ces molécules d'eau peuvent s'insérer dans la structure protéique et/ou s'échanger entre la couche de solvation de la protéine et l'intérieur de la protéine. Nous avons déterminé grâce à l'algorithme, le nombre de molécules d'eau contenues dans chacune des trois catégories de sphères au cours de la dynamique moléculaire pour AvGFP (1W7S). Les résultats sont consignés dans la Table III.7. Le nombre de molécules d'eau trouvées dans la sphère interne de la protéine par l'algorithme est similaire au nombre de molécules d'eau modélisée à l'intérieur de la protéine dans la structure cristallographique. De plus, comme en témoigne la figure III.23, nous ne constatons pas d'évolution significative du nombre de molécules d'eau contenues dans chacune des trois sphères d'intérêt pour la recherche des réseaux de liaisons hydrogène au cours de la simulation. Ce contenu est stable et ne dépend pas de façon significative de la protéine étudiée (table III.7). Le nombre croissant de molécule d'eau de la sphère interne à la première sphère souligne la solvation progressive de la protéine de son intérieur vers sa couche de solvation. L'ensemble de ces résultats indique que les phases d'équilibration puis de production de la simulation n'induisent pas une insertion importante de molécules d'eau dans la protéine qui aurait pu augmenter de façon aberrante le nombre de réseaux de liaisons hydrogène.

Protéine	Nombre de molécules d'eau		
	Sphère interne	Sphère de la protéine	1 <sup>ière</sup> sphère externe
AvGFP	15±2	313±24	1790±51
AvGFP-S205v	15±2	309±35	1731±59
AvGFP-S65T	13±2	307±24	1697±58
AvGFP-T203V-S205A	13±2	308±24	1698±48
AvGFP-T203V-S205V	13±2	308±23	1658±57

Tableau III.7 : Répartition des molécules d'eau entre les trois sphères considérées pour la recherche des liaisons hydrogène pour les différentes protéines étudiées.

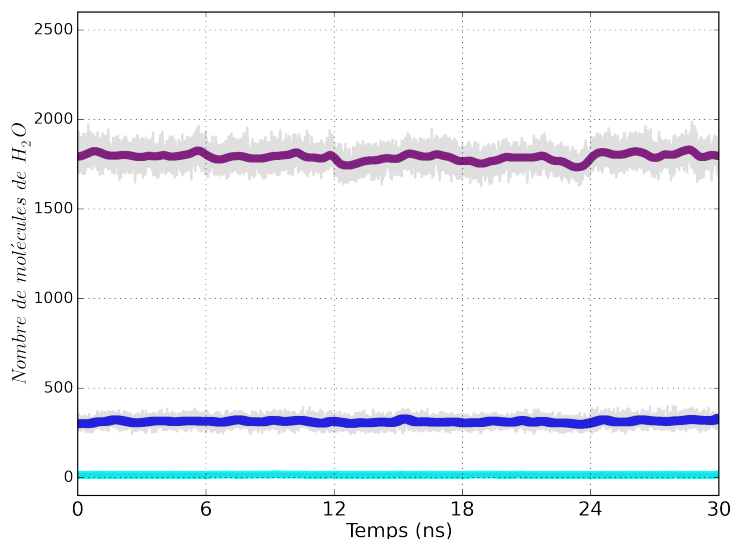


FIGURE III.23 : Distribution des molécules d'eau entre les différentes catégories de sphères considérées pour la recherche des LH le long de la dynamique pour AvGFP (1W7S) : sphères interne (cyan) et de bord (bleu) de la protéine, première sphère d'hydratation (mauve)

### D.1.c Distribution des liaisons hydrogène

Pour la recherche des LH, nous avons utilisé un critère usuel de distance et d'angle : Une liaison hydrogène existe entre un donneur D-H et un accepteur A si la distance D-A est inférieure à  $3,2 \text{ \AA}$  et l'angle  $\Omega (\widehat{DHA})$  compris entre  $120^\circ$  et  $180^\circ$ . La figure III.24 montre la distribution en distance et en angle de l'ensemble des LH, connectées au phénol du chromophore, détectées par l'algorithme dans le cas de AvGFP. Le maximum de la distribution des LH non intra-résidus se situe pour une distance de  $2,8 \text{ \AA}$  et un angle d'environ  $165^\circ$ , ce qui correspond à des LH assez fortes. Les liaisons les plus fortes détectées impliquent majoritairement des atomes lourds de résidus proches alors que la participation d'une molécule d'eau dans la liaison hydrogène a tendance à affaiblir celle-ci.

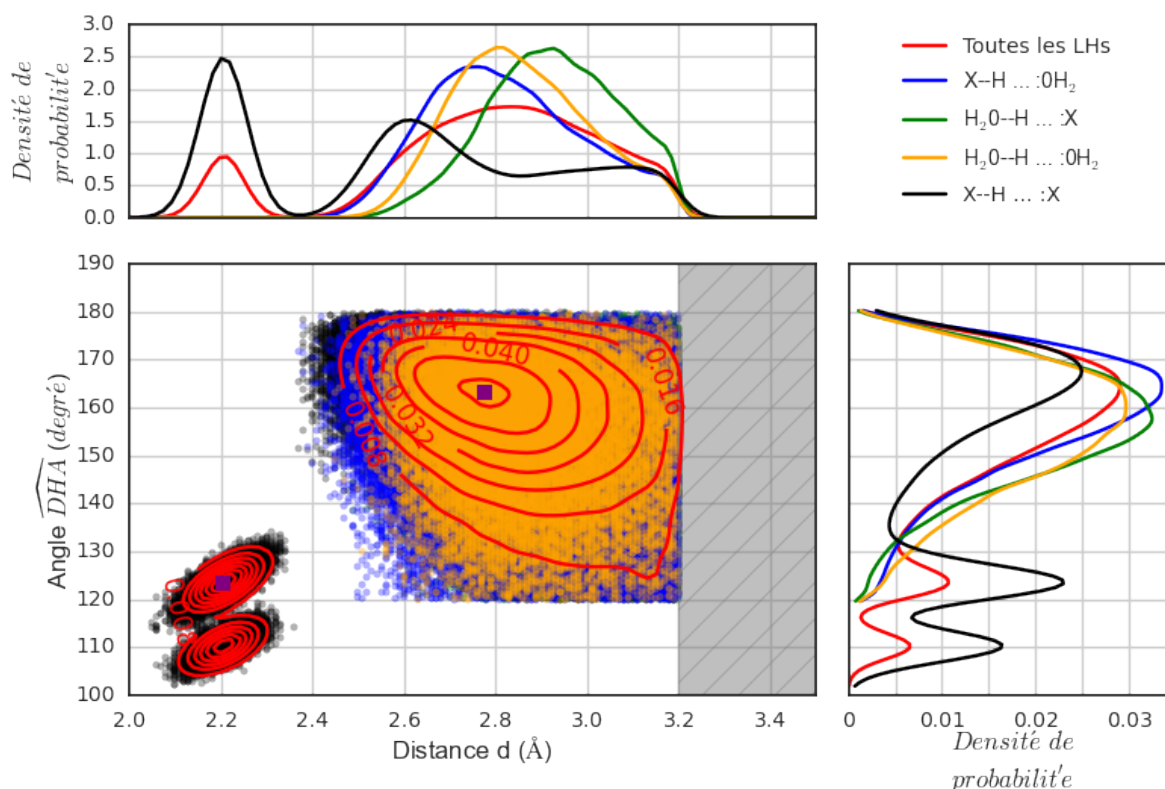


FIGURE III.24 : Distribution des LH, connectées au groupe hydroxyle du phénol du chromophore, détectées par l'algorithme lors de l'analyse de la dynamique de AvGFP (1W7S). Les deux distributions obtenues pour une distance  $d$  d'environ 2,2 Å et des angles de 110° et 125° correspondent aux LH intra-résidus (Asp, Glu, His). La distribution la plus large centrée à une distance de 2,8 Å et un angle d'environ 165° correspond aux LH entre résidus, entre un résidu et une molécule d'eau ou encore entre deux molécules d'eau.

#### D.1.d Distribution des réseaux de liaisons hydrogène

Un des objectifs finaux de l'algorithme est l'agrégation des LH connexes à un instant donné pour générer les réseaux de LH présents à cet instant. Dans le cas de AvGFP, l'algorithme rapporte en moyenne  $123 \pm 6$  réseaux de LH pour chaque instantané. Cette moyenne est similaire pour toutes les protéines étudiées ce qui suggère, à ce stade, que les mutations ne modifient pas de façon significative le nombre de réseaux par instantané. Une analyse plus détaillée de la trajectoire du nombre de réseaux au cours de la simulation indique que dans le cas de AvGFP le nombre moyen de réseaux est assez stable au sein d'une même dynamique de 6 ns et entre les 5 dynamiques. Pour les autres protéines, nous constatons une fluctuation légèrement plus importante du nombre de réseaux.

Protéine	# de réseaux de LH
AvGFP	123±6
AvGFP-S205V	123±7
AvGFP-S65T	122±6
AvGFP-T203V-S205A	122±7
AvGFP-T203V-S205V	123±7

Tableau III.8 : Nombre moyen de réseaux de liaisons hydrogène par instantané pour AvGFP et quatre de ses mutants étudiés.

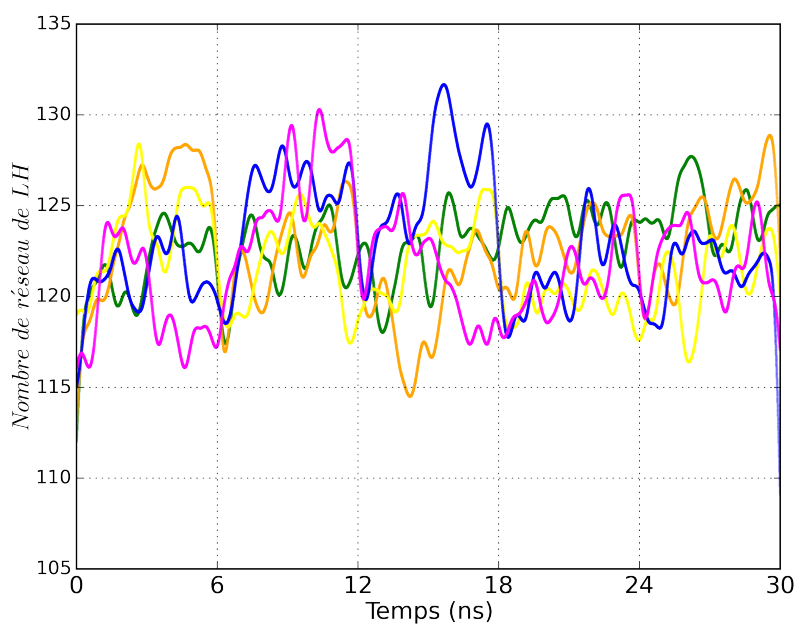


FIGURE III.25 : Trajectoires du nombre de réseaux de LH au cours de la simulation pour les simulations de AvGFP (vert), AvGFP-S65T (jaune), AvGFP-S205V (orange), AvGFP-T203V-S205A (bleu) et AvGFP-T203V-S205V (mauve)



## D.2 Réseau global de liaisons hydrogène connecté au phénol du chromophore

Pour valider l'algorithme mis au point, nous avons restreint notre analyse aux réseaux de liaisons hydrogène connectés au phénol du chromophore. Pour chacune des 5 protéines étudiées, nous avons analysés en détail le réseau global et les chemins de translocation de proton impliquant le phénol du chromophore.

### D.2.a AvGFP (1W7S)

#### D.2.a i Réseau global de liaisons hydrogène

D'une manière générale, le réseau global observé dans le cas de AvGFP met en lumière plusieurs chemins distincts ayant tous le glutamate 222 comme accepteur final (figure III.26). Deux chemins impliquent le phénol du chromophore et sont susceptibles de se prolonger jusqu'à la chaîne latérale de l'histidine 148. Les autres connectent l'autre atome du carboxylate du glutamate 222, OE1, avec la valine 68 et/ou la glutamine 69. Certains de ces chemins sont constitués d'un ensemble de liaisons hydrogène présentant une forte densité de probabilité et coexistent tout au long de la simulation. Dans ce contexte, Glu222 apparaît comme étant un atome central fortement ancré dans son environnement grâce à un ensemble de liaisons hydrogène dont la densité de probabilité est supérieure à 90%. Plusieurs groupes sont prépondérants dans cet ancrage : un ensemble de molécules d'eau ainsi que les hydroxyles des résidus Ser65 du chromophore et Ser205. Nous notons que malgré l'échange entre OE1 et OE2 observé au cours de la simulation et rapporté précédemment (voir paragraphe III.C.4.f), les chemins de translocation incluant le phénol du chromophore implique uniquement l'atome OE2.

De plus, la représentation sagittale indique que le réseau global de liaisons hydrogène observés dans AvGFP fait apparaître des molécules d'eau provenant uniquement de la structure cristallographique. Notons que le dédoublement des liaisons hydrogène impliquant les molécules d'eau retranscrit leur rotation au cours de la dynamique. La représentation sagittale fait également ressortir le caractère central et stable de ces molécules d'eau dans les réseaux. Notamment, la molécule Wat215 intervient dans le chemin de translocation susceptible de sous-tendre l'ESPT. Au cours de la dynamique, elle est maintenue par diverses liaisons hydrogène impliquant huit atomes distincts. De même, les trois autres molécules d'eau intervenant dans les chemins connectant Glu222/OE1 et Gln69/NE2 ou Val68/N s'interconnectent entre elles et sont égale-

ment maintenues par liaisons hydrogène avec des atomes des résidus voisins. Deux d'entre elles (Wat79 et Wat239) s'échangent au cours de la simulation, et cet échange s'effectue avec une faible densité de probabilité (<4%). Cette stabilité des molécules d'eau dans le cas de AvGFP va de pair avec la stabilité des liaisons hydrogène et des chemins de translocation dans lesquelles elles sont impliquées.

Enfin, l'algorithme nous permet de répertorier les réseaux de liaisons hydrogène s'étendant jusqu'aux sous-boîtes de la première sphère externe, à savoir se terminant par une molécule d'eau, ou encore ceux qui restent dans la sphère interne de la protéine. Dans le cas de avGFP, nous avons constaté que le phénol du chromophore est connecté avec le solvant durant 59,5% du temps de simulation alors qu'il reste confiné dans la sphère interne 33,8% du temps de dynamique. Une analyse plus fine de ces réseaux permettrait d'identifier les chemins de sortie ou d'entrée des protons provenant du solvant.

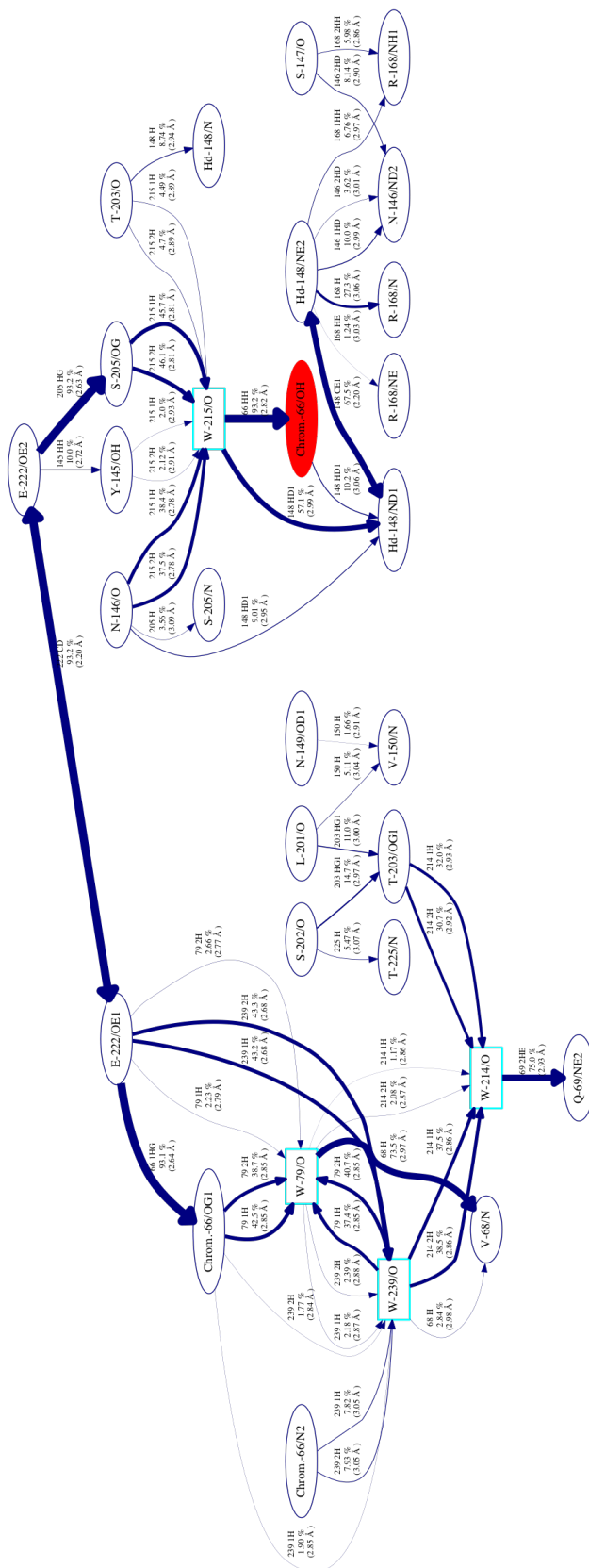


FIGURE III.26 : Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP (1W7S). Seules les LH dont la densité de probabilité est supérieure à  $T = 0,5\%$  sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge.

### D.2.a ii Chemin de translocation de protons impliquant le phénol du chromophore

Nous avons analysé plus en détail les chemins de translocation entre le phénol du chromophore et Glu222/OE2. Ces chemins sont au nombre de deux : le chemin de translocation usuel proposé dans la littérature et un chemin alternatif.

**Chemin de translocation usuel** : L'algorithme retrouve le chemin de translocation proposé dans la littérature impliquant, du donneur initial à l'accepteur final, l'hydroxyle du phénol (Chrom.66/OH), une molécule d'eau, ici Wat215, l'hydroxyle de Ser205 (Ser205/OG) et enfin OE2 de Glu222 (figure III.27). Les liaisons hydrogène qui composent ce chemin ont des densités de probabilité importantes, entre 91,8% et 93,2%. Les longueurs des liaisons hydrogène de ce chemin sont égales à  $2,82 \pm 0,14$  Å et  $2,81 \pm 0,13$  Å pour les liaisons Wat-215/O...H-Chrom.66/OH et Ser205/OG...H-Wat215/O, respectivement. La liaison hydrogène entre Glu222/OE2 et Ser205/OG est la plus courte avec des valeurs comprise entre 2,4 et 3,1 Å et une valeur moyenne de  $2,63 \pm 0,11$  Å (figure III.27).

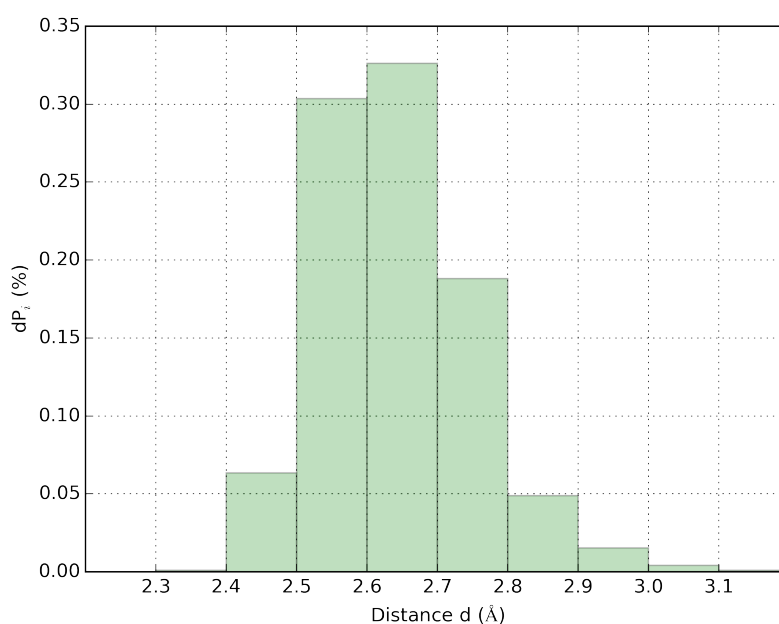


FIGURE III.27 : Distribution de distances de la liaison hydrogène entre Glu222/OE2 et Ser205/OG au cours de la dynamique de AvGFP (1W7S).

Les fortes densités de probabilité de ces diverses liaisons hydrogène soulignent la stabilité du chemin de translocation qu'elles forment et dont la densité de probabilité totale est de 91,8%. Sa longueur dynamique est  $8,27 \pm 0,23$  Å. Enfin, 10,2% du temps de dynamique, le chemin de

translocation peut partir de l'atome ND1 protoné de la chaîne latérale de l'histidine 148 alors en liaison hydrogène avec le phénol du chromophore et la densité de probabilité du chemin de translocation de His148/ND1-H à Glu222/O2 est réduite à 10,2%. Cette liaison hydrogène s'établit de façon très transitoire.

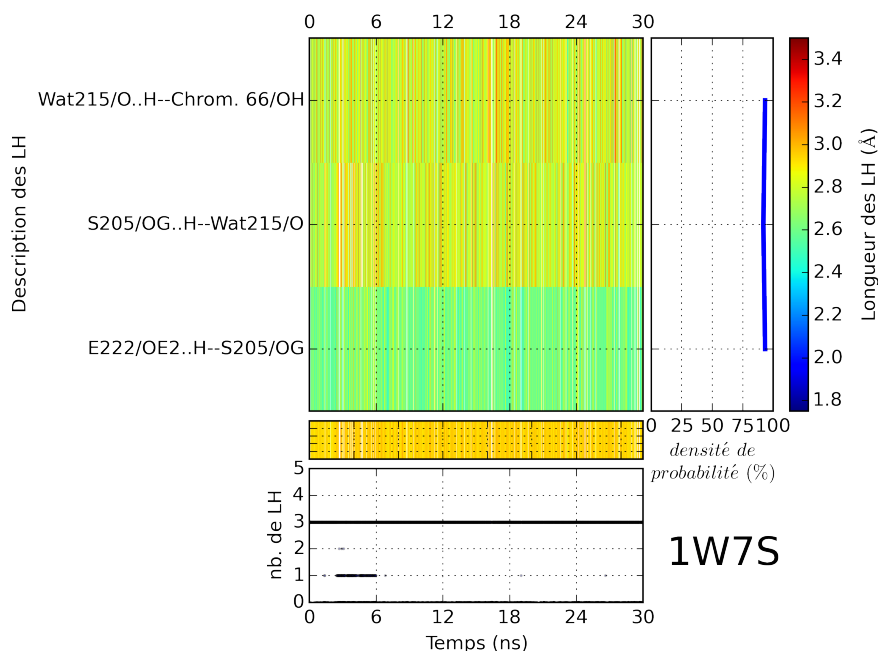


FIGURE III.28 : Représentation matricielle du chemin de translocation usuellement décrit dans la littérature entre Chrom.66/OH et Glu222/OE2 dans le cas de AvGFP (1W7S)

**Chemin de translocation alternatif** : La représentation sagittale met en évidence un chemin de translocation alternatif au chemin usuel. Ce chemin utilise comme relais l'hydroxyle de la Tyr145 au lieu de celui de la sérine 205. Ce réseau alternatif s'établit donc entre les différents groupes suivants : Chrom66/OH, Wat215, Tyr145/OH, Glu222/O2 (figure III.30.c). Sa densité de probabilité est faible et s'élève à 4,05%. Elle est due au caractère peu stable de la liaison hydrogène qui s'établit entre un doublet non liant de Tyr145/OH et un hydrogène de la molécule d'eau Wat215 et dont la densité de probabilité est de 4,12% (figure III.29). La longueur de ce réseau alternatif est  $8,47 \pm 0,15$  Å. Les résidus 145 à 148 sont localisés sur une partie faiblement structurée du brin  $\beta_7$  (voir paragraphe III.D.1.a) qui a la particularité d'avoir les deux chaînes latérales des résidus 146 et 147 tournées vers le solvant, celles des résidus 145 et 148 étant tournées vers l'intérieur du tonneau. L'analyse des différents instantanés mettant en évidence ce chemin alternatif indique que celui-ci résulte d'un changement conformationnel de la chaîne latérale et non d'un mouvement de grande ampleur du brin  $\beta_7$ . La réorientation de la chaîne latérale du résidu Tyr145 conduit à un déplacement d'environ 4 Å de cette dernière vers Glu222

et la molécule d'eau Wat215. Elle est concomitante avec la réorientation de deux acides aminés apolaires qui sont dans son environnement direct, Val61 et Leu220. La rotation de leur chaîne latérale respective d'environ  $120^\circ$  libère un volume suffisant pour accommoder Tyr145 qui établit une liaison hydrogène avec Glu222. Ces événements se produisent de façon transitoire et la durée cumulée de l'établissement ce chemin alternatif est de 1,44 ns.

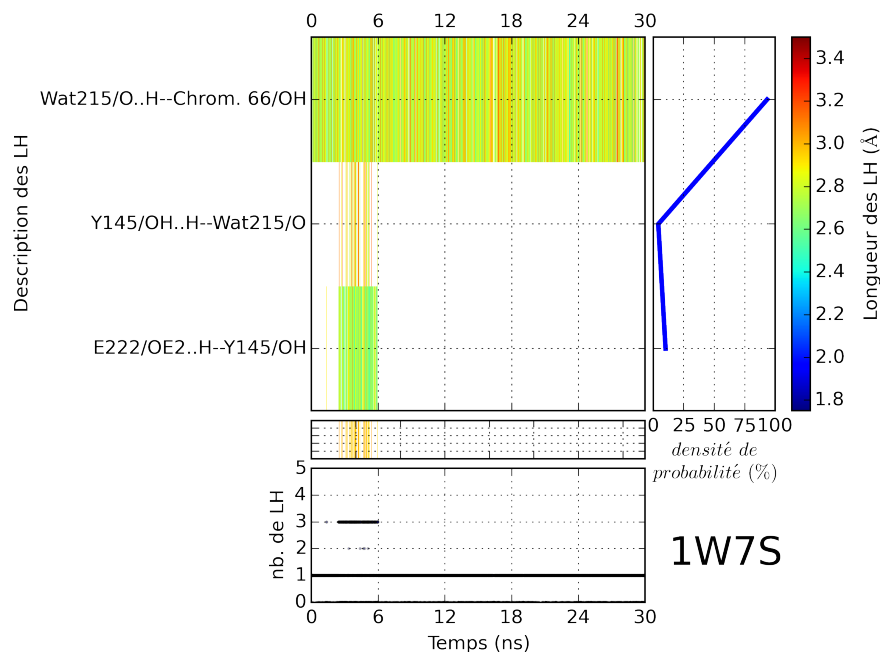
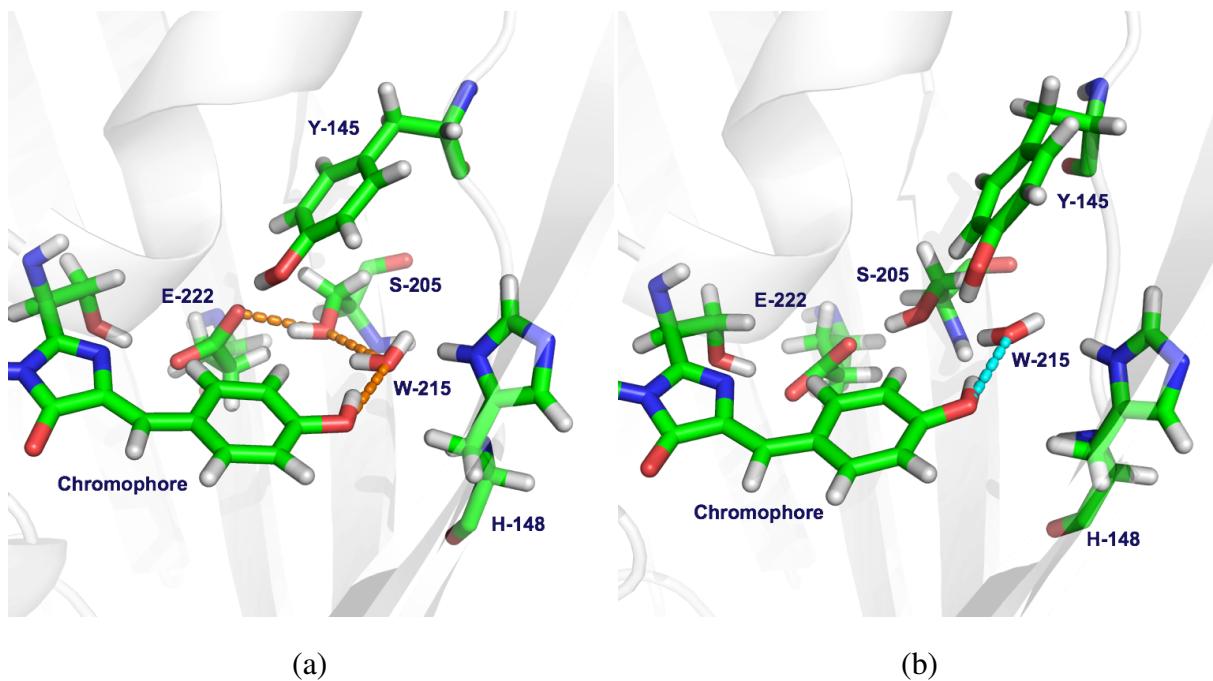
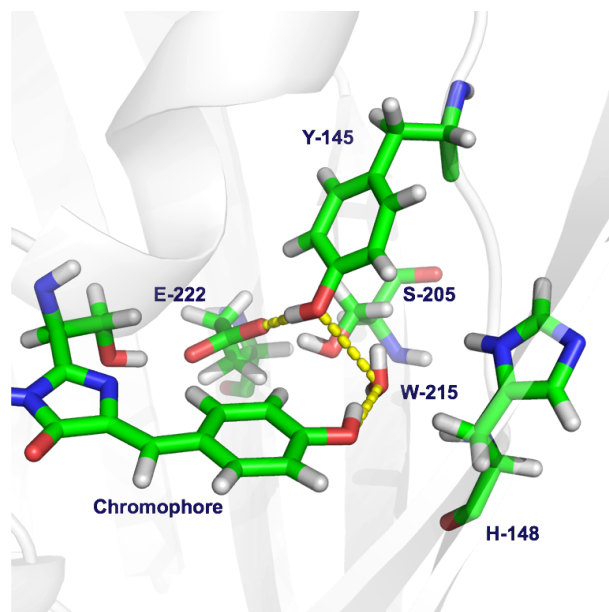


FIGURE III.29 : Représentation matricielle du chemins de translocation alternatif entre Chrom.66/OH et Glu222/OE2 dans le cas de AvGFP (1W7S)





(c)

FIGURE III.30 : Représentation tridimensionnelle (a) du chemin supportant l'ESPT usuellement décrit dans la littérature, (b) rupture transitoire du chemin de translocation ( $t = 2,09$  ns) et (c) chemin alternatif utilisant Tyr145/OH comme relais pour les chemins de translocation entre Chrom.66/OH et Glu222/OE2 retrouvés avec l'algorithme ( $TH > 0,5\%$ ),

**Récapitulatifs des chemins de translocation entre Chrom.66/OH et Glu222/OE2 :** Ainsi, dans le cas de AvGFP (1W7S), deux chemins de translocation de protons connectés au groupe phénol du chromophore ont été mis en évidence par l'algorithme. Le premier, majoritaire, est le chemin proposé dans la littérature pour sous-tendre l'ESPT observé dans cette protéine. Le second, minoritaire, implique l'hydroxyle de la tyrosine 145 suite à un changement conformationnel de sa chaîne latérale au lieu de l'hydroxyle de la sérine 205. Ce chemin alternatif existe lorsque His148/NE1 n'est ni en liaison hydrogène avec le phénol du chromophore ni transitoirement en liaison hydrogène avec la molécule d'eau pivot Wat215. En revanche, ce chemin alternatif peut exister en même temps que le chemin usuel. La non concomitance entre les deux chemins s'élève à 1,44% du temps de la simulation. Aussi, la densité de probabilité d'existence d'un chemin de translocation entre Glu222/OE2 et Chrom.66/OH dans AvGFP est de 93,2%, indépendamment du chemin suivi.

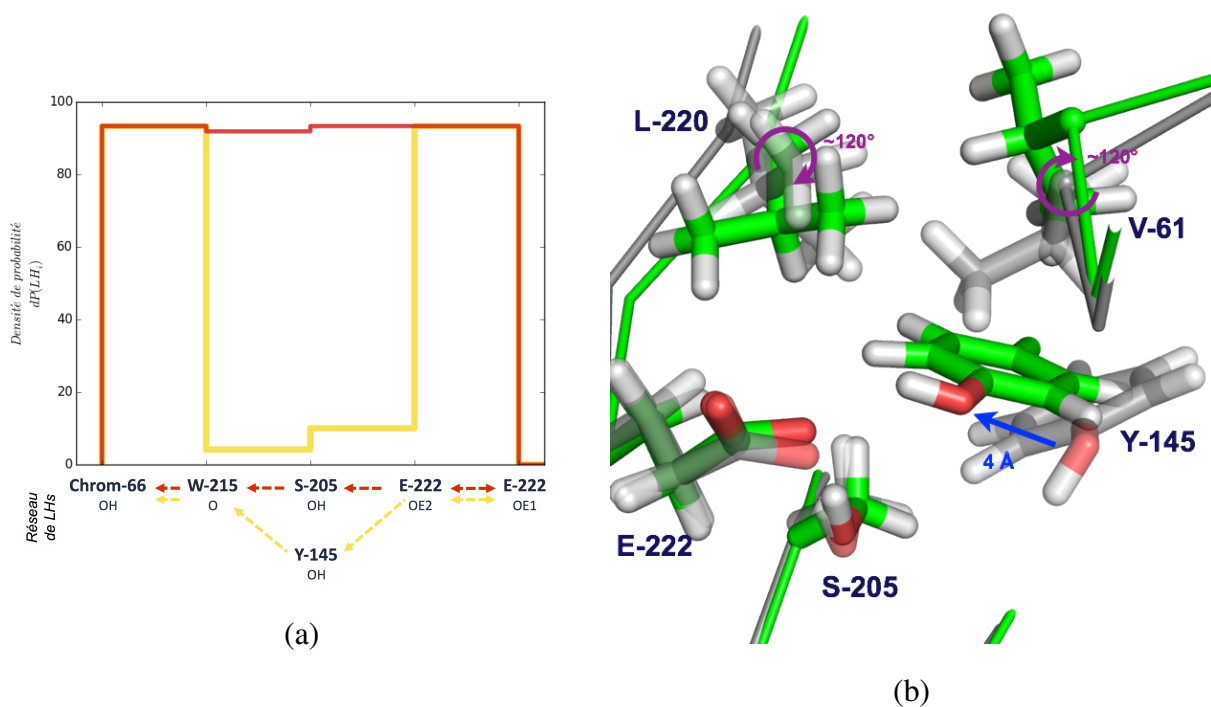


FIGURE III.31 : Schéma récapitulant les chemins de translocation entre Chrom66./OH et Glu222/OE2 trouvés dans la dynamique de AvGFP (1W7S), (a) Densité de probabilité des différentes LH constituant les chemins qui sont connectés au phénol du chromophore (Chrom.66/OH) et (b) Réarrangement du résidu Tyr145 pour l'établissement du chemin alternatif. Les instantanés gris et vert correspondent respectivement à  $t = 2$  ns et  $t = 2,5$  ns.



### **D.2.b AvGFP-S205V (2QLE)**

#### **D.2.b i Réseau global de liaisons hydrogène**

Comme attendu, la suppression de l'hydroxyle de la sérine 205 et la mutation de ce résidu par une valine rompt le chemin de translocation usuel précédemment observé dans avGFP (cf. figure III.32). La représentation sagittale du réseau global de liaisons hydrogène dans le mutant AvGFP-S205V révèle un chemin majoritaire entre Chrom.66/OH et Glu222 dont les atomes relais intermédiaires sont Thr203/OG1 et Wat62/O. Elle souligne de nouveau le rôle central du glutamate 222 comme accepteur final d'un ensemble de chemins de translocation de proton. Ces chemins ont comme donneurs de protons initiaux l'hydroxyle du chromophore ou l'atome ND1 de la chaîne latérale de l'histidine 148, comme pour AvGFP. Le groupe N-H de la valine 205 peut avec une faible densité de probabilité être un donneur de proton. Nous retrouvons également les chemins qui se propagent jusqu'aux résidus Val68 et Gln69, impliquant les mêmes atomes relais que pour AvGFP. Ces chemins mettent toujours en jeu des molécules d'eau intermédiaires, Wat13, Wat15 et Wat45 qui certes s'échangent entre elles mais restent ancrées à proximité du chromophore. La représentation sagittale met également en lumière l'échange au cours de la dynamique entre les atomes Glu222/OE1 et Glu222/OE2, due à la rotation du groupement carboxyle de ce résidu. Cet échange se traduit en particulier par la participation de l'atome OE1 comme accepteur final du chemin impliquant le phénol du chromophore. Enfin, le nombre total de réseaux de liaisons hydrogène atteignant la première sphère externe sont au nombre de 3973 soit 26,5% de temps de la dynamique, connectant le solvant au phénolate du chromophore.

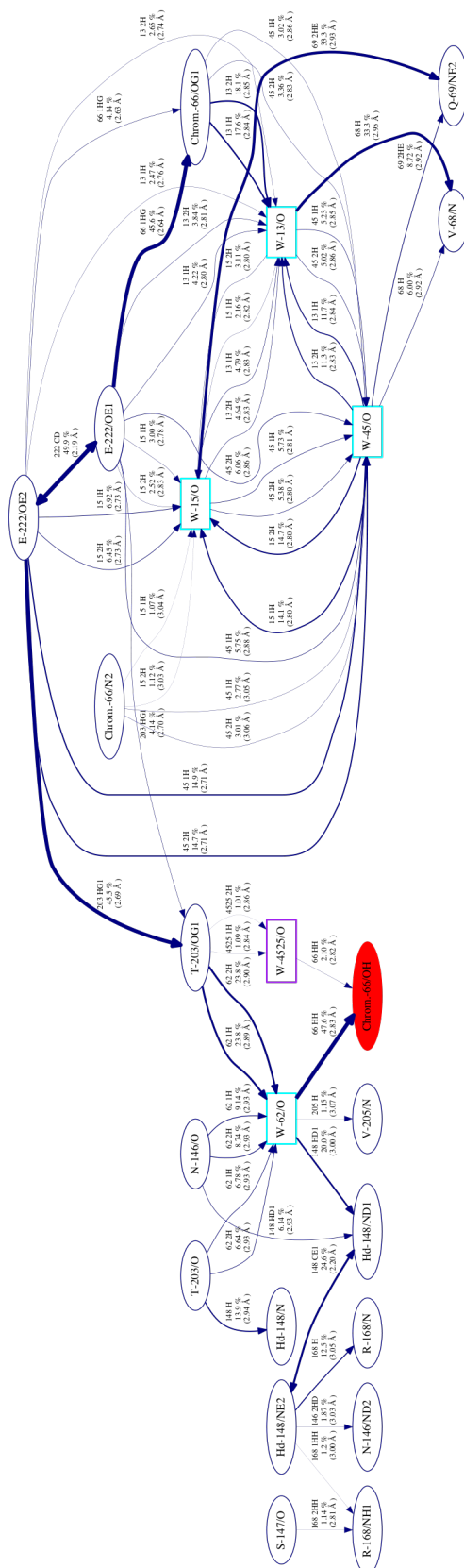
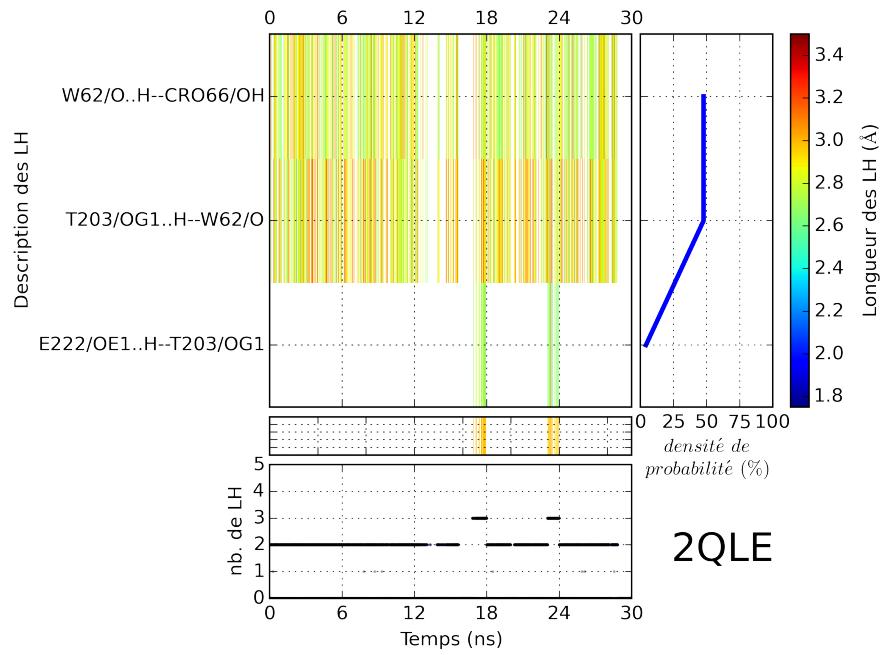


FIGURE III.32 : Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-S205V (2QLE). Seules les liaisons hydrogène dont la densité de probabilité est supérieure à  $T = 1\%$  sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge.

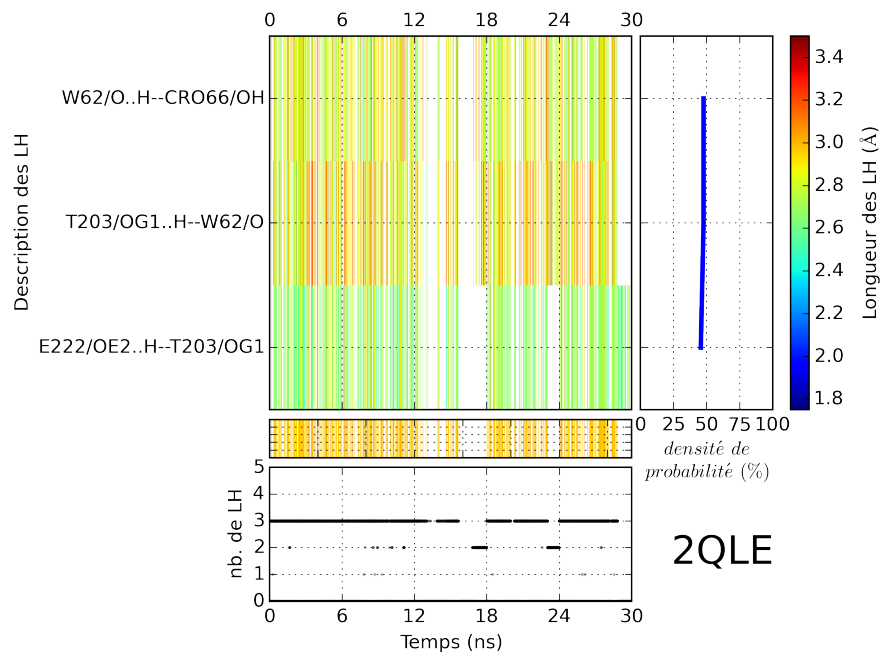
### **D.2.b ii Chemin de translocation de protons impliquant le phénol du chromophore.**

Deux chemins de translocation connectés au phénol du chromophore apparaissent sur la représentation sagittale : un majoritaire, l'autre minoritaire.

**Chemin de translocation principal :** La représentation sagittale indique l'existence d'un réseau de liaisons hydrogène prépondérant identique à celui proposé dans la littérature [Shu et al., 2007]. Il implique, du donneur initial à l'accepteur final, les atomes/groupes suivants : Chrom.66/OH, Wat62/O, Thr203/OG1 et Glu222/OEx, où OEx désigne OE1 ou OE2 (cf. figure III.35). Compte tenu de la rotation du carboxyle de Glu222 conduisant à un échange entre OE1 et OE2, le chemin de translocation s'établit avec l'atome OE2 durant 13,0 ns (43,3%) et avec l'atome OE1 1,23 ns (4,09%) (cf. figure III.33). Plusieurs pertes de continuité dans le chemin de translocation principal entre Glu222/OE2 et le phénol du chromophore sont constatées dont six ruptures d'une durée de plus de 0,2 ns, répertoriées sur la figure III.34. Ces ruptures correspondent à la rotation du groupement carboxyle de Glu222. Cependant, le chemin vers le chromophore n'est qu'en partie restauré, grâce à l'implication de Glu222/OE1 (cf. figure III.35.a) avec deux temps de présence transitoire entre 16,8 et 18 ns d'une part et 23,0 et 24 ns d'autre part. Globalement, le chemin de translocation entre Chrom.66/OH et Glu222/OEx présente une densité de probabilité d'existence de 47,4%.



(a)



(b)

FIGURE III.33 : Représentations matricielles du chemin de translocation principal entre Chrom.66/OH et Glu222, (a) chemin entre Chrom.66/OH et Glu222/OE1 et (b) a) Chemin entre Chrom.66/OH et Glu222/OE2 pour la dynamique de AvGFP-S205V (2QLE).

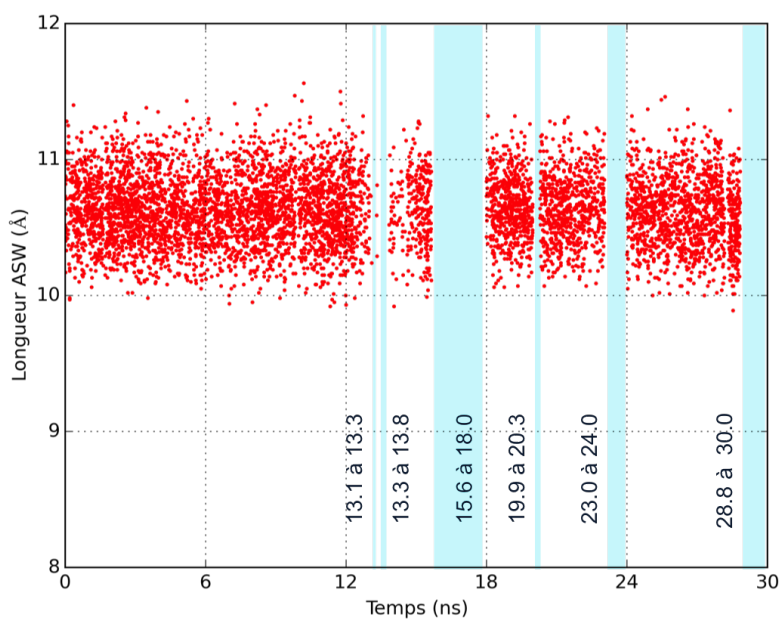


FIGURE III.34 : Longueur du chemin de translocation de proton entre Chrom66./OH et Glu222/OE2 dans AvGFP-S205V au cours de la simulation. Les zones de temps de rupture du chemin sont représentées en cyan et les temps de début et fin de la rupture sont précisés.

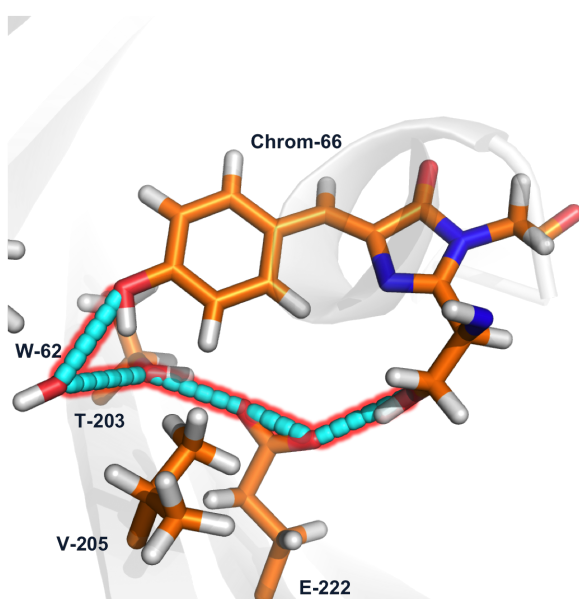
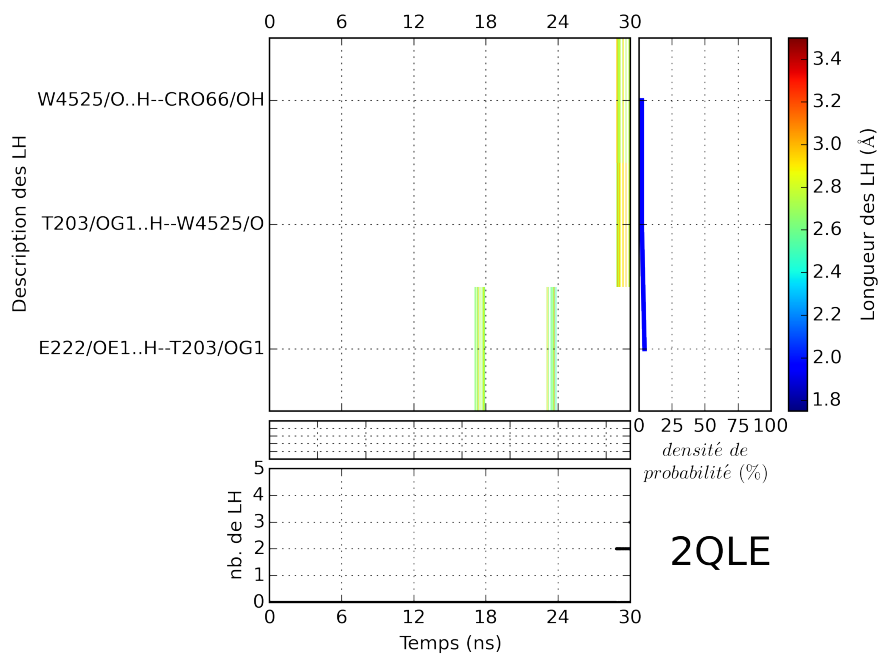
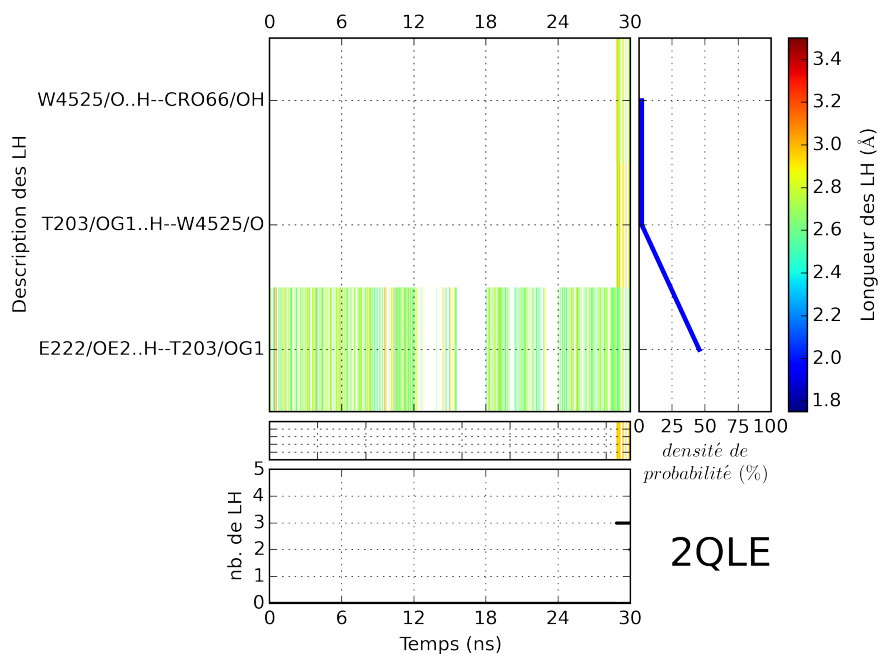


FIGURE III.35 : Représentation tridimensionnelle du chemin de translocation principal entre Chrom.66/OH et Glu222 pour AvGFP-S205V.

**Chemin de translocation secondaire :** La coupe sagittale de AvGFP-S205V (cf. figure III.32) suggère l'existence d'un chemin alternatif au chemin principal. Ce chemin secondaire semble emprunter un trajet similaire avec une molécule d'eau différente du chemin principal. Cette molécule d'eau provient de la dynamique moléculaire (Wat4525). Ce chemin secondaire passe par les atomes relais suivants Chrom.66/OH, Wat4525/O, Thr203/OG1 et Glu222/OEx et est donc constitué des mêmes composants que le chemin principal. Il apparaît uniquement lors de l'échange de la molécule d'eau cristallographique Wat62 avec la molécule d'eau Wat4525 issue de la dynamique moléculaire à partir de 28,8 ns.



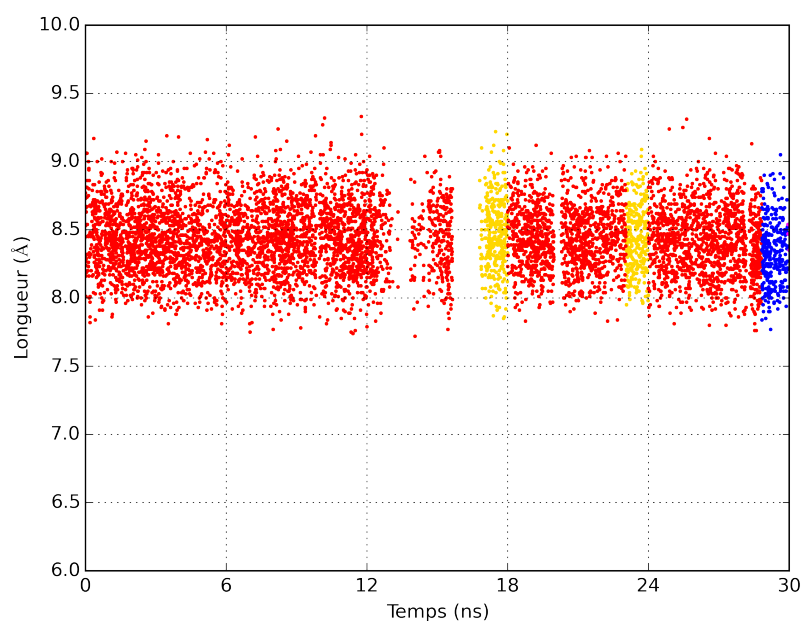
(a)



(b)

FIGURE III.36 : Représentation tridimensionnelle du chemin de translocation secondaire entre Chrom66/OH et Glu222 faisant intervenir une molécule d'eau cristallographique pour la dynamique de AvGFP-S205V (2QLE), (a) chemin entre Chrom.66/OH et Glu222/OE1, (b) chemin entre Chrom.66/OH et Glu222/OE2

**Récapitulatifs des chemins de translocation entre Chrom.66/OH et Glu222 :** Dans le cas de AvGFP-S205V, l'algorithme retrouve le chemin de translocation de protons proposé par Shu et coll. sur la base de la structure cristallographique. Ce réseau est constitué de Chrom.66/OH, lié à une molécule d'eau (Wat62 ou Wat45252) suivi de Thr203/OG1 et qui retrouve le groupe carboxyle de Glu222 (cf. figure III.37). Le chemin secondaire passant par la molécule d'eau Wat4525 est présent 2.10% du temps de la dynamique soit 632 ps. La présence du chemin secondaire n'est pas concomitante avec celle du chemin principal. Par conséquent la densité de probabilité d'existence d'un chemin entre Chrom.66/OH et le groupe carboxyle de Glu222 est de 49.5%. Indépendante de l'origine de la molécule d'eau relais, et de la nature de l'atome accepteur final (OE1 ou OE2) la longueur moyenne de ce chemin de translocation de proton est  $8,41 \pm 0,20 \text{ \AA}$  (cf. figure III.37)



*FIGURE III.37 : Longueur des chemins de translocation entre Chrom.66/OH et Glu222/OEx dans AvGFP-S205V (2QLE). Points rouges : chemin prépondérant avec Glu222/OE2 comme accepteur final (longueur moyenne du chemin,  $\langle l_{chemin} \rangle = 8,45 \pm 0,25 \text{ \AA}$ ). Points oranges : même chemin avec Glu222/OE1 comme accepteur final ( $\langle l_{chemin} \rangle = 8,43 \pm 0,25 \text{ \AA}$ ). Les chemins secondaires impliquant la molécule d'eau issue de la dynamique, Wat4525, sont représentés en bleu et violet suivant la nature de l'atome accepteur final, respectivement Glu222/OE1 ( $\langle l_{chemin} \rangle = 8,43 \pm 0,08 \text{ \AA}$ ) ou Glu222/OE2 ( $\langle l_{chemin} \rangle = 8,34 \pm 0,23 \text{ \AA}$ ).*

Enfin, la liaison interne au glutamate 222 entre OE1 et OE2 existe en permanence le long de la dynamique moléculaire. Elle n'intervient que 49,9% du temps de la dynamique dans le réseau de liaisons hydrogène connecté au phénol du chromophore. Nous nous attendons donc à



ce que, globalement, les atomes OE1 ou OE2 soient connectés ce même pourcentage de temps avec un ou plusieurs atomes eux-mêmes connectés par liaisons hydrogène à Chrom.66/OH. Une différence de 0,38% est trouvée entre le temps de participation de la liaison interne de Glu222 et la somme des densités de probabilité des chemins allant au phénol. Cette différence suggère que pour 59 instantanés, le chemin de translocation de protons s'établit par des chemins alternatifs. Ces chemins très transitoires n'ont pu être mis en avant lors de cette analyse.

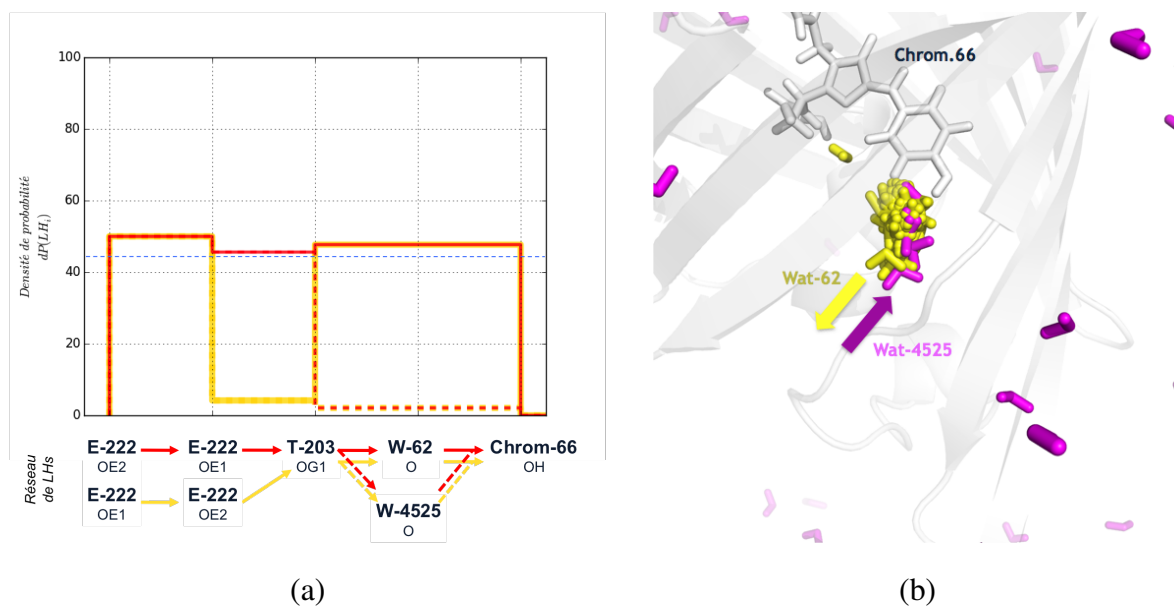


FIGURE III.38 : (a) Densité de probabilité des chemins de translocation entre Chrom.66/OH et Glu222/OEx obtenus lors de l'analyse de la dynamique moléculaire de AvGFP-S205V. (b) Echange de molécules d'eau observé à 28,8 ns de la dynamique entre Wat62 (jaune) et Wat4525 (violet) dans AvGFP-S205V.

### **D.2.c AvGFP-T203V-S205A (4OGS)**

#### **D.2.c i Réseau global de liaisons hydrogène**

Le remplacement de la thréonine en position 203 par une valine, en supplément de la mutation en position 205 qui visent toutes les deux à substituer un hydroxyle à la fois donneur et accepteur de proton par un groupe méthyle, induit logiquement la suppression du chemin de translocation passant par Thr203 dans AvGFP-S205V. L'analyse de la représentation sagittale montre que dans AvGFP-T203V-S205A plusieurs chemins de translocation s'établissent au cours de la simulation entre Chrom.66/OH et Glu222 (cf. figure III.39). Les LH qui composent ces chemins présentent des densités de probabilité ne dépassant pas 12%. Comme précédemment, on retrouve les chemins qui se propagent jusqu'aux résidus Val68 et Gln69 impliquant cinq molécules d'eau (Wat4, Wat40, Wat42, Wat48, Wat79). Ils sont constitués de LH qui apparaissent sur la coupe sagittale avec une faible densité de probabilité. Ces faibles densités sont des artéfacts. En effet, l'algorithme cherche les réseaux de LH qui à un moment ou un autre de la dynamique sont en connexion avec Chrom.66/OH. Il faut donc déjà que ces réseaux connectent à un instant donné le Chrom.66/OH à Glu222/OEx pour atteindre ensuite les réseaux allant jusqu'à Val68 et Gln69. A titre d'exemple, la liaison hydrogène entre Glu222/OE1 et Chrom.66/OG1 apparaît avec une densité de probabilité de 13,9% dans le réseau global de LH connectés à Chrom.66/OH, alors que sa densité de probabilité au cours de la dynamique moléculaire atteint 57,4%.

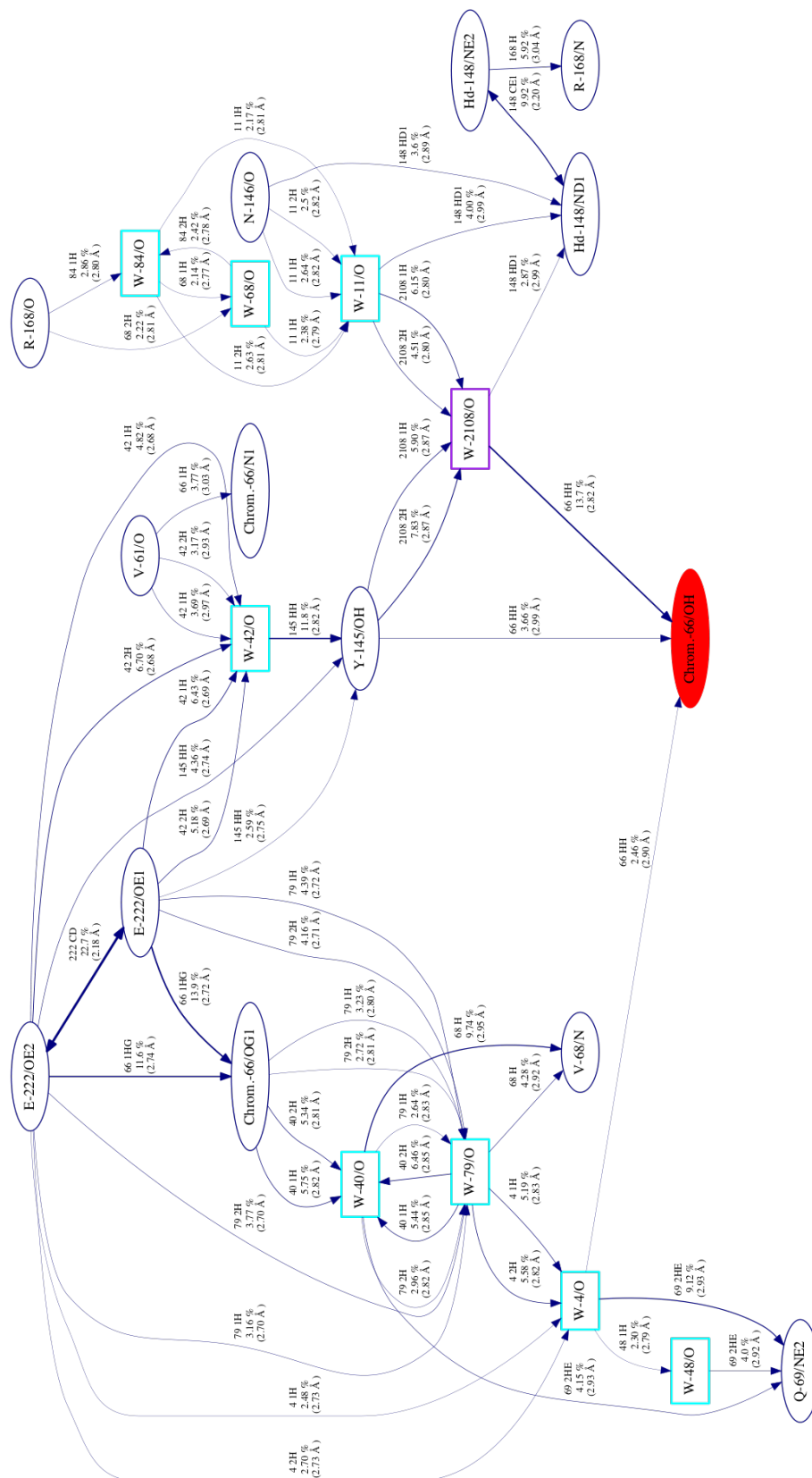


FIGURE III.39 : Coupe sagittale présentant le réseau global de LH connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-T203V-S205A (4OGS). Seules les LH dont la densité de probabilité est supérieure à  $T = 2\%$  sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge.

### D.2.c ii Chemin de translocation de protons impliquant le phénol du chromophore.

La densité de probabilité de la liaison intramoléculaire Glu222/OEx apparaissant sur la coupe sagittale indique que des réseaux de LH s'établissent entre Chrom.66/OH et Glu222/OEx durant 22,7% de la dynamique dans avGFP-T203V-S205A. La coupe sagittale présentée figure III.39 ne rapporte que les LH ayant une densité de probabilité supérieure à  $T = 2\%$ . Cette coupe laisse apercevoir que le total des densités de probabilité des LH menant à l'atome Chrom.66/OH n'est que de 19,8%. Cela indique que l'algorithme identifie des réseaux présents 2,88% du temps de la simulation qui impliquent des LH avec une densité de probabilité inférieure à 2%. Nous n'avons pas présenté dans ce manuscrit la coupe sagittale avec  $T = 0,5\%$  compte tenu de sa complexité.

**Les différents chemins de translocation :** Afin d'identifier les chemins de translocation effectifs entre Chrom.66/OH et Glu222/OEx, nous avons établi une liste de 9 chemins potentiels entre ces deux atomes dont les LH individuelles sont susceptibles d'être concomitantes au cours de la dynamique (cf. figure III.40). Ces chemins comprennent entre 1 et 3 atomes relais qui passent dans six cas sur neuf par le résidu Tyr145. Les molécules d'eau incluses dans ces chemins sont à la fois issues de la dynamique moléculaire (Wat2108, Wat3823) et de la structure cristallographique (Wat4, Wat42). La prise en compte de la rotation du groupement carboxyle du résidu Glu222 double le nombre de chemins de translocation potentiels à étudier. Ces hypothèses de travail additionnelles seront dénommées avec le suffixe bis.

- 1) E-222/OE1  $\Rightarrow$  W-42  $\Rightarrow$  Y-145  $\Rightarrow$  W-2108  $\Rightarrow$  Chrom-66/OH
- 2) E-222/OE1  $\Rightarrow$  W-42  $\Rightarrow$  Y-145  $\Rightarrow$  W-3823  $\Rightarrow$  Chrom-66/OH
- 3) E-222/OE1  $\Rightarrow$  W-42  $\Rightarrow$  Y-145  $\Rightarrow$  W-4  $\Rightarrow$  Chrom-66/OH
- 4) E-222/OE1  $\Rightarrow$  Y-145  $\Rightarrow$  W-2108  $\Rightarrow$  Chrom-66/OH
- 5) E-222/OE1  $\Rightarrow$  Y-145  $\Rightarrow$  W-3823  $\Rightarrow$  Chrom-66/OH
- 6) E-222/OE1  $\Rightarrow$  Y-145  $\Rightarrow$  W-4  $\Rightarrow$  Chrom-66/OH
- 7) E-222/OE1  $\Rightarrow$  W-42  $\Rightarrow$  W-4  $\Rightarrow$  Chrom-66/OH
- 8) E-222/OE1  $\Rightarrow$  W-4  $\Rightarrow$  Chrom-66/OH

FIGURE III.40 : Composition des chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx dans AvGFP-T203V-S205A (4OGS)

Les 18 chemins de translocation potentiels ont été étudiés. Leur densité de probabilité ainsi que les données statistiques sur leur longueur ont été regroupé table III.9. Parmi ces chemins, 5 sont exclus à cause de la non concomitance des LH qui les composent sur l'ensemble de la simulation. Le nombre de LH constituant les chemins dépend du chemin et varie de 2 à 4. Leur longueur varie entre 5,6 et 11,2 Å.

# chemin	Densité de probabilité (%)	$\langle l_{chemin} \rangle$ (Å)
1	6,60	11,17±0,26
1.bis	4,39	11,16±0,28
2	0,00	–
2.bis	0,00	–
3	0,00	–
3.bis	0,00	–
4	1,60	8,49±0,25
4.bis	2,03	8,51±0,24
5	0,05	8,58±0,32
5.bis	0,17	8,46±0,24
6	0,90	5,76±0,19
6.bis	2,16	5,69±0,21
7	0,48	8,34±0,21
7.bis	1,94	8,39±0,25
8	0,04	5,60±0,12
8.bis	0,00	–
9	0,50	5,67±0,20
9.bis	0,52	5,70±0,20

Tableau III.9 : Caractéristiques (densité de probabilité et longueur  $\langle l_{chemin} \rangle$ ) des 18 chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx dans AvGFP-T203V-S205A (4OGS).

La dynamique du chemin de translocation entre Chrom.66/OH et Glu222/OEx pour AvGFP-T203V-S205A est rendue complexe par la multiplication des chemins observés. Ces chemins peuvent être concomitants. Nous avons représenté la trajectoire de la longueur des 13 chemins restants au cours de la simulation (cf. figure III.41). Globalement, chaque chemin de translocation existe individuellement et pour des temps différents que les autres chemins observés, hormis pour les chemins de type 1 (1 et 1bis) et 4 (4, 4bis). Ces derniers coexistent pendant 0,9% du temps de la dynamique. De fait, il existe durant 20,5% du temps de simulation un chemin de translocation entre Chrom.66/OH et Glu222/OEx empruntant un des 13 chemins exposés ci-dessus. Cependant, cette probabilité est inférieure à la densité de probabilité de la liaison interne au Glu222 (22,7%). Ceci suggère qu'il existe encore d'autres chemins avec d'autres atomes relais que ceux que nous avons répertoriés permettant de connecter Chrom.66/OH et

Glu222/OEx, 2,2% du temps de simulation.

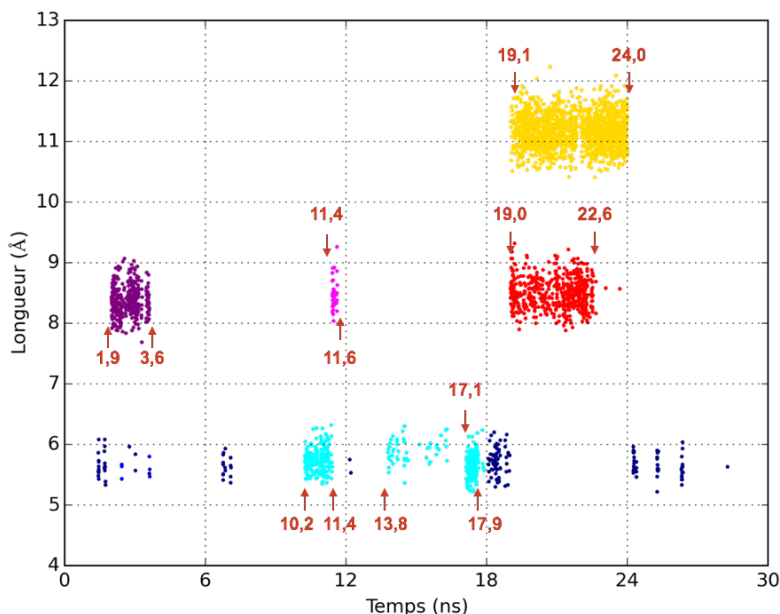


FIGURE III.41 : Trajectoire des longueurs des 7 types de chemin de translocation observés entre Chrom.66/OH et Glu222/OEx dans la dynamique de AvGFP-T203V-S205A (4OGS). La longueur des chemins en fonction du temps permet de les classer selon le nombre de LH qui les composent avec 2 LH, les types 6 (cyan), 8 (bleu), 9 (bleu foncé), 3 LH les types 4 (rouge), 5 (magenta), 7 (mauve) et 4 LH le type 1 (jaune). Les temps d'apparition des événements sont reportés en rouge (ns).

**Caractéristiques structurales des différents chemins de translocation :** Nous avons mené une analyse structurale des différents chemins de translocation observés. Nous les avons classés en fonction du nombre de LH constituant le chemin, ce qui revient à les classer en fonction de la longueur de ces chemins (cf. table III.10). Une légère augmentation de la longueur moyenne des LH qui composent ces chemins est constaté entre les chemins les plus longs et les chemins les plus courts. Les chemins les plus longs sont constituées de LH plus courtes et donc a priori plus forte. Ceci pourrait être corrélé avec la densité de probabilité de ces chemins qui est aussi plus importante. En effet, les chemins de type 1, les plus longs, sont majoritairement présents dans la dynamique de AvGFP-T203V-S205A et participe à hauteur de 51,9% à l'établissement d'un chemin de translocation entre Chrom.66/OH et Glu222/OEx. Ce type de chemin comprend le résidu Tyr145 dont la chaîne latérale est sujette à une réorientation ainsi que les molécules d'eau Wat42 et Wat2108 établissant des LH entre les atomes lourds du chemin.

Type de chemin	nb. LH / chemin	densité de probabilité %	ns	$\langle l_{chemin} \rangle$ (Å)
6, 8, 9	2	4,13	1,24	$5,68 \pm 0,18$
4, 5, 7	3	6,27	1,88	$8,46 \pm 0,25$
1	4	11,0	3,30	$11,17 \pm 0,27$

Tableau III.10 : Caractéristiques des 3 classes de chemins de translocation observés entre Chrom.66/OH et Glu222/OEx dans avGFP-T203V-S205A. Ces chemins sont classés selon le nombre de LH qui les constitue.

Le passage du type 1 au type 4 ou 5 est dû à la suppression d'une molécule d'eau (*i.e.* Wat42) au sein du chemin au profit d'une liaison hydrogène entre Glu222/OEx et Tyr145/OH (cf. figure III.42). Cette liaison est possible avec le rapprochement du groupement carboxyle de Glu222 vers l'hydroxyle de Tyr145.

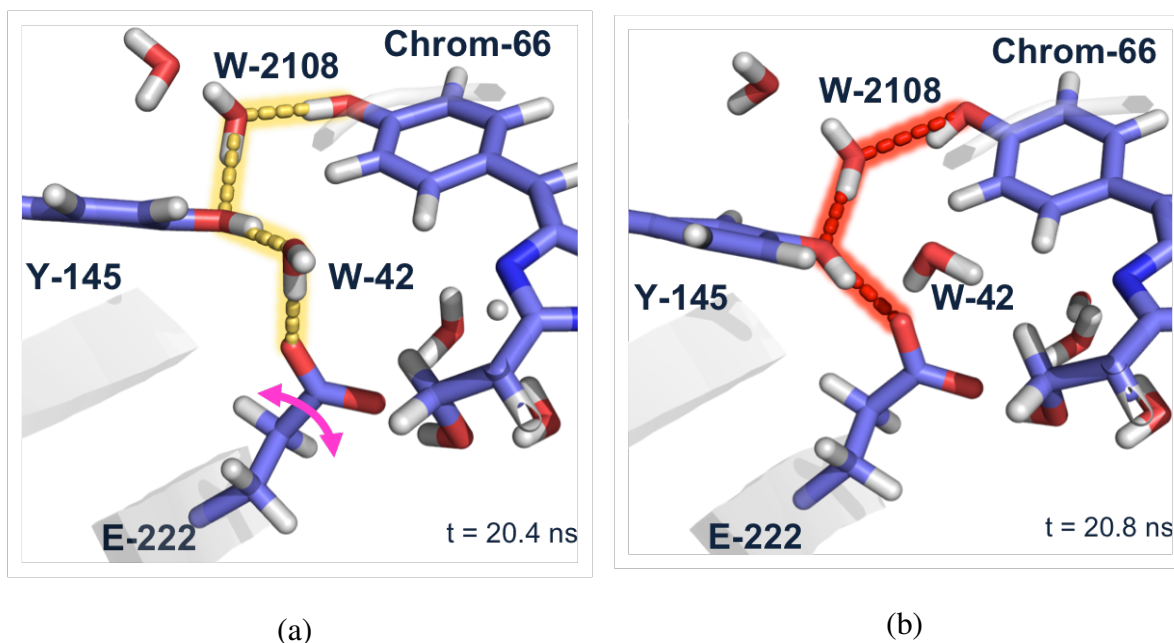


FIGURE III.42 : Passage entre les chemins de translocation, (a) de type 1 (en jaune) et (b) de type 4 ou 5 (en rouge type 4 uniquement) au cours de la dynamique de AvGFP-T203V-S205A.

Pour les chemins de type 1 et 4/5, l'intervention de la chaîne latérale de la Tyr145 s'effectue suite à une réorientation de la chaîne latérale de ce résidu (cf. figure III.43.a). Cette réorientation du cycle phénol de près de 90° déstabilise le brin  $\beta_7$  et appuie sur le résidu apolaire muté Ala205. Dans le cas des chemins de type 4 et 5, ceci permet une liaison hydrogène directe entre le phénol de Tyr145 et Glu222. Le chemin de type 5 est formé à partir du chemin 6 qui présentait une liaison hydrogène directe entre Tyr145/OH et Chrom.66/OH (cf. figure III.43.b). Cette liaison est rompue avec l'ajout d'une molécule d'eau issue de la dynamique. La molécule d'eau diffuse

vers l'intérieur du tonneau de AvGFP-T203V-S205A par la crevasse existant entre les brins  $\beta_7$  et  $\beta_{107}$ .

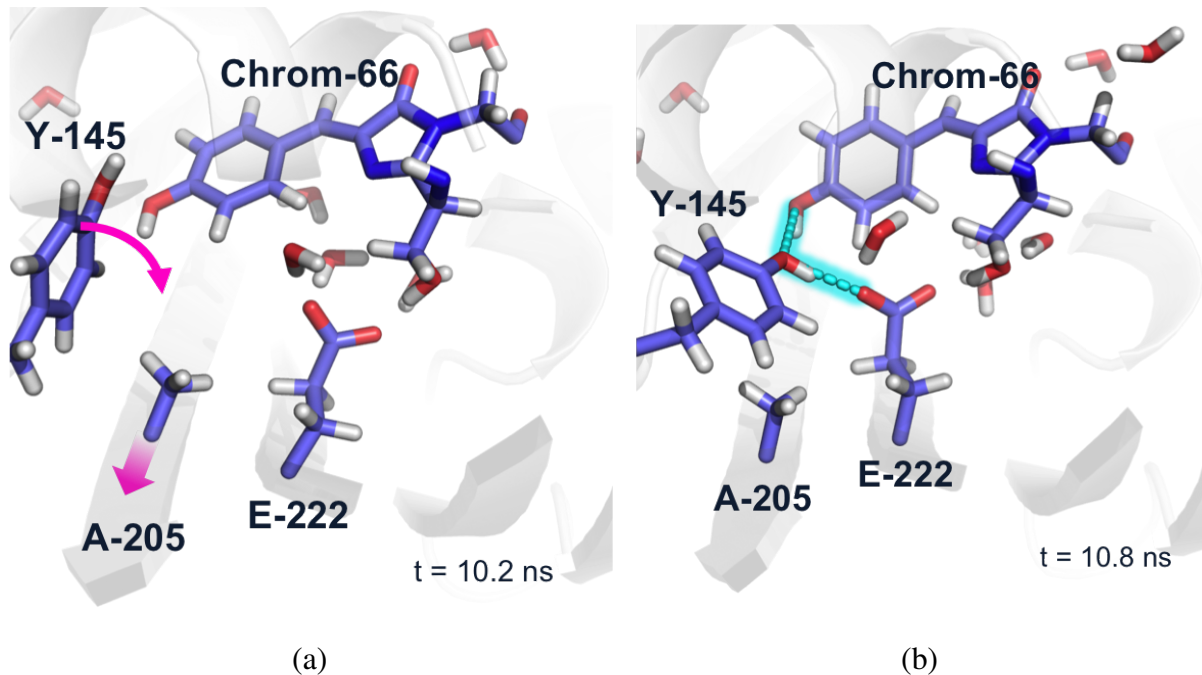


FIGURE III.43 : (a) Changement conformationnel de la chaîne latérale du résidu Tyr145 permettant l'établissement des chemins de type 1, 4/5 et 6 dans AvGFP-T203V-S205A. (b) Chemin de translocation de type 6 (en cyan).

**Caractéristiques des LH au sein des chemins de translocation** : Le chemin de translocation de type 1 est le chemin majoritaire. Il est composé de 3 atomes relais dont deux molécules d'eau (cf. figure III.44). Dans ce chemin, la densité de probabilité des différentes LH ne dépasse pas 13,7%, la liaison entre Glu222/OEx et Wat42/O étant la plus faible avec une densité de probabilité de 11,8%.



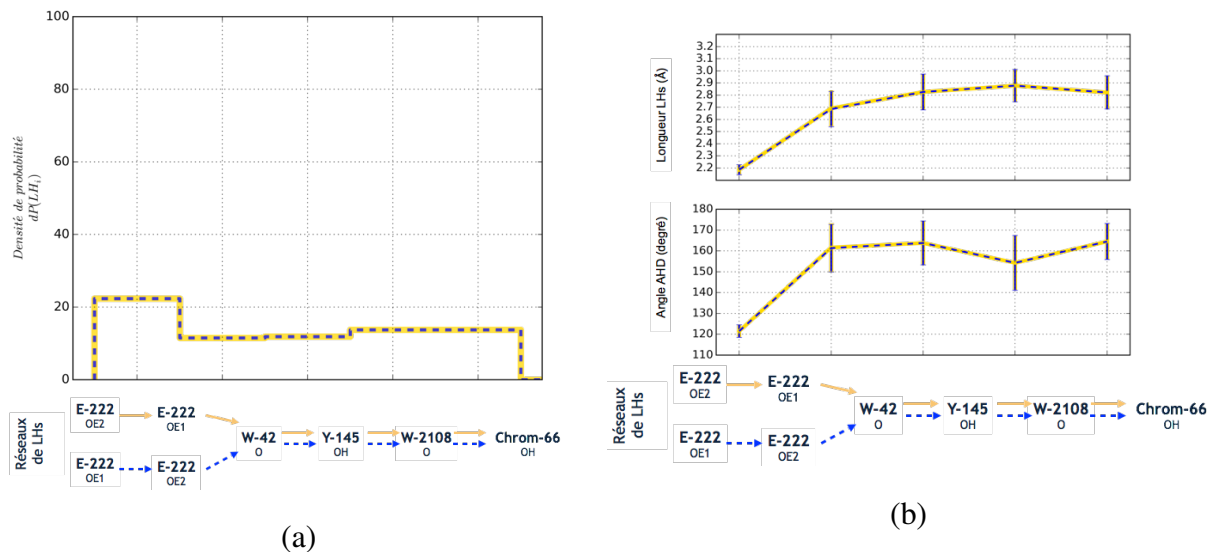


FIGURE III.44 : Densité de probabilité (a) et caractéristiques (distance  $d$  et angle  $\Omega$ ) (b) des LH intervenant dans le chemin de translocation de type 1 trouvé dans AvGFP-T203V-S205A. Le chemin de type 1 est représenté en jaune, celui de type 1.bis en bleu.

Les chemins de type 4, 5 et 7 sont composés d'un atome relais en moins par rapport au type 1. Dans le cas des types 4 et 5, il s'agit d'une molécule d'eau en moins tandis que dans le cas du type 7 c'est le résidu Ty145 qui ne participe plus au réseau. Ceci conduit à un raccourcissement de la longueur du chemin de translocation. Nous pouvons constater que les LH les moins fréquentes et donc celles qui constituent un point de faiblesse au niveau du chemin de translocation sont les liaisons localisées en amont du résidu Glu222.

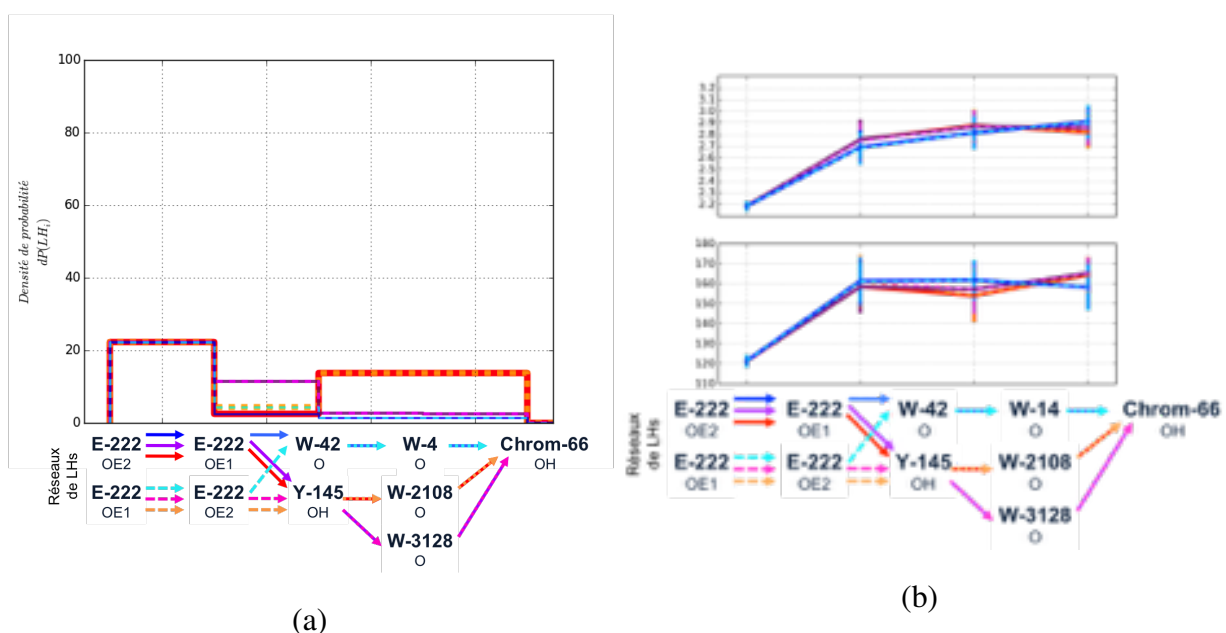


FIGURE III.45 : Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 4, 5 et 7 trouvés dans AvGFP-T203V-S205A. Les chemins de type 4 sont indiqués en rouge, 4.bis en orange, 5 en violet, 5. bis en magenta, 7 en bleu et 7.bis en cyan.

Les chemins de type 6, 8 et 9 sont constitués de 3 atomes, avec un seul relais. Ce sont les chemins de plus faible longueur. Ces réseaux présentent des LH de faible densité de probabilité, de 4%. Ici, la liaison hydrogène la plus faible est la liaison entre chrom.66/OH et son premier relais qui est soit Tyr145 soit une molécule d'eau (Wat14). Cette liaison est caractérisée par une longueur supérieure à 2.9 Å.

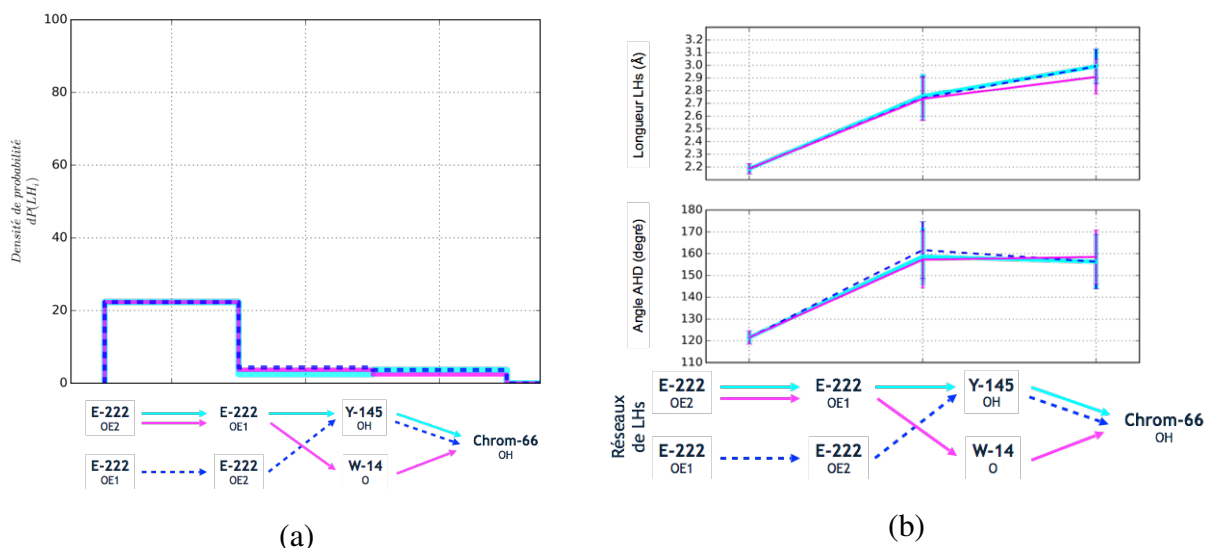


FIGURE III.46 : Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 6 et 8 trouvés dans AvGFP-T203V-S205A. Le chemin de type 6 est représenté en cyan, le 6.bis en bleu pointillé et le 8 en magenta.

#### **D.2.d AvGFP-T203V-S205V (4OGS-S205V)**

##### **D.2.d i Réseau global de liaisons hydrogène**

Expérimentalement, le remplacement de l'alanine en position 205 par une valine dans AvGFP-T203V-S205A conduit à une annihilation du transfert de proton à l'état excité. La coupe sagittale issue de l'analyse de la dynamique moléculaire de AvGFP-T203V-S205V fait apparaître un nombre faible de liaisons hydrogène. En effet, seulement 1872 liaisons hydrogène sont répertoriées au cours de la dynamique pour le réseau global de liaisons hydrogène connectées au phénol du chromophore. Dans AvGFP-T203V-S205A, il y en avait 70863. Ces liaisons sont de plus caractérisées par une très faible densité de probabilité qui n'excède pas 1%.

Sur la base de la coupe sagittale, un seul chemin de translocation entre Chrom.66/OH et Glu222/OEx est identifiable avec 2 atomes relais : il s'agit de deux molécules d'eau Wat4/O et Wat41/O. Compte tenu de la rotation fréquente du groupement carboxyle de Glu222 au cours de la dynamique, l'atome accepteur final est soit Glu222/OE2, soit Glu222/OE1. De plus, nous retrouvons en partie les chemins de translocation impliquant notamment l'hydroxyle en position 65 du chromophore. Dans le cas de cette protéine, seul le chemin se prolongeant jusqu'à Gln69 a été détecté, faisant intervenir 2 molécules d'eau.

Enfin, la connexion entre le phénol du chromophore et le solvant s'effectue au cours de 29 instantanés soit 0,19% du temps de simulation.

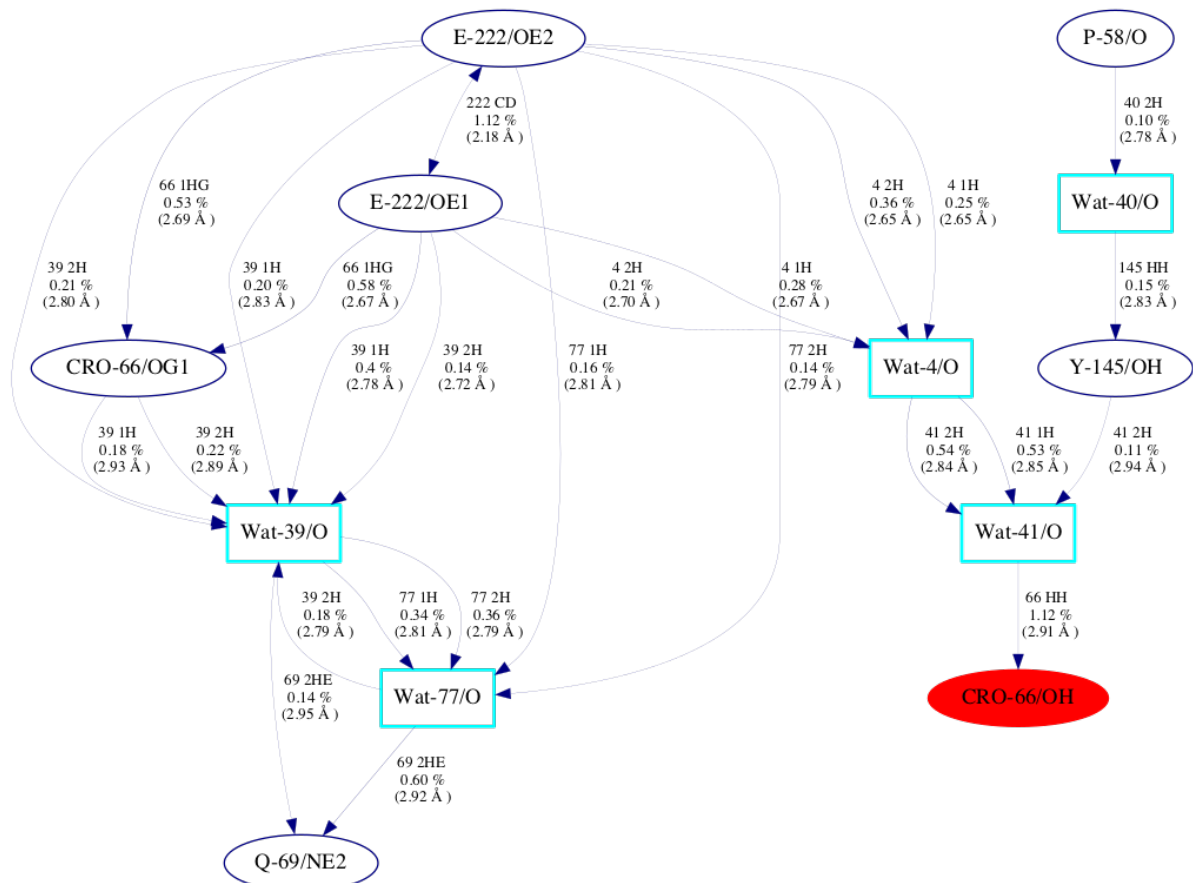


FIGURE III.47 : Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-T203V-S205V. Seules les LH dont la densité de probabilité est supérieure à  $T = 0,1\%$  sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge.

#### D.2.d ii Chemin de translocation de protons impliquant le phénol du chromophore

La concomitance des liaisons hydrogène du chemin de translocation entre Chrom.66/OH et Glu222/OEx est détectée pour 162 instantanés soit 1,08% du temps de la dynamique ce qui correspond à 324 ps sur une dynamique de 30 ns. On observe une légère différence de 0,04% avec la densité de probabilité interne au Glu222. Ceci suggère que l'algorithme détecte d'autres liaisons hydrogène, de faibles densité de probabilité et non répertoriées sur la coupe sagittale, participent à l'établissement d'un chemin de translocation connectant Chrom.66/OH et Glu222. Ce chemin s'établit au sein de la protéine au cours de 280 ps entre les temps 4,95 et 5,23 ns où il est repéré transitoirement. Sa longueur est de  $8,42 \pm 0,22 \text{ \AA}$  avec une longueur moyenne des liaisons hydrogène qui le compose égale à  $2,82 \pm 0,08 \text{ \AA}$ . Parmi les liaisons hydrogène de ce chemin, celle entre Glu222/OE2 et Wat41 a une longueur moyenne de  $2,66 \pm 0,11 \text{ \AA}$  soit la plus courte du chemin. Cette liaison hydrogène est aussi la moins probable au cours de la

dynamique. A noter que la longueur de liaison hydrogène est décroissante à mesure que le chemin de translocation se rapproche de Glu222, accepteur final de proton. Compte tenu de l'échange fréquent entre les atomes OE1 et OE2 de Glu222 au cours de la dynamique le chemin est fréquemment interrompu ce qui peut expliquer en partie sa très faible probabilité.

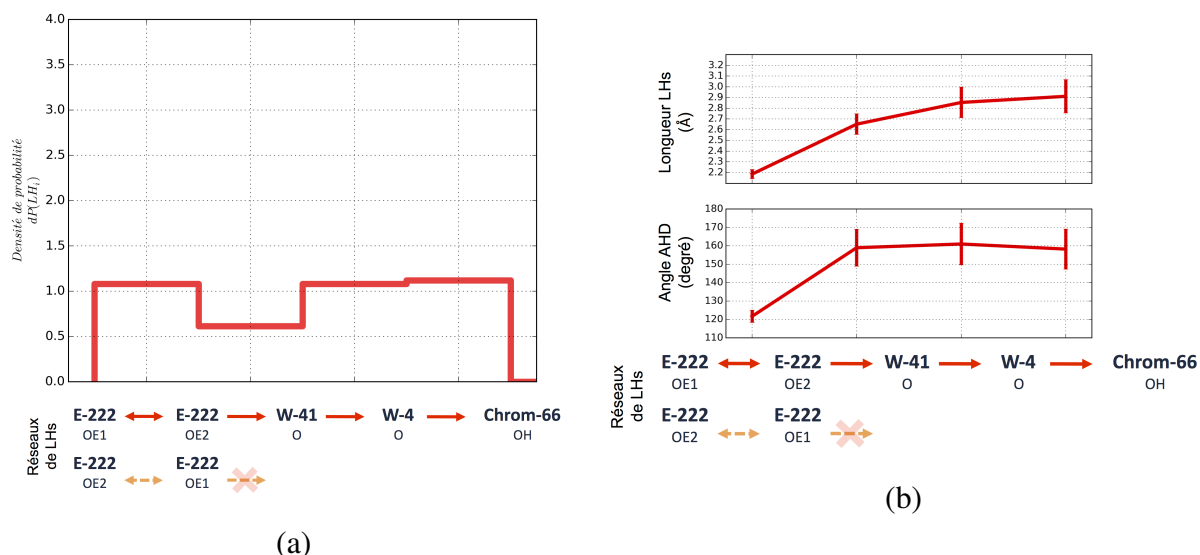


FIGURE III.48 : Densité de probabilité (a) et caractéristiques (distance et angle) (b) des LH intervenant dans les chemins de translocation de type 6 et 8 trouvés dans AvGFP-T203V-S205V.

La mutation S205V remplace un résidu de faible volume stérique par un résidu valine avec deux groupements méthyles supplémentaires. Cet encombrement stérique dans l'environnement direct de Tyr145 inhibe le réarrangement conformationnel subit par sa chaîne latérale observé précédemment dans le cas de AvGFP-T203V-S205A. La rotation de la chaîne latérale de Tyr145 est bloquée par deux valines, résidus apolaires, en position 58 et 205.

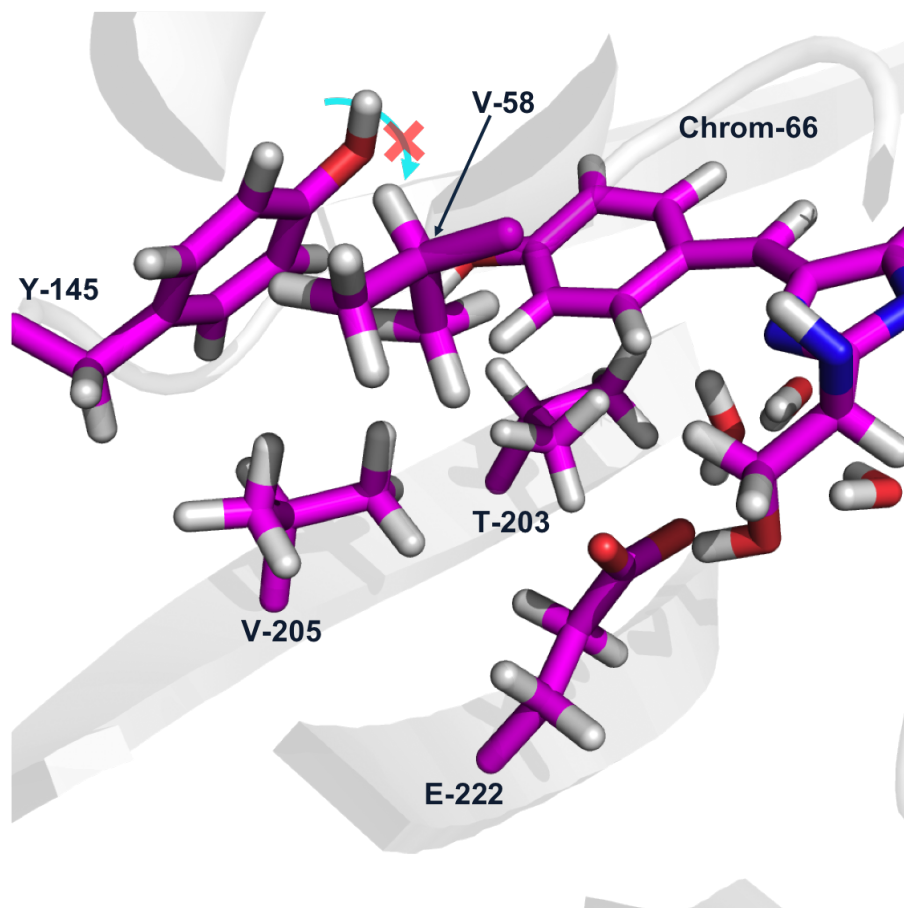


FIGURE III.49 : Mécanisme possible de blocage de la chaîne latérale de Tyr145 (flèche cyan) par les résidus apolaires Val58 et Val205 en présence d'une valine en position 205 dans AvGFP-T203V-S205V.

### D.2.e AvGFP-S65T (1C4F)

#### D.2.e i Réseau global de liaisons hydrogène

L'ajout d'un méthyle en position 65, remplaçant une sérine par une thréonine, conduit à la suppression du transfert de proton à l'état excité observé dans AvGFP [Stoner-Ma et al., 2008]. Cette mutation conduit à un changement conformationnel au niveau de ce résidu : l'hydroxyle adopte principalement une conformation 65OH-*up* alors que dans AvGFP, sa conformation est 65OH-*down*. L'analyse de la dynamique moléculaire menée sur la base de la structure cristallographique 1C4F de AvGFP-S65T obtenue à pH 4,6 rapporte un réseau global de LH connecté au phénol du chromophore constitué de 201520 LH. La représentation sagittale montre que Glu222 reste encore l'accepteur de proton final de divers chemins dont les donneurs de protons initiaux sont les mêmes que ceux observés notamment dans AvGFP, AvGFP-S205V ou encore AvGFP-T203V-S205A. Il s'agit des atomes Val68/N, Gln69/NE2, His148/ND1 et Chrom.66/OH. Nous constatons de plus que suite au changement conformationnel de l'hydroxyle en position 65 induit par la mutation S65T, cet hydroxyle participe moins à l'établissement des réseaux de LH que dans AvGFP. Il n'est plus en connexion avec les molécules d'eau se trouvant à proximité du chromophore qu'environ 9% du temps de simulation alors que dans AvGFP, il l'était plus de 80% du temps. Malgré ce changement conformationnel, l'hydroxyle en position 65 reste toujours en liaisons hydrogène avec Glu222/OEx, avec une densité de probabilité d'au minimum 58%. La représentation sagittale suggère que huit molécules d'eau sont impliquées dans le réseau global dont sept issues des structures cristallographiques. Toutes les molécules d'eau cristallographiques évoluent dans la poche de molécules d'eau se trouvant à proximité du chromophore près de l'imidazolinone, à l'exception de Wat14. Enfin, le phénol du chromophore est connecté 16,3% du temps de la dynamique à une molécule d'eau d'une sous-boîte de la première sphère externe (solvant) par liaisons hydrogène.

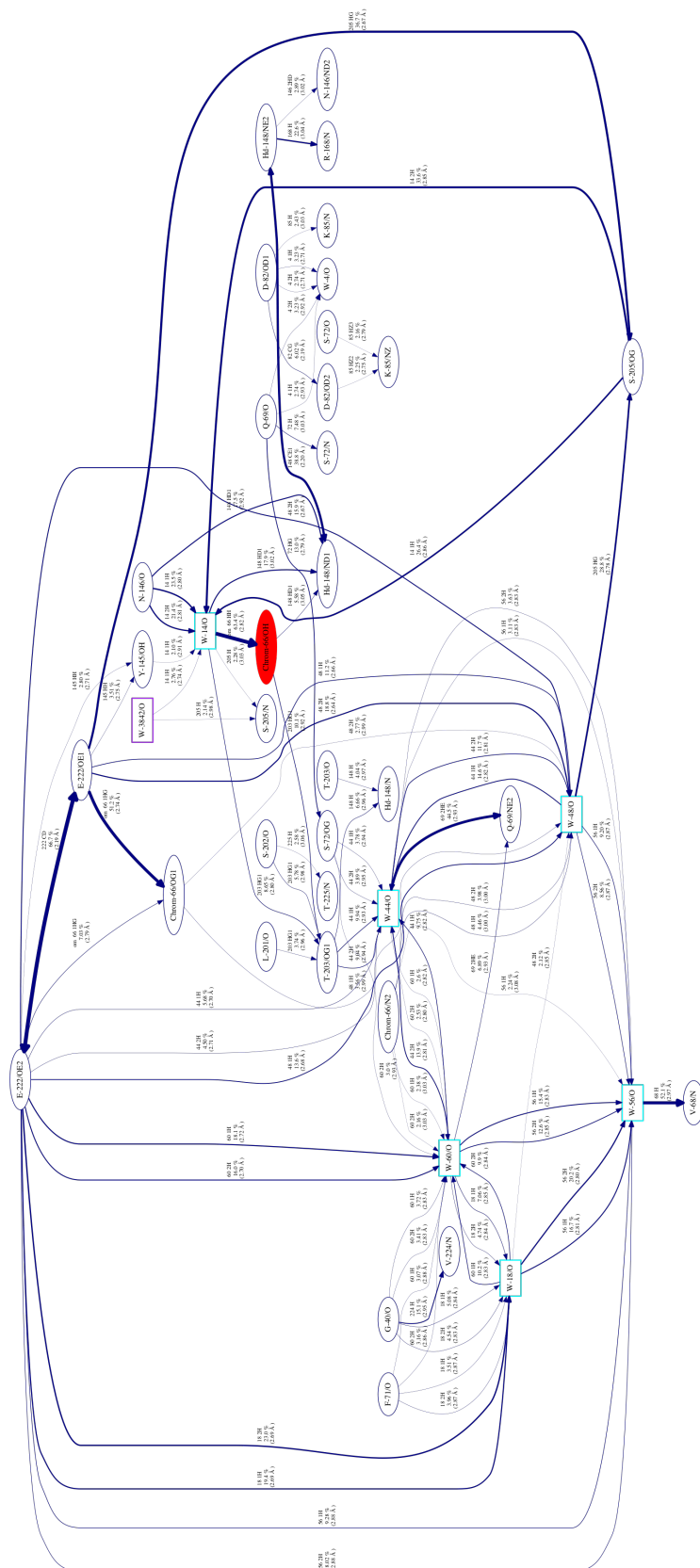


FIGURE III.50 : Coupe sagittale présentant le réseau global de liaisons hydrogène connecté à l'hydroxyle du phénol du chromophore dans le cas de AvGFP-S65T (1C4F). Seules les LH dont la densité de probabilité est supérieure à  $T = 2\%$  sont représentées. Les molécules d'eau d'origine cristallographiques sont représentées en cyan tandis que celles issues de la dynamique sont en violet. Le point de départ de la recherche des réseaux est représenté en rouge.



### D.2.e ii Chemin de translocation de protons impliquant le phénol du chromophore

Les données recueillies à partir de la coupe sagittale indiquent que des réseaux de LH s'établissent 66,7% du temps total de la dynamique si l'on se réfère à la liaison intramoléculaire Glu222/OE<sub>x</sub> (cf. figure III.50).

**Les différents chemins de translocation** : cette représentation sagittale permet d'établir une liste de quatre chemins dont les liaisons hydrogène sont susceptibles d'être concomitantes au cours de la simulation :

- 1) E-222/OE1 → S-205/OG → W-14 → Chrom-66/OH
- 2) E-222/OE1 → Y-145/OH → W-14 → Chrom-66/OH
- 3) E-222/OE1 → W-48 → S-205/OG1 → W-14 → Chrom-66/OH
- 4) E-222/OE1 → Chrom-66/OG1 → W-48 → S-205/OG1 → W-14 → Chrom-66/OH

FIGURE III.51 : Composition des chemins de translocation potentiels entre Chrom.66/OH et Glu222/OE<sub>x</sub> dans AvGFP-S65T (1C4F)

Compte tenu de la rotation du groupe carboxyle de Glu-222, nous avons au final 8 chemins possibles. Les chemins ayant Glu222/OE<sub>2</sub> comme accepteur final seront dénommés avec le suffixe bis. Le nombre d'atomes relais au sein de ces chemins n'est pas équivalent. On dénombre ainsi 2 atomes relais pour les chemins de type 1 et 2, 3 atomes relais pour les chemins de type 3 et enfin 4 atomes relais pour les chemins de type 4. Dans les chemins de type 3 et 4, 2 molécules d'eau cristallographiques font parties des relais.

Les huit chemins ont été analysés afin de déterminer leur densité de probabilité au cours de la dynamique (cf. table III.11). Ces données montrent que les chemins de translocation 1 et 3 ont des densités de probabilité d'existence supérieure à 25% du temps de dynamique. A l'inverse, les chemins 1.bis, 3.bis et 4bis ayant l'atome Glu222/OE<sub>2</sub> comme accepteur final n'existent pas au cours de la dynamique. Enfin, les chemins 2, 2.bis et 4 ont des densités de probabilité inférieures à 5%.

# chemin	Densité de probabilité (%)	$\langle l_{chemin} \rangle$ (Å)
1	29,46	8,35±0,25
1.bis	$6,67 \times 10^{-3}$	9,00±0,00
2	2,43	8,57±0,26
2.bis	1,43	8,54±0,24
3	28,67	11,11±0,27
3.bis	0,00	–
4	4,97	14,24±0,30
4.bis	$6,67 \times 10^{-3}$	114,38±0,00

Tableau III.11 : Caractéristiques (densité de probabilité et longueur  $\langle l_{chemin} \rangle$ ) des huit chemins de translocation potentiels entre Chrom.66/OH et Glu222/OEx et de leur densité de probabilité dans AvGFP-S65T (1C4F). Les chemins 1 bis et 4 bis peuvent être considérés comme inexistantes.

L'analyse des 5 chemins restants indique que les chemins 2 et 2.bis sont simultanés 1,42% du temps de dynamique (*i.e.* 213 instantanés). Le chemin 4 quant à lui existe en même temps que le chemin 3. Cette concomitance s'effectue durant 4,97% (*i.e.* 746 instantanés) du temps de simulation.. Ainsi, la densité de probabilité d'existence d'un chemin entre Chrom.66/OH et Glu222/OEx passant par un des 5 chemins décrits ci-dessus est de 60,57%. Cette densité est inférieure à la densité de probabilité de la liaison interne de Glu222 (66,70%) laissant supposer qu'il existe d'autres chemins entre Chrom.66/OH et Glu222/OEx que ceux identifiés ici, constitués de LH de densité de probabilité inférieure à 2%.

Nous avons analysé la suite des évènements lors de l'établissement d'un chemin de translocation entre Chrom.66/OH et Glu222/OEx. La figure III.52 montre que le chemin 1 présentant 2 atomes relais est suppléé 1 fois par les chemins de type 2 et 2 fois par le chemin 3. Ce dernier présente un atome relai supplémentaire, une molécule d'eau venant s'insérer entre Ser205/OG1 et l'accepteur final Glu222/OE1. Ce chemin 3 apparaît 546 ps après l'interruption du chemin 1 dans le premier cas et directement dans le second cas. Le passage du chemin 1 à 2 implique un échange entre l'hydroxyle de Ser205 et celui de Thr203.

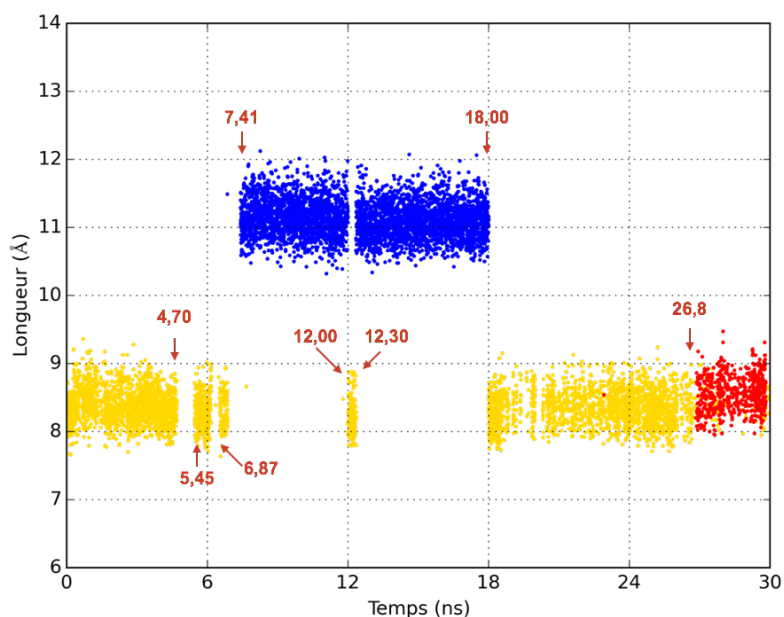


FIGURE III.52 : Trajectoire des longueurs des 3 types de chemins de translocation observés entre Chrom66/OH et Glu222/OEx dans la dynamique de AvGFP-S65T (1C4F). Le chemins 1 est indiqué en jaune, les chemins de type 2 en rouge (Glu222/OE2 accepteur final) et magenta (Glu222/OE1 accepteur final) et le chemin de type 3 en bleu. Les temps d'apparition des événements sont reportés en rouge (ns).

**Caractéristiques structurales des différents chemins de translocation** : nous constatons que le passage du chemin 1 au chemin 3 est directement corrélé à la rotation de la chaîne latérale de Ser205, avec dans un premier temps une orientation de l'hydroxyle vers Glu222/OE1. La rotation de l'hydroxyle est concomitante avec la rotation des chaînes latérales des résidus Thr203 et Leu207 permettant une libre rotation de la chaîne latérale de Ser205. Cette rotation qui rompt la liaison hydrogène avec Glu222/OE1 offre un espace suffisant pour l'insertion entre ces deux résidus de la molécule d'eau Wat48, située initialement sous le chromophore (cf. figure III.53).

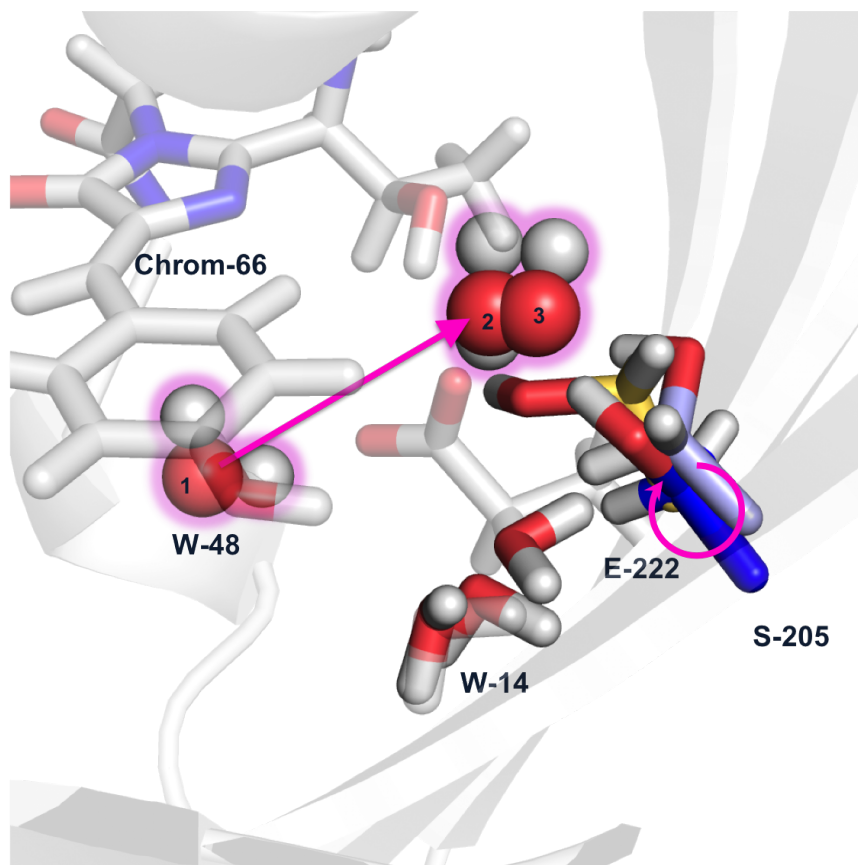


FIGURE III.53 : Mécanisme de passage des chemins 1 à 3 avec le sens de déplacement de Wat48 numéroté et fléché ainsi que la rotation de Ser205. Instantané pris à 6,2 (doré), 7,2 (bleu léger) et 8 ns (bleu intense).

La commutation du chemin 1 vers le chemin 2 apparaît vers 26,8 ns de la dynamique et correspond à une phase durant laquelle l'hydroxyle de Ser205 présente une dynamique importante. Cette commutation implique la substitution de l'hydroxyle de la Ser205 avec celui du phénol de Tyr145. Or, ce dernier est majoritairement à une distance supérieure à 3,2 Å de la molécule d'eau Wat14 avec laquelle il établira finalement une liaison hydrogène. Une analyse des structures à différents instantanés suggère un déplacement égale ou supérieur à 1,5 Å pour les résidus Ser205 et Ala206. La séquence des événements présentée figure III.54 montre que le retrait en direction du solvant de la chaîne carbonée de Ser205 précède la rotation de la chaîne latérale de Tyr145. Cette rotation en direction de Glu222 est complétée par une rotation du proton de Tyr145 afin d'établir une liaison hydrogène avec Glu222. Ainsi, le mécanisme de passage entre les chemins 1 à 2 suggéré par la dynamique est un réarrangement conformationnel de Tyr145 aidé par le retrait vers le solvant des résidus Ser205 et Ala206 tandis que le mécanisme de passage entre les chemins 1 à 3 serait dû à l'accommodation de la molécule d'eau Wat48

suite au mouvement de l'hydroxyle de Ser205.

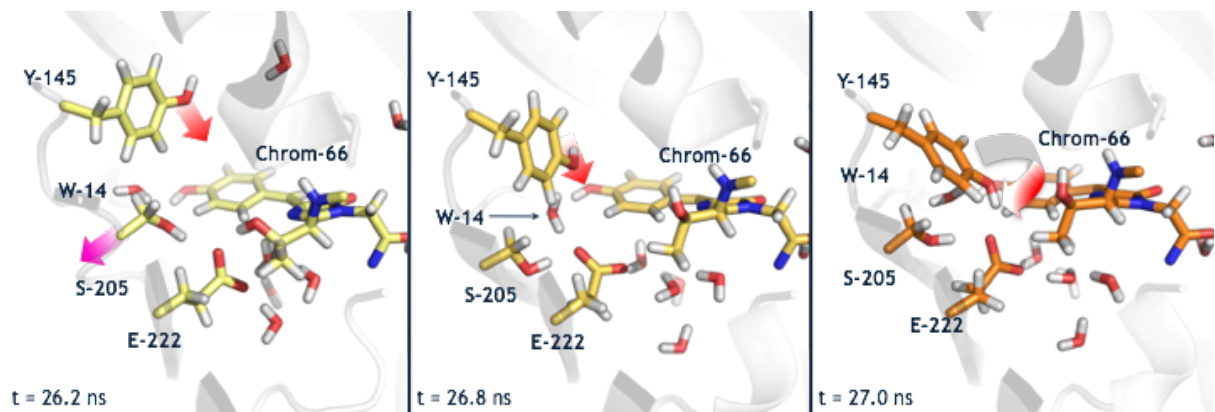
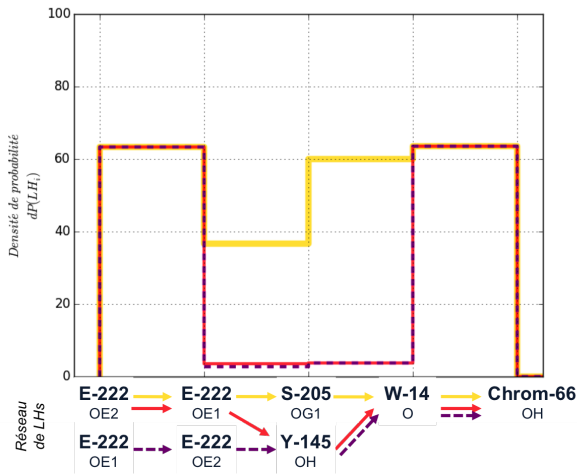
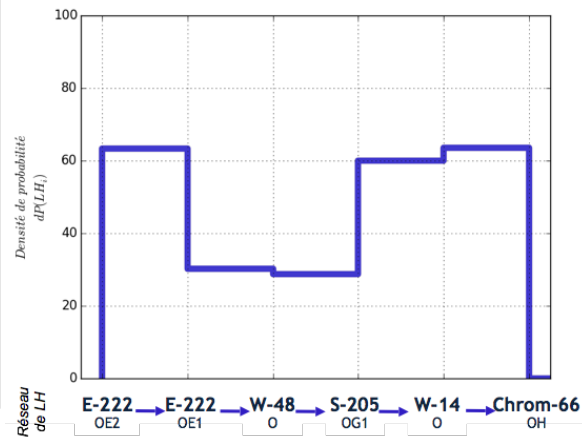


FIGURE III.54 : Mécanisme de passage des chemins 1 à 2 avec le sens de déplacement de Ser205 (magenta) et le mécanisme de rotation de Tyr145 (rouge).

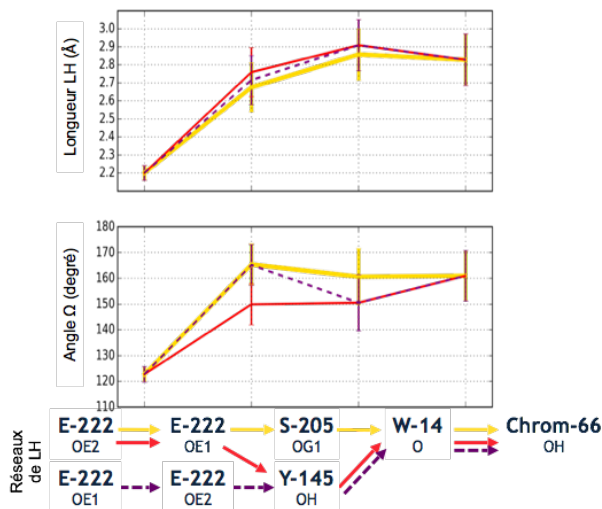
**Caractéristiques des liaisons hydrogène au sein des chemins de translocation** : l'analyse des densités de probabilité des LH des divers chemins de translocation observés dans AvGFP-S65T montre que ces chemins présentent une faiblesse principalement située sur la liaison hydrogène reliant Glu222/OEx et son atome relai partenaire. Cette liaison hydrogène est invariablement la plus courte des liaisons présentes dans les différents chemins. On note que l'angle  $\Omega$  des LH varie significativement dans les chemins de type 2 (rouge et violet sur la figure III.55) suggérant que l'établissement de la LH avec Tyr145/OH présente des contraintes.



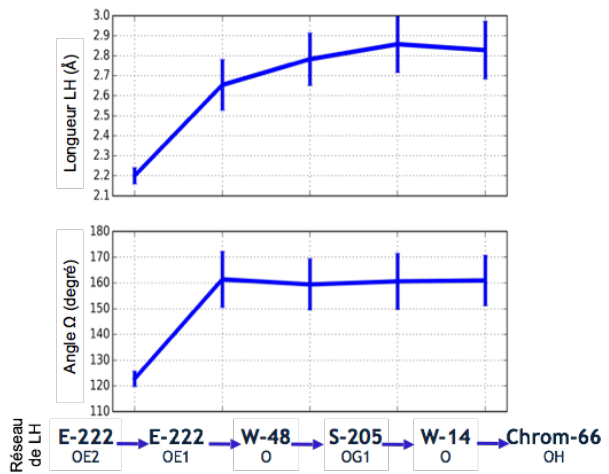
(a)



(b)



(c)



(d)

FIGURE III.55 : (a/b) Densité de probabilité et (c/d) caractéristiques (longueur  $d$  et angle  $\Omega$ ) des LH intervenant dans les chemins de translocation trouvés dans AvGFP-S65T. Le chemin 1 est représenté en jaune, le chemin 2 en rouge, le chemin 2 bis en violet et le chemin 3 en bleu.

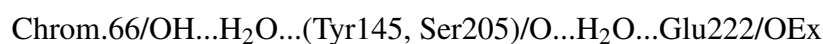
## E Discussion

Grâce à l'algorithme de recherche et d'agrégation des LH au sein d'une protéine mis au point au cours de ces travaux, nous avons identifié le réseau global de LH connectées au phénol du chromophore pour chacune des 5 protéines fluorescentes étudiées. Au sein de ce réseau, un ou plusieurs chemins de translocations entre le phénol et la chaîne latérale du glutamate 222, accepteur de proton final, ont été repérés.

Nous retrouvons les chemins proposés dans la littérature pour sous-tendre l'ESPT pour les protéines AvGFP, AvGFP-S205V et AvGFP-T203V-S205A. Dans le cas de AvGFP, le chemin de translocation de protons proposé dans la littérature est également le chemin prépondérant trouvé par l'algorithme. Il s'établit entre le phénol du chromophore et l'atome OE2 de Glu222, en impliquant deux relais, l'hydroxyle de Ser205 et une molécule d'eau cristallographique. Suite à la mutation de la Ser205, l'algorithme détecte un nouveau relais dans AvGFP-S205V. Il s'agit de l'hydroxyle de Thr203. Le chemin de translocation ainsi généré est également celui proposé dans la littérature sur la base de la structure de ce mutant [Shu et al., 2007] et étayé par des études de dynamique moléculaire [Simkovitch et al., 2013]. Celles-ci mettent en évidence l'échange de la molécule d'eau cristallographique avec une molécule du solvant au cours de la dynamique. Cet échange s'effectue 16 fois sur 60 ns de simulation. De façon similaire, notre algorithme fait apparaître cet échange. Notre approche a permis de quantifier la densité de probabilité de ce chemin alternatif à 2,10%. Dans le cas de AvGFP-T203V-S205A, l'algorithme met en évidence différents chemins de translocation. Parmi ceux-ci, certains utilisent uniquement des molécules d'eau comme relais. Ces molécules proviennent de la structure et sont celles qui se situent à proximité du groupe imidazolinone du chromophore. Ce mécanisme est proche de celui présenté par Wineman-Fisher et coll. [Wineman-Fisher et al., 2014]. Ces derniers proposent, en effet, sur la base d'une étude de dynamique moléculaire, l'existence d'un chemin impliquant comme relais une molécule cristallographique et une autre provenant du solvant et s'insérant occasionnellement à l'intérieur du tonneau. Grâce au suivi par l'algorithme de la dynamique des LH, nous estimons la densité de probabilité de l'ensemble des chemins impliquant uniquement soit une soit deux molécules d'eau à 0,54% et 2,42% respectivement.

Par ailleurs, sur l'ensemble des 5 protéines étudiées, les chemins de translocation détectés par l'algorithme présentent une topologie différente en termes de nombre de relais impliqués. De 1 à 3 relais sont nécessaires pour former un chemin de translocation entre Chrom.66/OH et Glu222/OEx. Pour autant, certains relais sont similaires d'une protéine à une autre et *in fine*, le nombre d'atomes relais distincts, autres que des molécules d'eau, participant aux chemins dans les diverses protéines, est restreint (cf. figure III.56). Il s'agit des groupes hydroxyle de Ser205, de Thr203 et de Tyr145. Quatre typologies distinctes de chemin de translocation ont été détectées. La première classe de chemin présente une typologie similaire au chemin de translocation usuel proposé dans la littérature. Ce type de chemin connecte, par LH, Chrom.66/OH à une molécule d'eau puis à un atome d'oxygène d'un des trois résidus Tyr145, Thr203 ou Ser205 et enfin Glu222/OEx. Cette typologie a été observée dans toutes les protéines à l'exception de AvGFP-T203V-S205V. Ces chemins ont des longueurs d'environ 8 à 10 Å et le chemin de plus forte densité de probabilité est le chemin usuel observé dans AvGFP empruntant Ser205 (91,8%). Il s'agit également du chemin le plus court (~ 8,3 Å) parmi tous ceux de cette typologie. Pour toutes les protéines présentant cette typologie de chemin, nous constatons que la liaison hydrogène la plus faible en termes de densité de probabilité d'existence est celle entre Glu222/OEx et l'hydroxyle d'un des trois résidus Ser205 ou Tyr145 ou Thr203, alors que la distance entre atomes lourds de cette liaison est la plus courte du chemin. Une étude plus poussée de l'ancrage de ces deux résidus au sein de la protéine par leurs résidus voisins permettrait de proposer des pistes sur les facteurs structuraux et dynamiques à l'origine de cette faiblesse.

Une seconde classe de chemin incorpore une molécule d'eau supplémentaire entre l'hydroxyle de Tyr145 ou de Ser205 et Glu222/OEx selon le schéma :

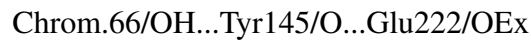


Ce type de chemin est observé dans AvGFP-T65S et AvGFP-T203V-S205A avec des densités de probabilité de 33% et 11% respectivement. Ces chemins sont caractérisés par une longueur de 13,3 Å. Dans ce type de chemin, les liaisons présentant la plus faible densité de probabilité restent entre Glu222/OEx et la molécule d'eau et entre cette molécule d'eau et l'hydroxyle de Ser205 ou Ty145.

Une troisième classe de chemin de translocation n'utilise pas de molécules d'eau comme

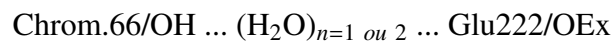


relais mais uniquement un résidu. Ce type de chemin est observé uniquement dans avGFP-T203V-S205A et présente une densité de probabilité très faible, inférieure à 1% :



Cette faible probabilité peut provenir des tensions stériques importantes qu'implique une telle typologie.

Enfin, certains chemins transitent uniquement par une ou 2 molécules d'eau.



Ils ont été observés dans AvGFP-T203V-S205A et dans AvGFP-T203V-S205V avec des densités de probabilité très faibles, inférieures à 2%. La dynamique des molécules d'eau dans la protéine peut être un des facteurs à l'origine du caractère transitoire et peu fréquent de ces chemins.

L'algorithme met en évidence un nouveau relais, à savoir l'hydroxyle de Tyr145. Ce relais n'a pas été rapporté dans la littérature. Ce relais est utilisé lorsque la dynamique de la chaîne latérale de Tyr145 n'est pas bridée. En effet, les dynamiques moléculaires menées sur les différentes protéines indiquent que le passage des chemins de translocation par Tyr145/OH nécessite une rotation de la chaîne latérale qui requiert de l'espace et de fait ne peut pas être observée pour les mutants ayant un résidu valine en position 205.

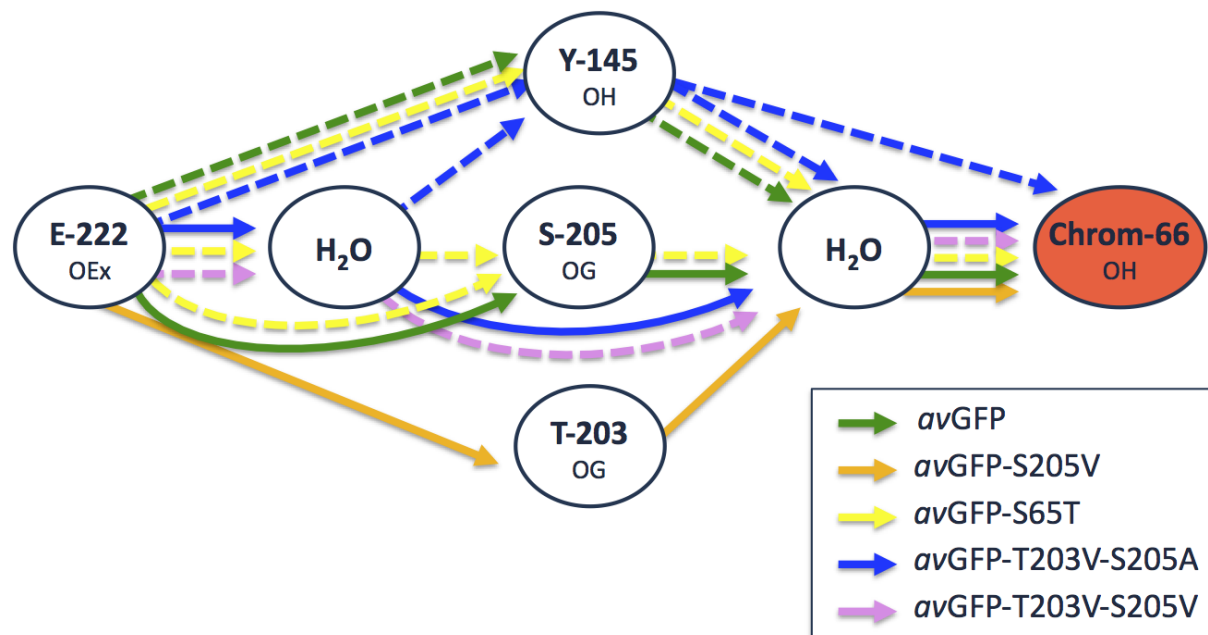


FIGURE III.56 : Schéma synoptique des divers chemins de translocation trouvés par l'algorithme pour AvGFP et ses mutants (chemins de la littérature en trait plein et ceux mis en évidence par l'algorithme en trait pointillé).

Nous avons mis en regard les densités de probabilité totales des chemins de translocation rapportées par l'algorithme et les vitesses relatives de l'ESPT pour chaque protéine (cf. figure III.57). Nous obtenons un accord raisonnable entre nos résultats et ceux expérimentaux rapportés dans la littérature. D'une manière générale, les protéines caractérisées par une vitesse d'ESPT faible sont aussi celles pour lesquelles l'algorithme rapporte des chemins de translocation à partir du phénol du chromophore de faibles densités de probabilité, à l'exception de AvGFP-S65T. Les décalages observés peuvent résulter du fait que la démarche développée ici vise à rechercher uniquement les chemins physiques potentiels susceptibles de supporter une translocation de protons. Elle ne tient pas compte de la probabilité de transfert de proton d'un donneur de proton initial vers un accepteur final. Le cas de AvGFP-S65T est particulier, car l'analyse par l'algorithme de la dynamique moléculaire effectuée en partant de la structure cristallographique 1C4F identifie des chemins de translocation avec une densité de probabilité totale de plus de 60%. Ceci diverge des résultats expérimentaux qui suggèrent l'absence d'ESPT dans cette protéine. Plusieurs hypothèses pourraient expliquer l'origine d'une telle divergence : une mauvaise définition des états de protonation et/ou de la conformation locale de certains résidus sur la structure de départ ou encore une mauvaise appréciation du pH de cristallisation conduisant à un choix des états de protonation inapproprié.

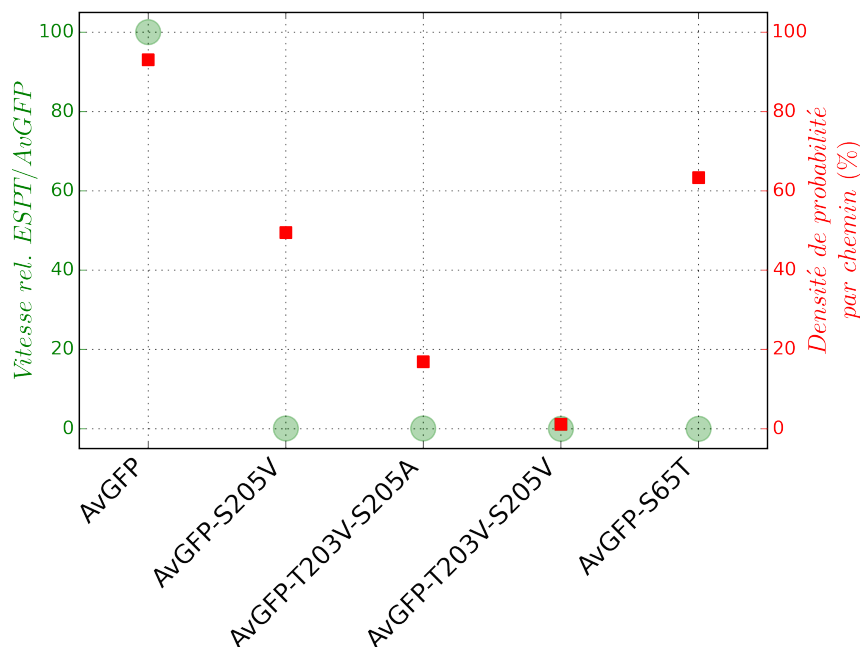


FIGURE III.57 : Valeur de  $k_{PT}$  relative à AvGFP et densité de probabilité des chemins de translocation (Chrom-66/OH à Glu222/OEx) pour chacune des dynamiques étudiées.

## F Conclusion

En somme, la cohérence des résultats obtenus au cours de ces travaux avec ceux rapportés dans la littérature nous confortent dans la capacité de l'algorithme à identifier les réseaux de LH et les chemins de translocation au sein de la protéine et à en dégager des informations sur leur dynamique. En effet, même si nous retrouvons des chemins similaires à ceux de la littérature obtenus sur la base de l'analyse de la structure cristallographique couplée ou non à des simulations de dynamique moléculaire, notre approche offre l'avantage d'une analyse d'ensemble des LH en réseaux. A cela s'ajoute la faculté d'identifier les chemins de translocation à savoir l'ensemble des LH pouvant supporter un transfert de proton d'un donneur initial à un accepteur final et de quantifier leur probabilité d'existence. L'analyse détaillée des différents chemins offre également la possibilité de pointer les liaisons les plus faibles qui, suivant le contexte de l'étude menée, pourraient être des sites de mutation privilégiés. Par la suite, il serait intéressant d'avoir en supplément une vue d'ensemble de la densité de probabilité totale de chaque liaison hydrogène individuelle constituant un réseau global. L'algorithme fournit pour le moment la probabilité d'existence de ces liaisons quand elles sont connectées à un atome d'intérêt. Ceci permettrait de recueillir des informations de l'incidence des mutations sur la stabilité des différentes LH du réseau global : une liaison hydrogène peut, suite à une mutation, ne plus être connectée à l'atome d'intérêt tout en conservant sa stabilité.

# IV CHAPITRE

## Etude structurale de protéines fluorescentes de *Branchiostoma lanceolatum* par cristallographie aux rayons X

### A Contexte structural

#### A.1 Diversité des organismes contenant des protéines fluorescentes

L'écrasante majorité des informations structurales connues sur les protéines fluorescentes ont été obtenues pour des protéines d'organismes marins du phylum des cnidaires (*Cnidaria*) : méduses de la classe des hydrozoaires (*Hydrozoa*) [Prasher et al., 1992], coraux et anémones de mer de la classe des anthozoaires (*Anthozoa*) [Matz et al., 1999].

Après la résolution des structures cristallographiques de AvGFP [Yang et al., 1996] et de AvGFP-S65T [Ormö et al., 1996] en 1996, environ 450 structures de PF de cnidaires ont été déposées dans la PDB. Depuis des PF ont été identifiées dans des organismes d'autres phylums [Shagin et al., 2004, Deheyn et al., 2007] et un petit nombre de structures de ces protéines ont été déterminées, 5 du phylum des arthropodes (*Arthropoda*) depuis 2006, et 6 du phylum des chordés (*Chordata*) depuis 2013. L'intérêt de ces structures est, à travers la diversité génétique, de pouvoir identifier de nouveaux déterminants structuraux de fluorescence et ainsi de proposer de nouvelles pistes pour l'évolution dirigée.

## A.2 Structures de protéines fluorescentes du phylum *Arthropoda*

Les 5 structures de PF d'arthropodes déposées dans la PDB présentent un chromophore de type GYG comme, apparemment, toutes les PF d'arthropodes [Masuda et al., 2006]. Deux d'entre elles sont des structures de PF naturelles de zooplancton, l'une verte *CopGFP* ( $\lambda_{abs./em.} = 482/502$ ) et l'autre jaune-vert *CpYGFP* ( $\lambda_{abs./em.} = 509/517$ ) provenant des copépodes *Pontellina plumata* [Wilmann et al., 2006] et *Chiriduis poppei* [Suto et al., 2009]. L'évolution dirigée de *CopGFP* a mené à un mutant à la maturation particulièrement rapide, TurboGFP ( $\lambda_{abs./em.} = 482/502$ ). L'étude par mutagenèse de *CpYGFP* a conduit à s'intéresser au mutant *CpYGFP-H52T* ( $\lambda_{abs./em.} = 495/511$ ) qui montre un déplacement hypsochromique de son maximum d'absorption. La mutation H52T démontre que le déplacement bathochromique du maximum d'absorption observé chez *CpYGFP* est dû à une interaction de type  $\pi - \pi$  entre le groupement phénol du chromophore et le résidu His52, similaire à l'interaction  $\pi - \pi$  entre Tyr203 et le chromophore chez les *AvYFP* qui, elle, a été introduite par mutagenèse rationnelle. Il apparaît que les PF d'arthropodes sont plus proches de celles des chordés que des cnidaires [Shagin et al., 2004]. La liste des structures cristallographiques de PF du phylum *Arthropoda* résolues est fournie dans la table IV.1.

Protéine	Code PDB	Résolution (Å)	Groupe d'espace	Nb. molécules par unité asymétrique	$\lambda_{abs.}^{max}$ (nm)	$\lambda_{em.}^{max}$ (nm)	QY
<i>CopGFP</i> <sup>a</sup>	2G3O	2,10	$P3_2$	6	482	502	N.D.
<i>CpYGFP</i> <sup>b</sup>	2DD7	1,90	$C222_1$	2	509	517	N.D.
<i>CpYGFP-H52T</i> <sup>b</sup>	2DD9	2,30	$P2_12_12_1$	4	495	511	N.D.
TurboGFP <sup>c</sup> (pH 7,5)	3G6X	2,00	$C222_1$	4	482	502	0,53
TurboGFP <sup>c</sup> (pH 4,6)	3G6Y	1,60	$P12_11$	4	482	502	N.D.

Tableau IV.1 : Listes des structures de PF du phylum *Arthropoda*, (a) [Wilmann et al., 2006], (b) [Evdokimov et al., 2006], (c) [Suto et al., 2009]

### A.3 Structures de protéines fluorescentes du phylum *Chordata*

Les premiers résultats structuraux sur les protéines fluorescentes provenant de vers marins du genre *Branchiostoma* (phylum *Chordata*) ont été obtenus très récemment [Pletnev et al., 2013, Bomati et al., 2014]. Les six structures résolues proviennent soit de l'espèce *B. lanceolatum* qui colonise la Méditerranée, la mer Noire, la Manche, la mer du Nord et la côte est des Etats-Unis d'Amérique, soit de l'espèce *B. floridae* qui est uniquement localisée dans le golfe du Mexique. En mettant de côté les gènes des autres espèces de *Branchiostoma* encore peu caractérisées, toutes les PF de chordés présentent un chromophore de type GYG, à l'exclusion de deux protéines de *B. floridae* possédant un chromophore GYA, qui ne semble pas fluorescent [Yue et al., 2016]. La liste des structures de PF du genre *Branchiostoma* résolues par cristallographie aux rayons X est fournie dans la table IV.2.

Protéine	Code PDB	Résolution (Å)	Groupe d'espace	Nb. molécules par unité asymétrique	$\lambda_{abs.}^{max}$ (nm)	$\lambda_{em.}^{max}$ (nm)	QY
<i>lanRFP</i> <sup>a</sup>	4JF9	2,33	<i>P</i> 6 <sub>4</sub> 22	2	521	592	0,10
<i>lanRFPdam</i> <sup>a</sup>	4JEO	2,35	<i>P</i> 6 <sub>4</sub> 22	2	N.D.	N.D.	N.D.
<i>lanRFP-ΔS83</i> <sup>a</sup>	4JGE	1,94	<i>P</i> 6 <sub>4</sub> 22	2	514	594	N.D.
<i>lanGFP</i> <sup>a</sup>	4HVF	1,70	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	502	511	N.D.
<i>bfloGFP-c1</i> <sup>b</sup>	4DKM	1,95	<i>C</i> 121	8	493	521	1,50 × 10 <sup>-2</sup>
<i>bfloGFP-a1</i> <sup>b</sup>	4DKN	1,35	<i>C</i> 222 <sub>1</sub>	2	497	512	1,00

Tableau IV.2 : Listes des structures de PF du phylum *Chordata*, (a) [Pletnev et al., 2013], (b) [Bomati et al., 2014]

Pletnev et coll. ont résolu deux structures de PF naturelles de *B. lanceolatum*, la PF verte *lanGFP* et la PF rouge *lanRFP* [Pletnev et al., 2013]. En outre, ils ont résolu la structure de *lanRFP* endommagée par les rayons X appelée *lanRFPdam* et celle du mutant de délétion *lanRFP-ΔS83* dont les vitesses de maturation et de repliement sont accélérées.

La structure de *lanRFP* montre deux caractéristiques remarquables. D'une part, la protéine présente une glutamine (Gln211) à la position conservée du résidu catalytique de formation du chromophore des PF (Glu222 chez *AvGFP*). D'autre part, les modifications post-traductionnelles menant habituellement aux chromophores rouges (modification d'une liaison peptidique en groupement acylimine, ou rupture de la chaîne principale) sont absentes. L'extension de conjugaison électronique est en fait assurée par une liaison covalente encore jamais observée entre

l'atome de carbone central du pont méthylène du chromophore et le groupement phénol du résidu Tyr62 (figure IV.1). Cette liaison covalente remarquable est responsable de l'émission de fluorescence de *lanRFP* dans le rouge ( $\lambda_{abs./em.} = 521/592$ ) ainsi que du large déplacement de Stokes ( $\sim 70$  nm) comme démontré par les mutants *lanRFP-Y62S* ( $\lambda_{abs./em.} = 514/524$ ) et *lanRFP-Y62F* ( $\lambda_{abs./em.} = 511/526$ ).

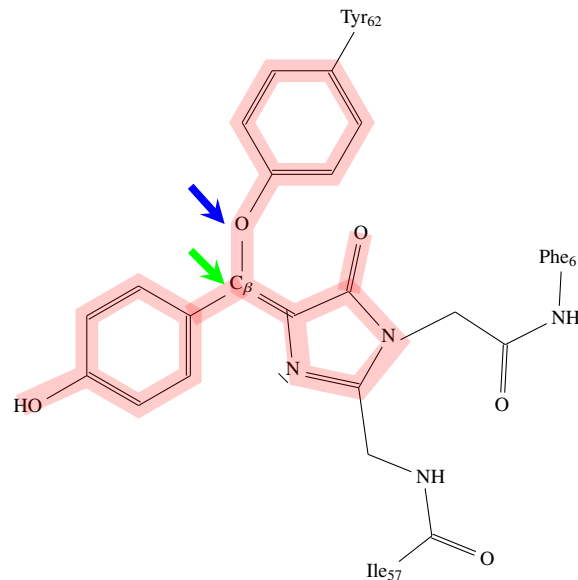


FIGURE IV.1 : Structure chimique du chromophore de *lanRFP* présentant une liaison covalente entre Tyr59/C $\beta$  (vert) et Tyr62/O (bleu) [Pletnev et al., 2013]

Bomati et coll. ont résolu deux structures de PF vertes issues de l'espèce *B. floridae* : *bfloGFP-c1* et *bfloGFP-a1* [Bomati et al., 2014]. Ces deux PF présentent des valeurs de rendement quantique très différentes. L'une est très fluorescente, l'autre quasiment pas. La comparaison de leurs structures permet donc de rationaliser les déterminants structuraux qui gouvernent les processus de fluorescence. La différence la plus remarquable entre ces structures est l'orientation opposée du noyau indole du résidu Trp157 au voisinage du chromophore entre les deux structures.

## A.4 Objectifs

Une collaboration entre Nathan Shaner (Scintillon Institute, San Diego, USA) et Antoine Royant (Institut de Biologie Structurale, Grenoble, France) a été établie en 2013 afin de comprendre pourquoi *lanYFP* et mNeonGreen sont des PF particulièrement brillantes [Shaner et al., 2013]. La protéine fluorescente *lanYFP* provient également de *B. lanceolatum* (gène naturel *bIFP-Y3*, code d'accèsion Genbank EU482389). Elle présente un rendement quantique ( $QY = 0,95$ ) et un coefficient d'extinction molaire ( $\epsilon = 150 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ ) particulièrement élevés. L'évolution de *lanYFP* ( $\lambda_{abs./em.} = 513/524$ ), PF naturellement tétramérique [Shaner et al., 2013], a été effectuée en quatre étapes successives alternant réduction de l'état d'oligomérisation et restauration de la fluorescence (figure IV.2) [Shaner et al., 2013].

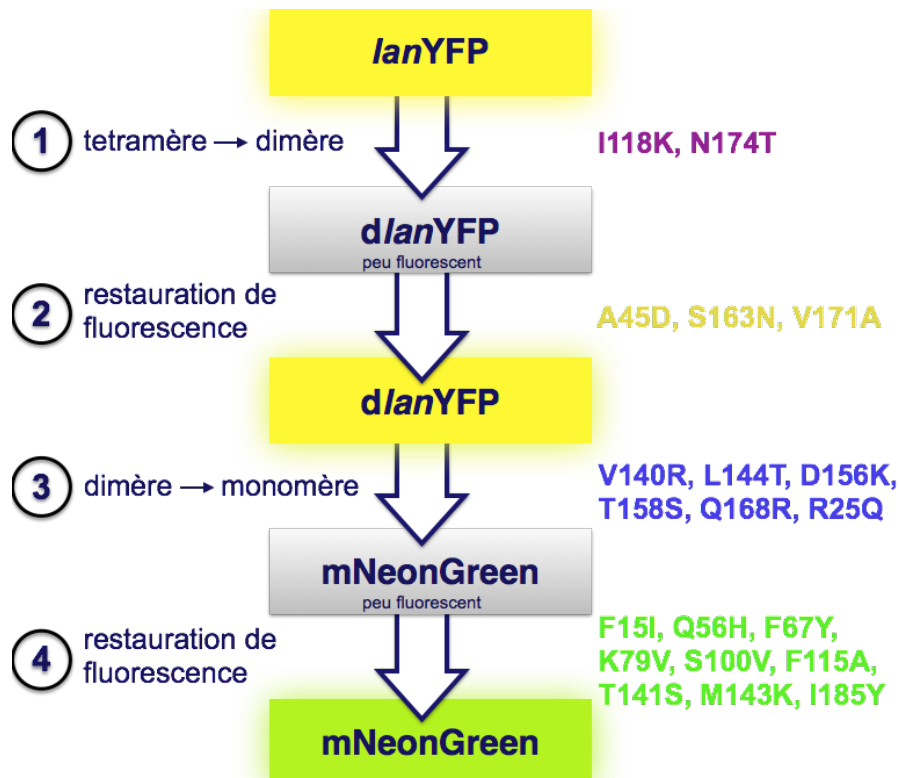


FIGURE IV.2 : Décomposition du processus d'évolution dirigée de *lanYFP* en *mNeonGreen*



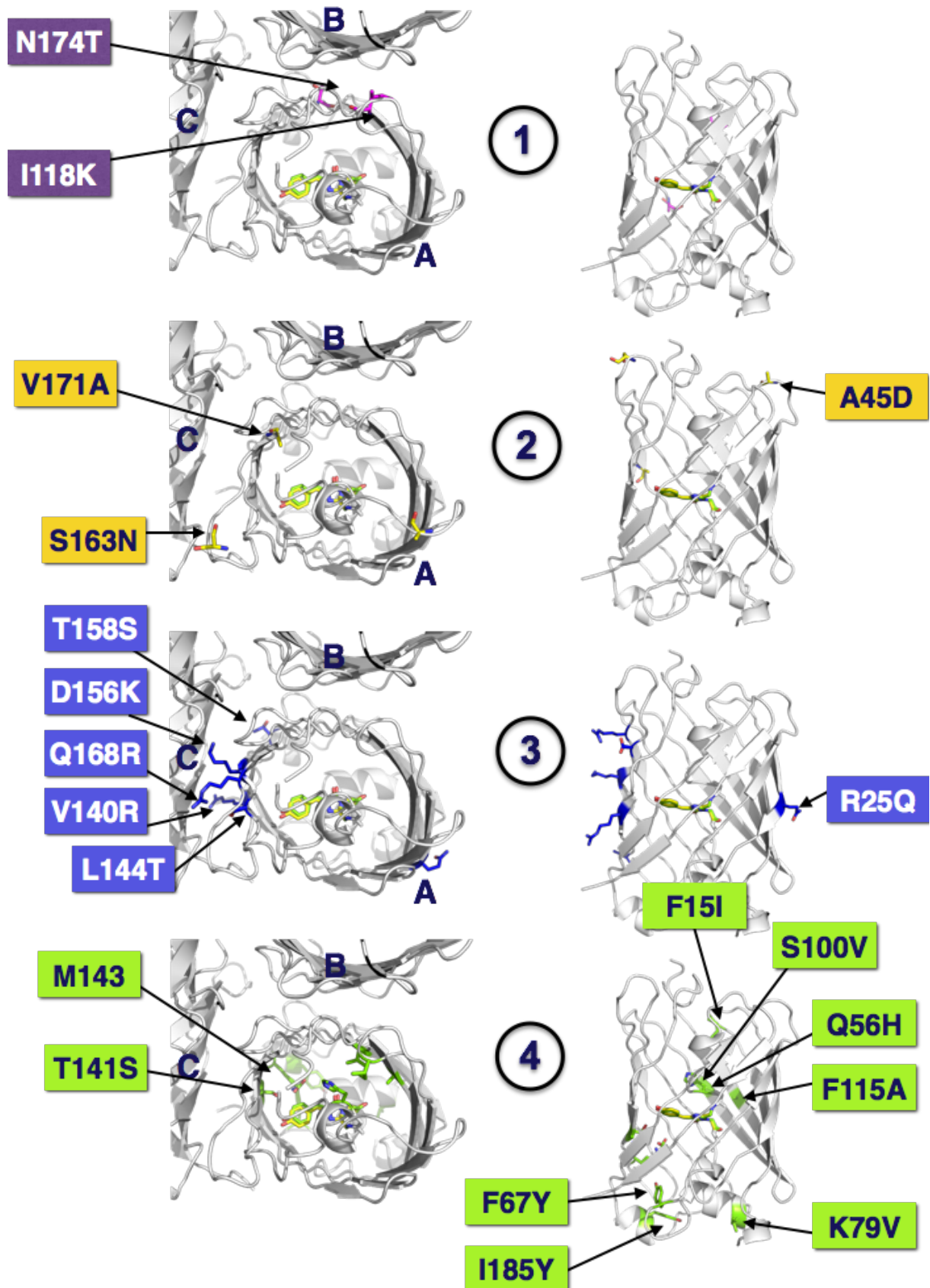


FIGURE IV.3 : Représentation tridimensionnelle du processus d'évolution de lanYFP en mNeonGreen décrivant les étapes de réduction du degré d'oligomérisation (étapes 1 [violette] et 3 [bleue]) et de restauration des propriétés d'émission de fluorescence (étapes 2 [jaune] et 4 [vert]).

Les PF tétramériques sont habituellement décrites comme quatre chaînes polypeptidiques A, B, C et D qui forment deux types d'interfaces : une hydrophile entre les monomères A et B, ou C et D (dite "interface A/B"), une autre hydrophobe entre les monomères A et C, ou B et D (dite "interface A/C").

L'**étape 1** a consisté à rompre l'interface A/B de *lanYFP* en introduisant les mutations I118K et N174T choisies en s'inspirant de la structure de *CopGFP*, dont la séquence est la plus proche de *lanYFP* parmi les structures de PF connues [Wilmann et al., 2006]. Cette étape a permis l'obtention d'un mutant dimérique mais peu fluorescent de *lanYFP*. L'**étape 2** est une opération de restauration de la fluorescence du mutant dimérique de *lanYFP*, avec l'introduction des mutations A45D, S163N et V171A. Cette étape a généré *dlanYFP* ( $\lambda_{abs./em.} = 513/524$ ), mutant dont les propriétés de brillance sont restaurées à 65% (QY = 0,90,  $\epsilon = 125 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ ). L'**étape 3** est l'étape de monomérisation du mutant intermédiaire brillant *dlanYFP*, où 6 mutations (V140R, I144T, D156K, T158S, Q168R et F192Y) permettent la rupture de l'interface A/C. Le mutant résultant n'est pas fluorescent et la **4<sup>ème</sup> étape** a permis de restaurer la fluorescence de ce mutant monomérique de *dlanYFP* en introduisant une série de 9 mutations (F15I, R25Q, Q56H, F67H, K79V, S100V, F115A, T141S, M143K et I185Y) disséminées sur toute la séquence protéique. A la fin de ce cycle d'évolution a été obtenu le mutant mNeonGreen ( $\lambda_{abs./em.} = 505/517$ ), avec un rendement quantique (QY = 0,80) et un coefficient d'extinction molaire ( $\epsilon = 116 \text{ mM}^{-1} \cdot \text{cm}^{-1}$ ) encore élevés, qui permettent de considérer mNeonGreen comme la PF étalon de la gamme spectrale jaune-vert. Au cours de cette étape, on note un décalage hypsochrome de 8 nm du maximum d'absorption de mNeonGreen responsable de son "verdissement". En résumé, mNeonGreen est une version monomérique de *lanYFP* avec une perte de brillance de seulement 35%, présentant un temps de maturation particulièrement rapide, inférieur à 10 minutes.

Afin de rationaliser le processus d'évolution de *lanYFP* en mNeonGreen, nous avons résolu les structures de *lanYFP* et mNeonGreen. Nous avons également étudié les structures de *lanYFP* et mNeonGreen par simulations de dynamique moléculaire afin d'identifier les résidus qui stabilisent, ou déstabilisent, le chromophore.

## **B Structure de protéines fluorescentes de *Branchiostoma lanceolatum***

### **B.1 Cristallisation**

Cette partie décrit les différents essais de cristallisation appliqués aux protéines fluorescentes mNeonGreen et *lanYFP*, qui ont été exprimés et purifiés comme décrit au paragraphe VI.A. Les deux PF ont été cristallisés en utilisant la technique de la diffusion de vapeur par goutte suspendue (voir paragraphe VI.B.1.c). L'ordre de présentation des expériences de cristallisation correspond à l'ordre d'apparition des cristaux au cours de ma thèse.

#### **B.1.a Cristallisation de mNeonGreen à pH acide**

Les efforts pour cristalliser mNeonGreen à pH acide ont nécessité de tester près de 400 conditions avec la plateforme de criblage de cristallisation à haut débit (HTX) du laboratoire européen de biologie moléculaire (EMBL) [Dimasi et al., 2007] et plus de 900 conditions manuelles. La meilleure condition de cristallisation a été identifiée avec un criblage manuel en utilisant les kits de cristallisation PEGRx 1, PEGRx 2, Index et Crystal Screen Cryo (Hampton Research, Aliso Viejo, USA) avec une concentration de protéine égale à 30 mg.mL<sup>-1</sup>. Les cristaux obtenus lors de ce criblage sont apparus après 7 à 15 jours à partir d'une solution de protéine digérée à la trypsine. Des oursins ont été obtenus avec la condition n°46 du kit PEGRx 1 composée de 0,1 M de citrate de sodium dihydraté à pH 5,0 et 18% w/v de PEG 20000. Cette condition a été affinée avec un criblage utilisant des concentrations de 14 à 22% de PEG 20 000 et des pH de 4,5 à 5,5. Cela a permis d'obtenir des aiguilles fines avec une solution tamponnée à pH 4,5 et 14% de PEG 20000. Le processus de cristallisation a été ralenti avec l'ajout de 0,2  $\mu$ L d'un crible d'additifs Additive Screen HT (Hampton Research, Aliso Viejo, USA). L'additif Cymal-7 a alors permis l'obtention de cristaux en aiguilles épaisses (figure IV.4). La condition de cristallisation finale est composée de 0,1 M de citrate de sodium dihydraté pH 4,50, 14% de PEG 20000 et de 6,8 mM de Cymal-7.

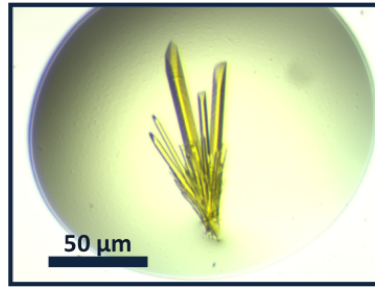


FIGURE IV.4 : Cristaux de mNeonGreen à pH 4,50.

### B.1.b Cristallisation de *lanYFP* à pH physiologique

La condition optimisée de cristallisation a été obtenue après 200 essais manuels de cristallisation. L'identification de cette condition de cristallisation a été réalisée avec deux kits de cristallisation PEGRx 1 et PEGRx 2 (Hampton Research, Aliso Viejo, USA) et des concentrations en protéine entre 10 et 30 mg.mL<sup>-1</sup>. Les premiers cristaux sont apparus au bout de 15 jours dans la condition n°9 de Peg Rx2 composée de 0,19 mM de Cymal-7, 0,1 M d'HEPES pH 7,5 et 40% v/v de PEG 400. Le criblage autour de cette condition a été réalisé avec des solutions tampons entre pH 7,0 et 8,0 et des concentrations de PEG 400 entre 30 et 45%, mais sans amélioration notable. La meilleure condition identifiée est celle de la condition n°9 de PEGRx 2 avec une concentration de protéine de 10 mg.mL<sup>-1</sup>.

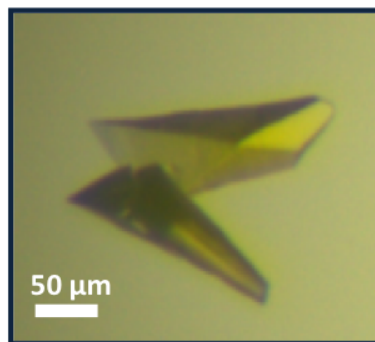


FIGURE IV.5 : Cristaux de *lanYFP* entremêlés.

### B.1.c Cristallisation de mNeonGreen à pH physiologique

Le criblage a été effectué avec la plateforme HTX de l'EMBL [Dimasi et al., 2007] avec des solutions de protéine digérée à la trypsine à des concentrations 10, 20 ou 30 mg.mL<sup>-1</sup>. Les premiers cristaux sont apparus au bout de 72 h dans la condition contenant 20% PEG 8000

et 100 mM d'HEPES pH 7,5 (crible Wizard I+II, Rigaku corporation, Tokyo, Japon) à des concentrations de 10 mg.mL<sup>-1</sup>.

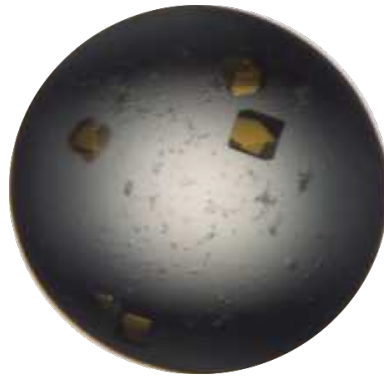


FIGURE IV.6 : Cristaux de mNeonGreen à pH physiologique photographiés après 9 jours sur la plateforme du HTX

Un criblage manuel a été effectué autour de cette condition en utilisant des concentrations de 18 à 22% PEG 8000, et des pH de 7,2 à 7,8. Les cristaux obtenus diffèrent légèrement par rapport aux cristaux de la plateforme HTX avec des bipyramides imbriquées (figure IV.7.a). La morphologie des cristaux a été restaurée en effectuant une étape de digestion à la trypsine directement dans la goutte de cristallisation (figure IV.7.b). La meilleure condition de cristallisation est composée de 20% de PEG 8000 et 100 mM d'HEPES à pH 7,5 avec une concentration en protéine de 10 mg.mL<sup>-1</sup>.

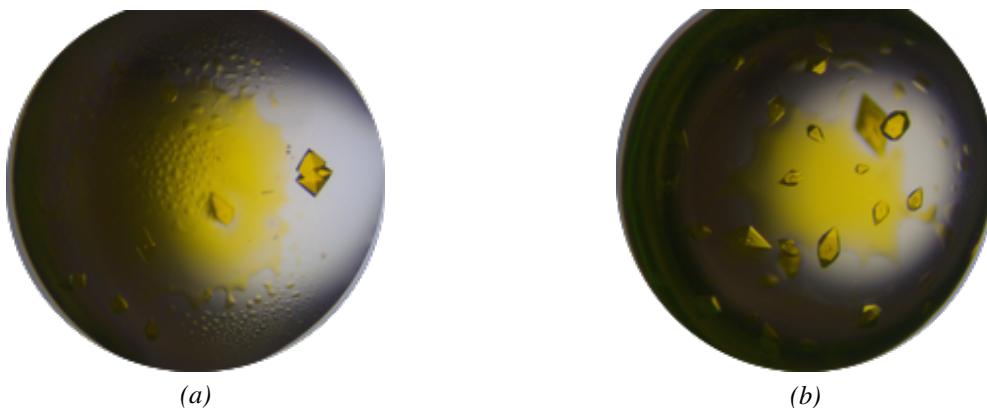


FIGURE IV.7 : Cristaux de mNeonGreen obtenus avec trypsinisation de la protéine effectuée (a) avant et (b) pendant la cristallisation

Nous avons vérifié que les cristaux obtenus à différents pH rendaient compte de l'équilibre acido-basique du chromophore en enregistrant des spectres d'absorption UV-visible *in crystallo*. Ces spectres montrent que l'acidification de la liqueur mère par un saut de pH 4,5 à pH 8,0 fait augmenter la fraction protonée du chromophore dans le cristal. La bande d'absorption

caractéristique de cette espèce protonée est à 403 nm alors que la bande correspondant à la forme déprotonée est à 506 nm (figure IV.8).

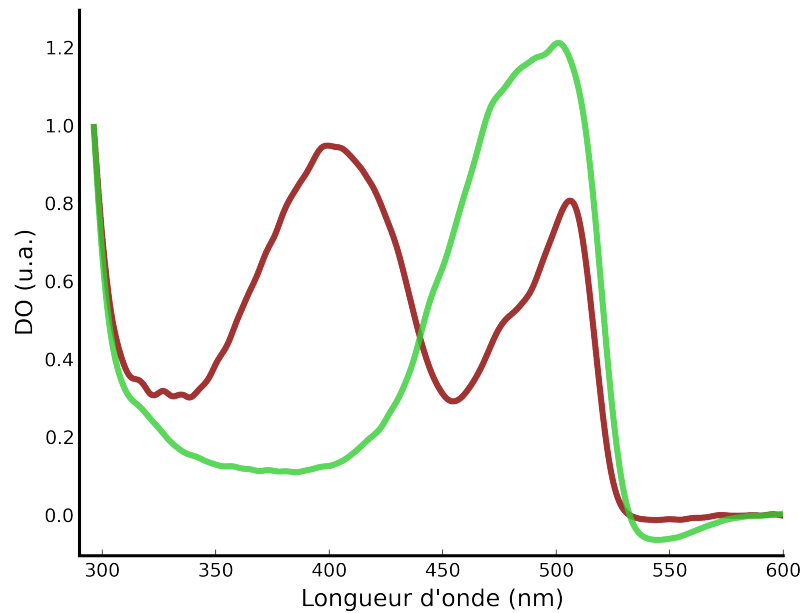


FIGURE IV.8 : Spectres d'absorption UV-visible sur cristaux de mNeonGreen à pH acide (pH 4,5 - rouge foncé) et physiologique (pH 8,0 - vert)

## B.2 Structures cristallographiques de *lan*YFP et mNeonGreen

### B.2.a Collecte des données de diffraction et résolution des structures

Les données de diffraction ont toutes été collectées à l'Installation Européenne de Rayonnement Synchrotron de Grenoble (ESRF, *European Synchrotron Radiation Facility*). Les statistiques de réduction des données sont présentées en table IV.3.

La résolution des structures des protéines fluorescentes de *B. lanceolatum* a été effectuée selon la méthode de remplacement moléculaire (voir paragraphe VI.B.2.f ii). La structure qui a servi de modèle est celle de *lan*RFP (code PDB [4JF9](#)), dont la séquence est la plus proche de celle de mNeonGreen (69% d'identité de séquence). L'opération de phasage a été réalisée avec le logiciel PHASER-MR [[McCoy, 2007](#)]. La première structure à avoir été résolue est celle de mNeonGreen à pH acide. Les structures de mNeonGreen et de *lan*YFP à pH physiologique ont été résolues en utilisant le modèle de mNeonGreen à pH acide.

L'affinement des structures a été mené de façon itérative en reconstruisant le modèle dans le logiciel de visualisation de structures COOT (voir paragraphe VI.B.2.f iii) puis l'affinement a été réalisé dans REFMAC5 dont les statistiques sont consignées dans la table IV.3. Les structures résolues au cours de ma thèse ont été déposées dans la banque de données PDB (*Protein Data Bank*) :

- 5LTP pour la structure de mNeonGreen à pH acide
- 5LTR pour la structure de mNeonGreen à pH physiologique
- 5LTQ pour la structure de *lan*YFP à pH physiologique

Protéine	<i>lan</i> YFP	mNeonGreen (pH 4.5)	mNeonGreen (pH 8.0)
Code PDB	5LTQ	5LTP	5LTR
<b>Paramètre de maille et statistique des données collectées</b>			
Longueur d'onde (Å)	0,976	0,976	0,954
Température (K)	100	100	100
pH	7,5	4,5	8,0
Groupe d'espace	P2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P6 <sub>5</sub> 22
Maille élémentaire			
a, b, c (Å)	78,7 197,2 115,1	75,9 127,6 146,9	72,3 72,3 154,0
$\alpha, \beta, \gamma$ (°)	90,0 90,5 90,0	90,0 90,0 90,0	90,0 90,0 120,0
Résolution (Å)	49,29 - 2,05 (2,12 - 2,05)	46,34 - 1,70 (1,76 - 1,70)	48,56 - 1,21 (1,25 - 1,21)
Nb. total de réflexions	835 157 (86 173)	670 077 (65 962)	1 330 789 (131 417)
Nb. de réflexions uniques	217 268 (21 713)	155 051 (15 390)	72 944 (7 142)
Multiplicité	3,8 (4,0)	4,3 (4,3)	18,2 (18,4)
Complétude (%)	99,38 (99,55)	98,87 (99,05)	99,86 (99,92)
$I/\sigma$ moyen (I)	13,5 (2,5)	19,4 (1,9)	18,8 (2,1)
Facteur de Wilson (Å <sup>2</sup> )	28,5	22,4	12,9
$R_{meas.}^{\dagger}$	0,088 (0,659)	0,054 (0,844)	0,088 (1,705)
$CC_{1/2}^{\#}$	0,998 (0,730)	1,000 (0,678)	1,000 (0,729)
<b>Paramètre d'affinement de structure</b>			
Nb. de molécules par u. a.*	16	6	1
Résolution (Å)	49,29 - 2,05	46,34 - 1,70	48,56 - 1,21
$R_{cris.} / R_{libre}$	0,216 / 0,244 (0,296 / 0,321)	0,181 / 0,205 (0,275 / 0,301)	0,158 / 0,184 (0,316 / 0,346)
Nb. d'atomes			
Protéine	27 491	10 879	1 876
Chromophore	304	114	19
Molécules d'eau	553	911	230
Agitation thermique B (Å <sup>2</sup> )			
Protéine	38,4	27,5	15,1
Chromophore	34,2	18,8	9,5
Molécules d'eau	27,8	32,5	25,8
Déviations quadratiques moyennes			
Longueurs des liaisons	0,006	0,008	0,009
Angles des liaisons	1,29	1,50	1,59

Tableau IV.3 : Table des données cristallographiques. les valeurs entre parenthèse correspondent aux coquilles de plus haute résolution),(#) la coupure en résolution est basée sur le  $CC_{1/2}$ , (†)  $R_{meas.} = R_{merge} \times [N/(N - 1)]^{1/2}$ , où  $N$  est la multiplicité. (\*) u.a. : unité asymétrique



### B.2.b Structure de mNeonGreen à pH acide

La structure de mNeonGreen à pH acide a été déterminée à 1,70 Å de résolution dans le groupe d'espace  $P2_12_12_1$  (voir table IV.3). L'unité asymétrique de la maille du cristal contient 6 molécules. Comme mNeonGreen est supposée être monomérique [Shaner et al., 2013], cet arrangement est-il artificiel ? L'observation de l'hexamère suggère qu'il peut être décrit comme un dimère de trimère. En utilisant le serveur PISAe [Krissinel and Henrick, 2007], j'ai caractérisé les surfaces d'interaction entre les monomères de l'unité asymétrique (table annexe VII.F). Les quatre surfaces d'interaction maximales au sein des trimères sont de 622, 667, 682 et 726 Å<sup>2</sup>, inférieures aux surfaces d'interaction rencontrées usuellement chez les tétramères naturels de PF. On verra plus tard au paragraphe IV.B.2.c que les surfaces d'interaction de *lanYFP* sont supérieures à 840 Å<sup>2</sup>. On note que les surfaces d'interaction entre trimères sont de 482 Å<sup>2</sup> maximum. Il apparaît donc que l'état hexamérique que nous avons observé à pH acide pour mNeonGreen, voir tétramérique, n'a pas de pertinence physiologique (figure IV.9).

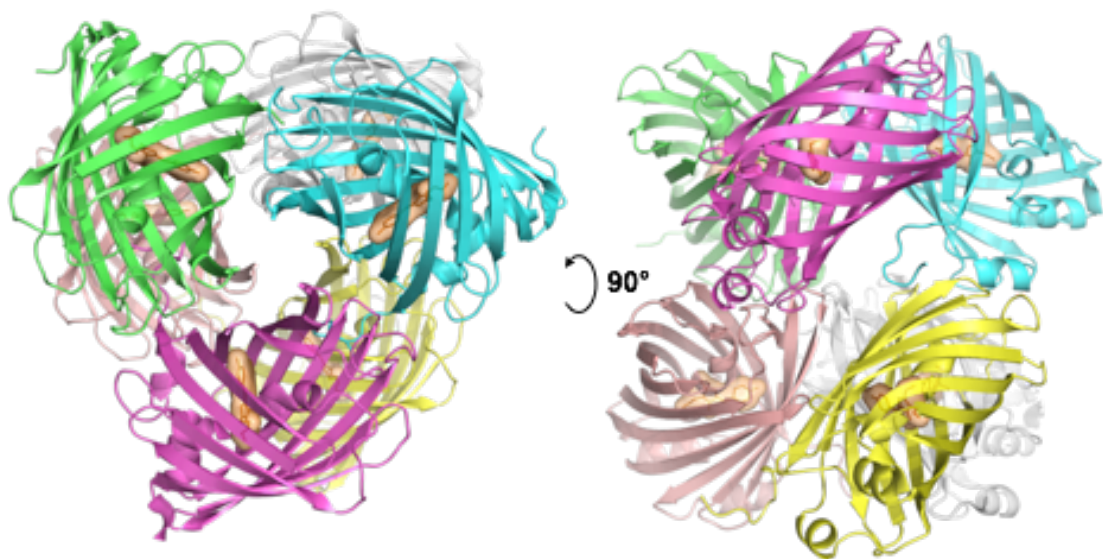


FIGURE IV.9 : Arrangement hexamérique de mNeonGreen dans l'unité asymétrique des cristaux obtenus à pH acide.

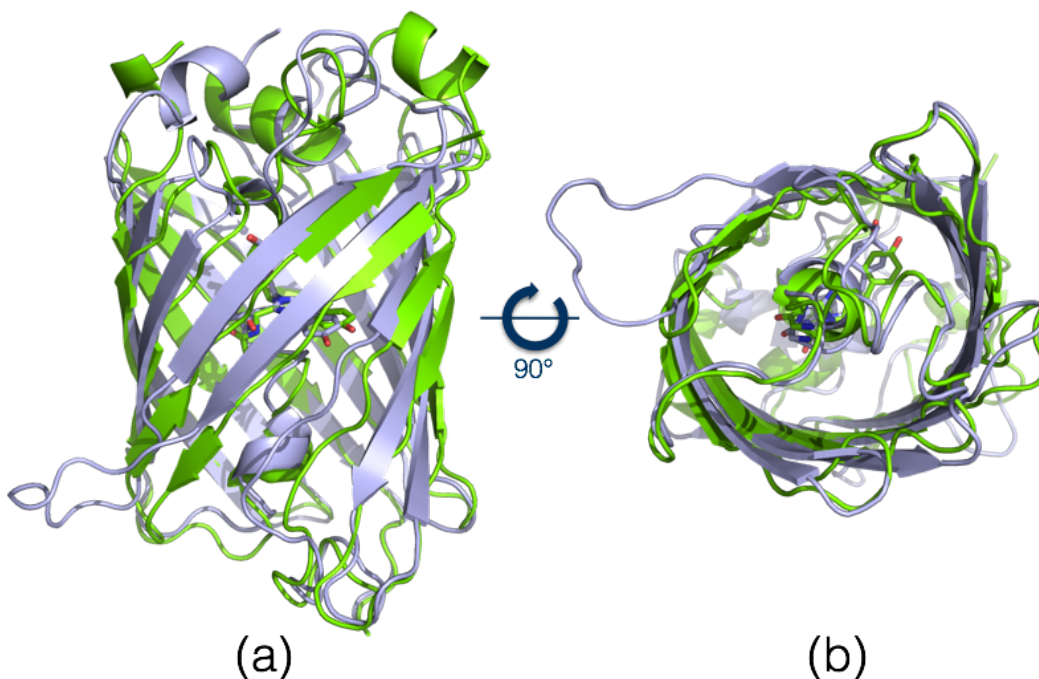


FIGURE IV.10 : Comparaison structurale entre AvGFP code PDB 1EMA [Ormö et al., 1996] (bleu clair) et de mNeonGreen à pH acide (vert).

Si l'on compare les structures tertiaires de mNeonGreen et de *A. victoria* (1EMA [Ormö et al., 1996]), on remarque que les différences structurales sont minimales (figure B.2.b.a) avec un RMSd égale à 1,4 Å. En effet, la présence du tonneau de 11 feuillets  $\beta$  ainsi qu'un chromophore supporté par une hélice  $\alpha$  coaxiale est confirmé. Toutefois, la boucle reliant les feuillets  $\beta$  10 et 11 est plus courte dans le cas de mNeonGreen ce qui structurellement conduit à l'obtention d'un tonneau plus compact (délétion de 4 acides aminés). Par ailleurs, on remarque que la position du chromophore chez AvGFP est différente dans mNeonGreen (figure .b).

### B.2.c Structure de *lanYFP* à pH physiologique

La structure de *lanYFP* a été déterminée à 2,05 Å de résolution dans le groupe d'espace P2<sub>1</sub> par remplacement moléculaire en utilisant comme modèle la structure de mNeonGreen à bas pH, avec 16 monomères dans l'unité asymétrique. *lanYFP* formant physiologiquement des tétramères [Shaner et al., 2013], l'unité asymétrique est constituée de tétramères de tétramères. La nomenclature des 4 tétramères de *lanYFP* a été interprétée à partir de celle classiquement utilisée pour DsRed et autres PF naturellement tétramérique (figure IV.11) [Yarbrough et al., 2001].

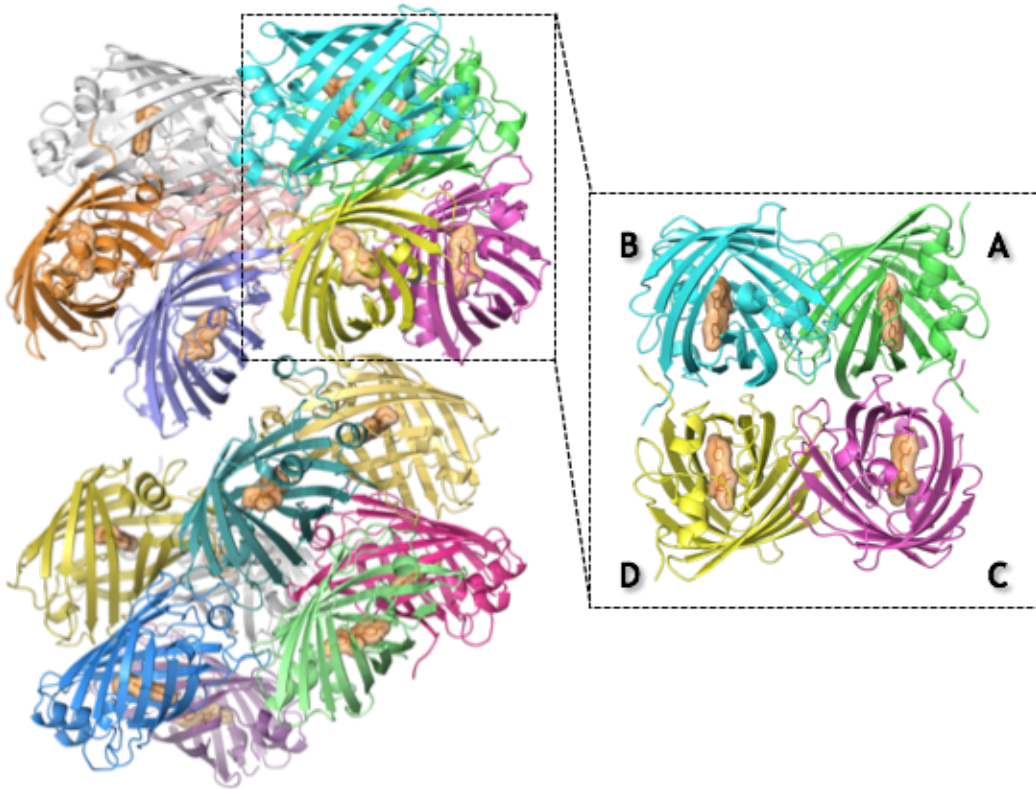


FIGURE IV.11 : Arrangement du 16-mère de lanYFP dans l'unité asymétrique avec zoom sur un des tétramères.

La caractérisation des surfaces d'interaction entre les monomères de l'unité asymétrique réalisée avec le serveur PISAe [Krissinel and Henrick, 2007] montre que les interfaces A/B et A/C ont des surfaces minimales respectives de 840 et 1020 Å<sup>2</sup>, respectivement (figure IV.12.a) à l'exception d'une des deux l'interfaces A/C du tétramère MNOP, ce qui s'explique par une des extrémités C terminales non modélisée en raison de son agitation. Les interactions hydrophiles au sein des interfaces s'effectuent principalement par des liaisons hydrogène, seules les interfaces de type A/C sont complétées par des ponts salins (figure IV.12.b) suggérant que cette interface est la plus forte de l'arrangement. Les interfaces entre tétramères ont une valeur maximale de 317 Å<sup>2</sup> ce qui confirme que l'état d'oligomérisation est tétramérique .

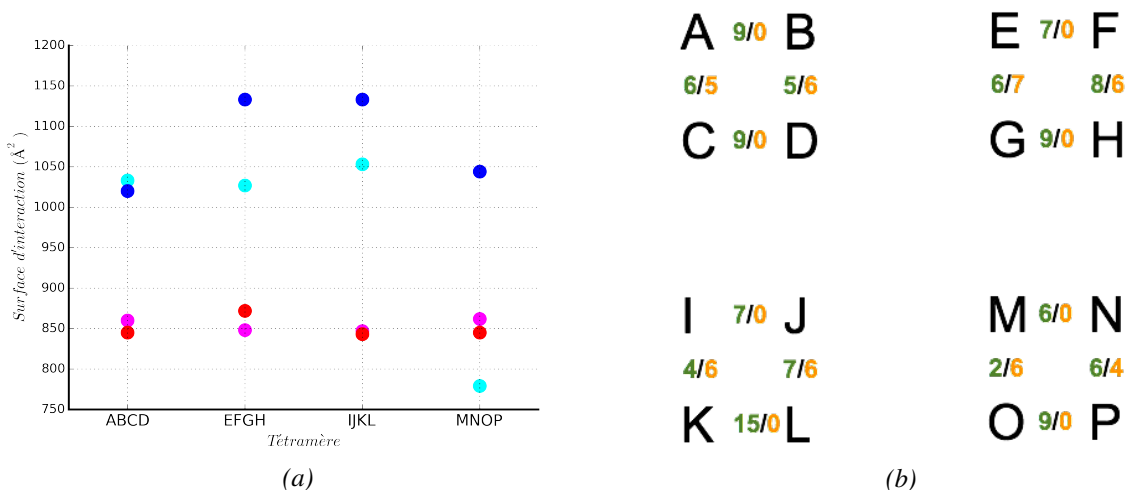


FIGURE IV.12 : Description des interfaces de lanYFP entre les tétramères ABCD, EFGH, IJKL et MNOP, (a) surface d'interaction (magenta/rouge pour l'interface A/B et bleu/cyan pour l'interface A/C), (b) nombre d'interaction hydrophiles à chaque interfaces A/B et A/C (vert : liaisons hydrogène, orange : ponts salins)

L'utilisation de PISAe m'a permis d'identifier précisément les résidus impliqués dans les deux interfaces. L'interface A/B est composée des résidus 14, 16, 17, 19, 85, 87, 89, 91, 95, 97, 99, 101, 103, 116, 118, 120-124, 146, 147, 150, 152, 172, 174 et 176. L'interface A/C, des résidus suivants 137-140, 142, 144, 145, 154, 156, 158, 166, 168, 172, 189, 190, 192, 194-196, 211-219.

### B.2.d Structure de mNeonGreen à pH physiologique

La structure de mNeonGreen à pH physiologique a été déterminée à 1,21 Å de résolution dans le groupe d'espace P6<sub>5</sub>22 par remplacement moléculaire en utilisant comme modèle la structure de mNeonGreen à pH acide. Contrairement aux deux structures décrites précédemment, l'unité asymétrique du cristal est composée d'une seule molécule, cela est de bon augure pour appuyer le fait que mNeonGreen est monomérique en solution [Shaner et al., 2013]. Pour le vérifier, nous avons calculé les surfaces d'interaction (serveur PISAe [Krissinel and Henrick, 2007]) avec les molécules symétriques, ces surfaces sont de 742, 562, 218 et 98 Å<sup>2</sup>, inférieures aux surfaces d'interaction de 840 Å<sup>2</sup> trouvées pour les deux structures précédentes.

### B.2.e Comparaison des structures de lanYFP et mNeonGreen à pH physiologique

La comparaison de ces deux structures va nous permettre de rationaliser le processus d'évolution dirigée de la protéine tétramérique lanYFP en mNeonGreen, protéine monomérique (figure IV.2).

L'étape 1 a consisté à passer d'une protéine tétramérique à une protéine dimérique en introduisant les mutations N174T et I118K. Dans la structure de *lanYFP*, Ile118 est situé à proximité des résidus Asn101 et Arg103 d'un autre monomère. Sa mutation en lysine introduit un groupement chargé et plus volumineux qui vient déstabiliser deux îlots d'interaction de l'interface A/B, ce qui a apparemment suffi pour passer du tétramère *lanYFP* à un mutant dimérique peu fluorescent (figure IV.13.a) [Shaner et al., 2013]. Dans l'étape 2, l'introduction des mutations A45D, S163N et V171A soit à la surface de la protéine, soit à un endroit très proche permettent de restaurer significativement la fluorescence de la protéine. Comme ces mutations sont éloignées du chromophore, il nous faut considérer que ces mutations sont également des mutations de repliement.

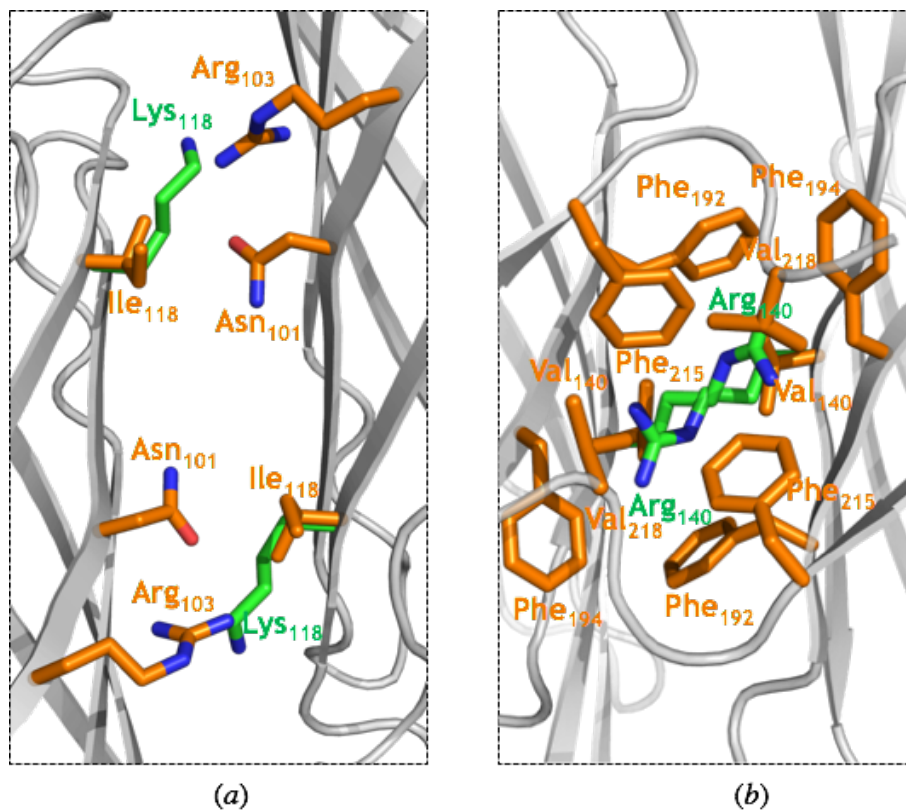


FIGURE IV.13 : Interfaces (a) A/B et (b) A/C de *lanYFP* (chaîne latérale orange) représentant deux mutations cruciales pour la monomérisation de *mNeonGreen* (chaîne latérale vertes)

L'étape 3 a permis de monomériser le mutant dimérique fluorescent *dlanYFP* en introduisant les six mutations R25Q, L114T, V140R, D156K, T158S et Q168R. La mutation qui déstabilise le plus l'interface A/C est la mutation V140R qui perturbe un large îlot hydrophobe composé de six phénylalanines et de deux valines (figure IV.13.b). Par ailleurs, les surfaces électrostatiques des monomère A et C de *lanYFP* (figure IV.14.a) montre une complémentarité des surfaces chargées positivement et négativement. La déstabilisation de l'interface A/C est complétée par l'ajout de charges positives grâce aux mutations D156K et Q168R (figure IV.14.c.). Ces mutations entraînent une répulsion électrostatique des interfaces A et C en modifiant en profondeur la polarité de cette face du monomère A, région jaune sur les figures IV.14.b et IV.14.c. Les mutations R25Q, L144T et T158S n'ont pas de rôle évident d'après la structure.

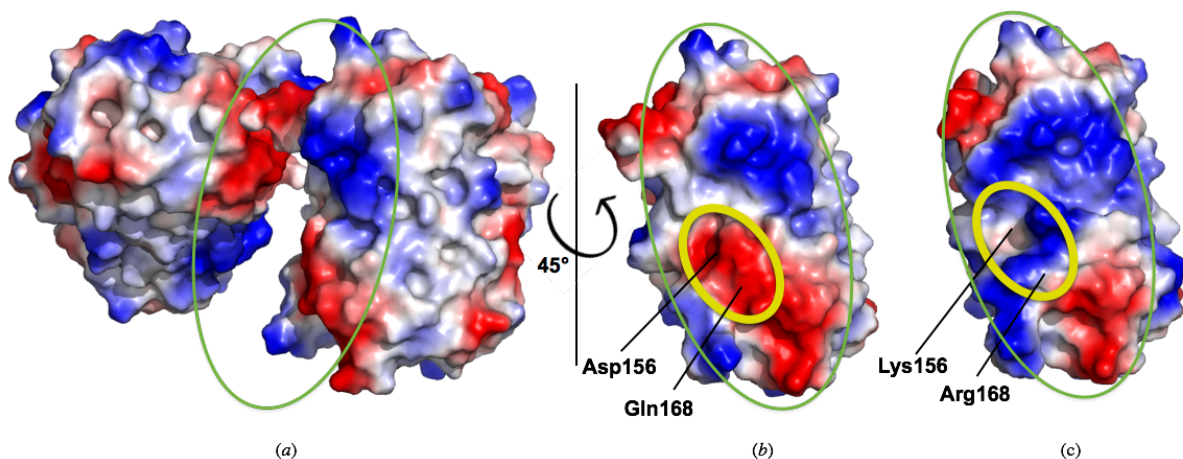


FIGURE IV.14 : Représentation des surfaces électrostatiques (positive en rouge et négative en bleu) de *lanYFP* et *mNeonGreen* à l'interface A/C (cerclée de vert), (a) vue éclatée du dimère de *lanYFP*, (b) face du monomère A de *lanYFP* à l'interface avec C et (c) face du monomère de *mNeonGreen*

L'étape 4 a consisté à restaurer des propriétés de fluorescence fortes chez la protéine monomérique issue de l'étape 3. Pour comprendre comment le processus s'est déroulé, nous avons comparé l'environnement du chromophore chez *lanYFP* et *mNeonGreen*. A pH physiologique, les structures de *lanYFP* et *mNeonGreen* présentent deux changements significatifs, un de chaque côté du chromophore.

Il existe un site de fixation aux ions chlorures situé à 4,5 Å du cycle phénolate du chromophore de *lanYFP* (figure IV.16.a). L'absence d'ions chlorure dans la liqueur mère suggère que l'affinité de *lanYFP* pour les ions chlorure est forte, ce qui rappelle le résultat obtenu pour la YFP de *AvGFP* [Galiotta et al., 2001]. Dans *mNeonGreen* ce site de fixation existe également, mais pour une occupation partielle (30%). Cela est dû à la présence de Lys143 (Met143 chez *lanYFP*)

qui présente la particularité d'être partiellement carboxylée (figure IV.15). Il est curieux de noter que ce type de modification est potentiellement présente à hauteur de 1,3% chez les macromolécules [Jimenez-Morales et al., 2014]. Il apparaît toutefois que cette carboxylation proche du chromophore n'est pas responsable de la modulation des propriétés spectroscopiques du chromophore (baisse du QY et du  $\epsilon$ , décalage vers le bleu du maximum de fluorescence) car la mutation M143K a été introduit tardivement dans le processus d'évolution. De facto, cette modification fortuite a entraîné une réduction de la sensibilité de mNeonGreen aux ions chlorures par rapport à *lanYFP*.

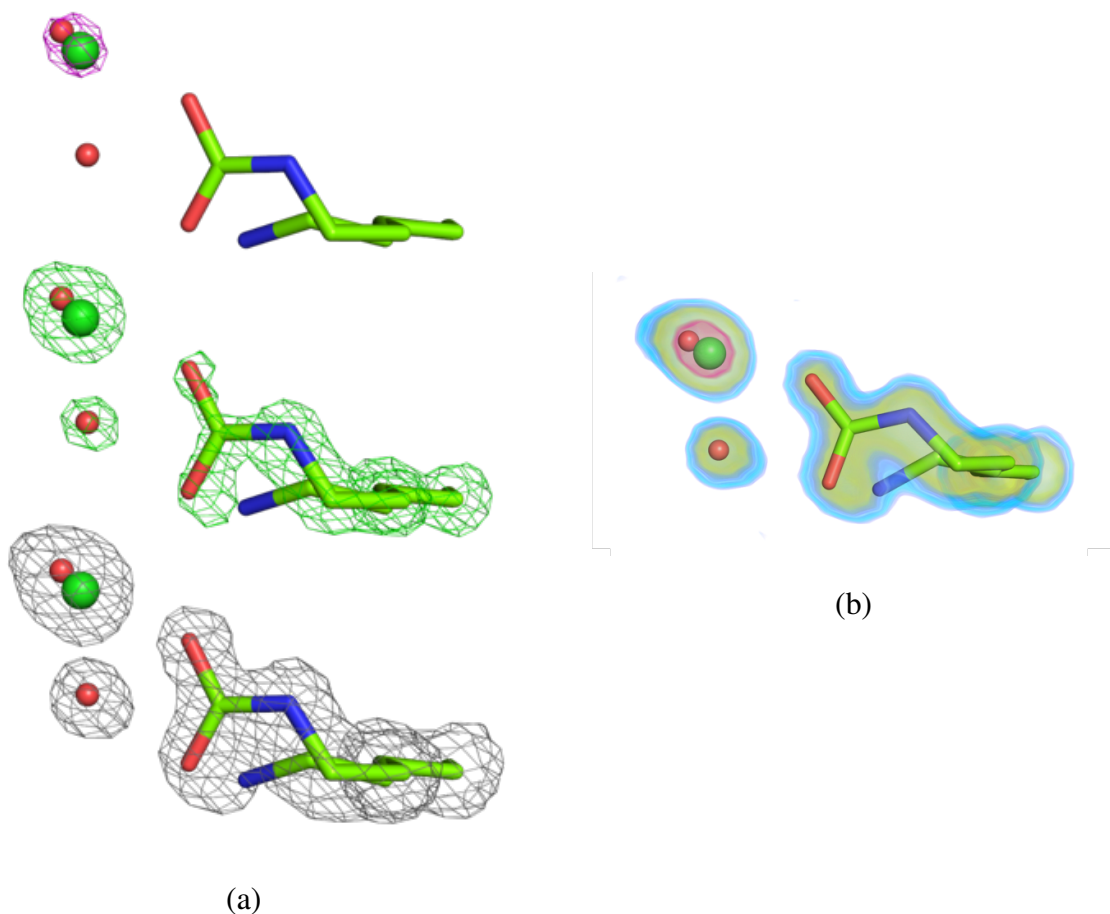


FIGURE IV.15 : Densité électronique autour du résidu 143 chez mNeonGreen à pH physiologique. (a) Représentation classique de la densité en maille contourée à un niveau de  $\sigma$  de 5,0, 2,0 et 1,0 (de haut en bas) montrant progressivement la présence de l'ion chlore (dans la carte à 5,0  $\sigma$ ), du groupement carboxylate (carte à 2,0  $\sigma$ ) et de la conformation non carboxylée de la lysine (carte à 1,0  $\sigma$ ). (b) Représentation volumique des trois cartes superposées (magenta : 5,0  $\sigma$ , vert : 2,0  $\sigma$ , bleu : 1,0  $\sigma$ ).

Les changements structuraux les plus significatifs se situent de l'autre côté du chromophore (figure IV.16.b). Le groupement carbonyle du résidu Pro55 est significativement déplacé de 1,2 Å entre mNeonGreen et *lanYFP* dans une direction parallèle au chromophore, modifiant l'interaction doublet non-liant/ $\pi$  entre les cinq atomes du cycle imidazolinone du chromophore et l'oxygène du carbonyle de Pro55.

Dans le cas de *lanYFP*, l'oxygène du carbonyle de Pro55 est en interaction symétrique avec les cinq atomes du cycle imidazolinone du chromophore (pointillés rouges sur la figure IV.16.b) tandis qu'elle est asymétrique dans le cas de mNeonGreen (pointillés bleus sur la figure IV.16.b). Cette modification de l'interaction polaire avec le nuage électronique du chromophore module très certainement les moments de transition dipolaire d'absorption et d'émission du chromophore ayant pour conséquence le déplacement de 7 nm dans le bleu des maxima d'absorption et d'émission. L'origine du déplacement de Pro55 est dû aux effets concertés de l'introduction de trois mutations Q56H, S100V et F115A lors de l'**étape 4** de restauration de fluorescence. La mutation F115A permet la rotation de la chaîne latérale de Tyr102 en diminuant l'encombrement stérique et mène à l'établissement d'une liaison hydrogène courte de 2,6 Å entre l'oxygène du phénolate de Tyr102 et le carbonyle de His56 exerçant une force sur la chaîne principale des résidus 55 et 56 (figure IV.16.b). En conséquence, le résidu Trp157, qui est en interaction CH/ $\pi$  avec Pro55, est également poussé dans la même direction.



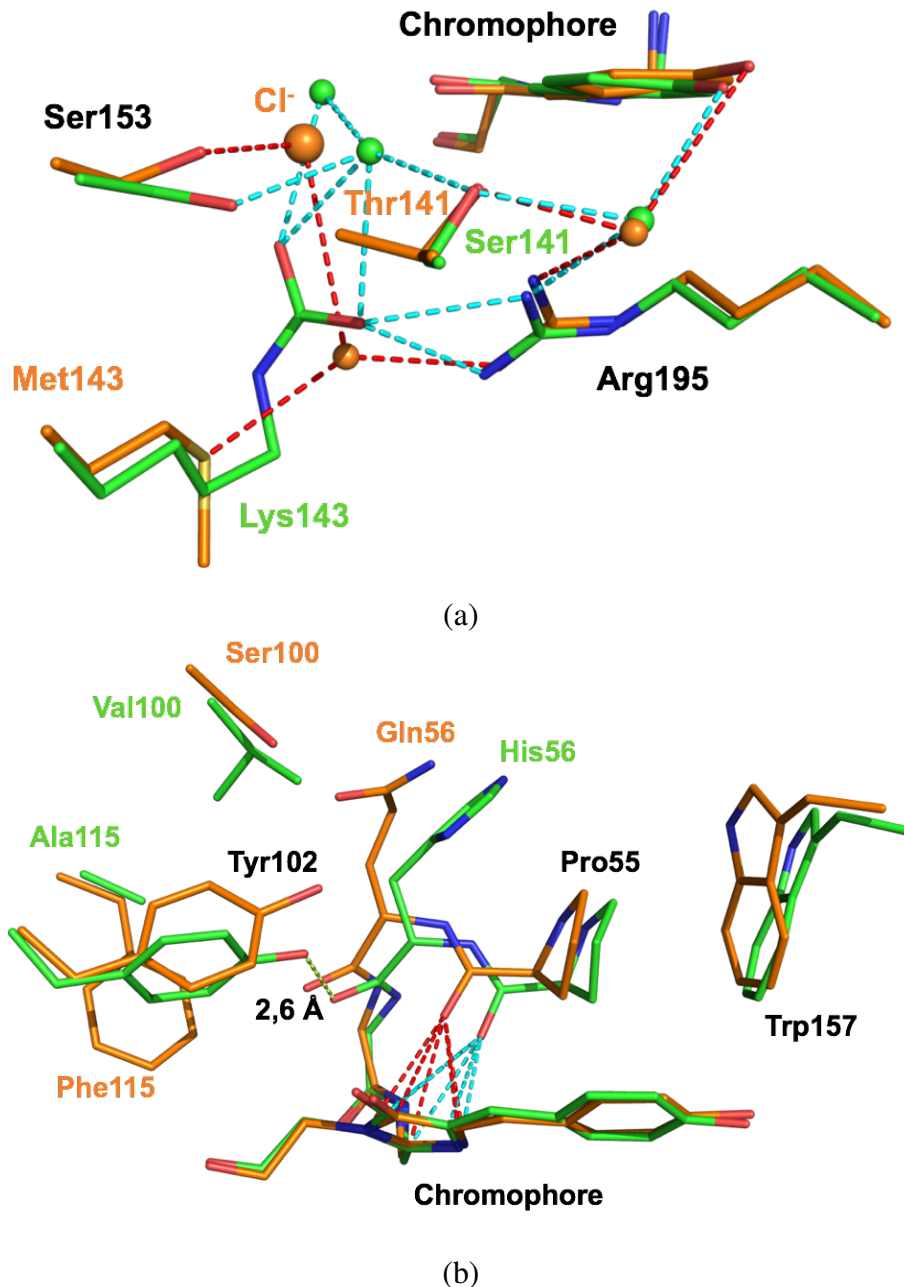


FIGURE IV.16 : Comparaison de l'environnement du chromophore pour lanYFP (orange) et mNeonGreen (vert), (a) site de fixation de l'ion chlorure de lanYFP et de la lysine carboxylée pour mNeonGreen à pH physiologique, (b) différence localisée sur la face opposée du chromophore. La liaison hydrogène entre Tyr102 et le groupement carbonyle de Lys56 est représentée par des pointillés jaune. L'interaction doublet non-liant /  $\pi$  entre le carbonyle de Pro55 et le cycle de l'imidazolinone du chromophore est représenté par des pointillés rouges pour lanYFP et cyans pour mNeonGreen.

### B.2.f Comparaison des structures de mNeonGreen

Pour comprendre pourquoi mNeonGreen est peu fluorescente en dessous de son  $pH_{1/2}$ , nous avons comparé les structures à pH 4,5 et 8,0. Il y a une seule modification qui se restreint au résidu 143. En effet, à bas pH, la lysine n'est pas du tout carboxylée et laisse donc la place à un site de fixation d'un ion chlorure. Cette différence s'explique par le fait que les lysines carboxylées ne se forment qu'à pH basique [Jimenez-Morales et al., 2014].

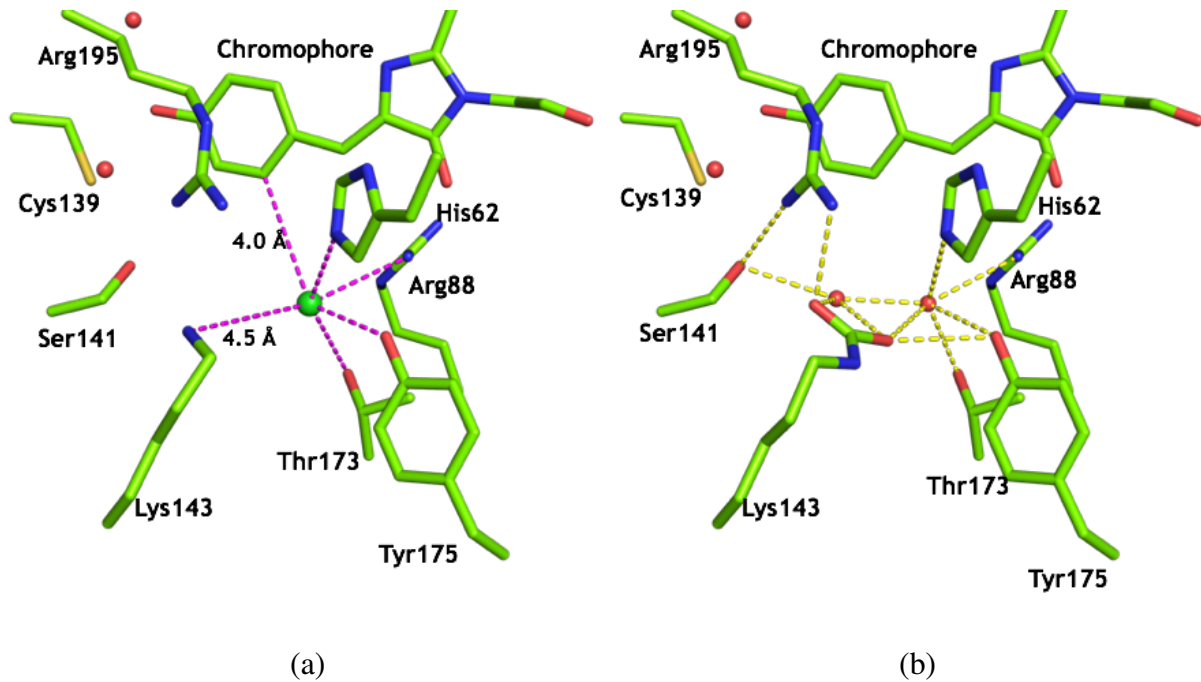


FIGURE IV.17 : Environnement du chromophore de mNeonGreen à pH acide (a) et physiologique (b)

### B.3 Dommages spécifiques aux rayons X subis par mNeonGreen

La première collecte de diffraction à température cryogénique (100 K) sur des cristaux de mNeonGreen à pH physiologique nous a révélé une caractéristique inattendue, avec le changement de couleur du volume du cristal irradié par les rayons X de jaune-vert à orange (figure IV.18). Ce phénomène est à notre connaissance remarquable puisque habituellement les rayons X induisent des processus de photo-blanchiment dans les cristaux de protéines colorées. Afin de comprendre les mécanismes structuraux responsables du changement de couleur, une étude des dommages spécifiques des rayons X couplant des expériences de spectroscopie optique et de cristallographie aux rayons X.

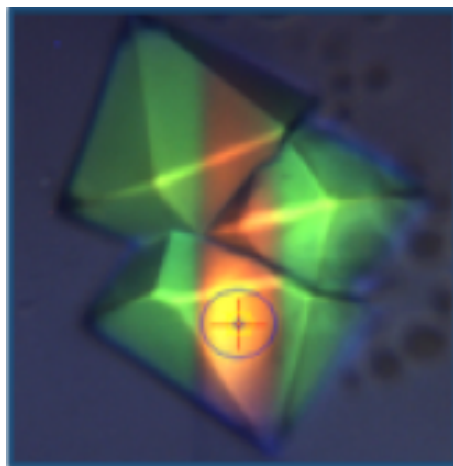


FIGURE IV.18 : Cristal de mNeonGreen observé sur la caméra de microdiffractomètre MD2 de la ligne de lumière de l'ESRF après une collecte de données diffraction (715 kGy) [de Sanctis et al., 2012]. Le faisceau de rayons X semble laisser une trace orange sur le cristal de mNeonGreen et bleu foncé dans le solvant, y signalant la présence d'électrons solvatés.

#### B.3.a Spectroscopie d'absorption UV-visible *in crystallo*

Une caractérisation spectroscopique en absorption UV-visible des zones irradiée et non irradiée a été réalisée en transportant un cristal de mNeonGreen dans l'azote liquide de la ligne de lumière ESRF-ID29 sur la plateforme ID29S-Cryobench [von Stetten et al., 2015]. La dose de radiation déposée dans le volume du cristal de mNeonGreen a été calculée à 389 kGy (figure IV.19). Le spectre correspondant à la zone non irradiée montre un ensemble de pics d'absorption entre 420 et 520 nm qui correspondent à la forme déprotonée du chromophore GYG. La forme globale du spectre UV-visible *in crystallo* du chromophore protoné est différente de ceux enregistrés en solution en raison de la saturation du pic, due à une densité optique voisine

de 2,0. Le spectre de la zone irradiée (389 kGY) montre la diminution de cet ensemble de pics au profit d'un nouveau pic centré autour de 384 nm. Ce pic rappelle la bande d'absorption de la forme protonée du chromophore que l'on observe à pH acide (figure IV.8).

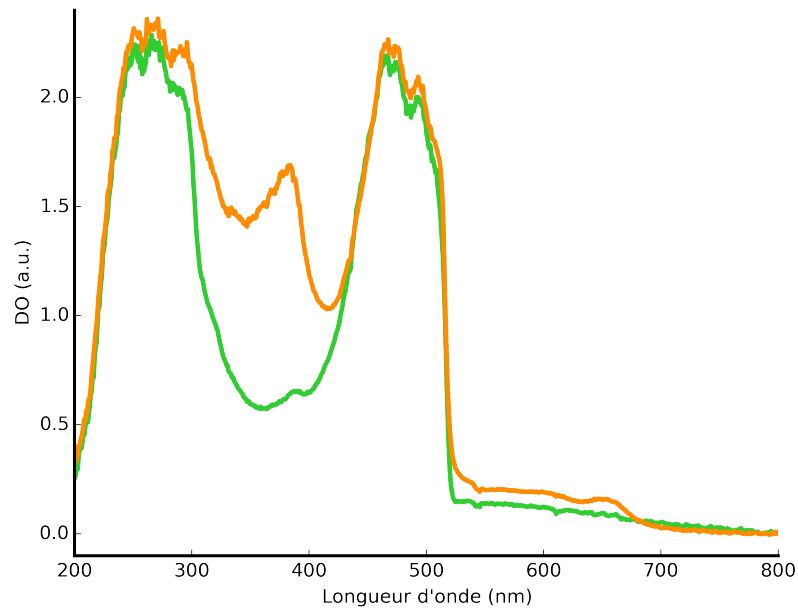


FIGURE IV.19 : Spectres d'absorption UV-visible des volumes de cristal de mNeonGreen irradié (orange) et non irradié (vert) par les RX

### B.3.b Spectroscopie Raman sur ligne de lumière

La caractérisation de l'effet des rayons X sur les cristaux de mNeonGreen a été complétée par une expérience de spectroscopie Raman en utilisant le montage dédié développé sur la ligne de ID29 de l'ESRF (en cours de publication von Stetten et al.) qui a permis d'enregistrer séquentiellement des spectres Raman avec des doses croissantes de rayons X.

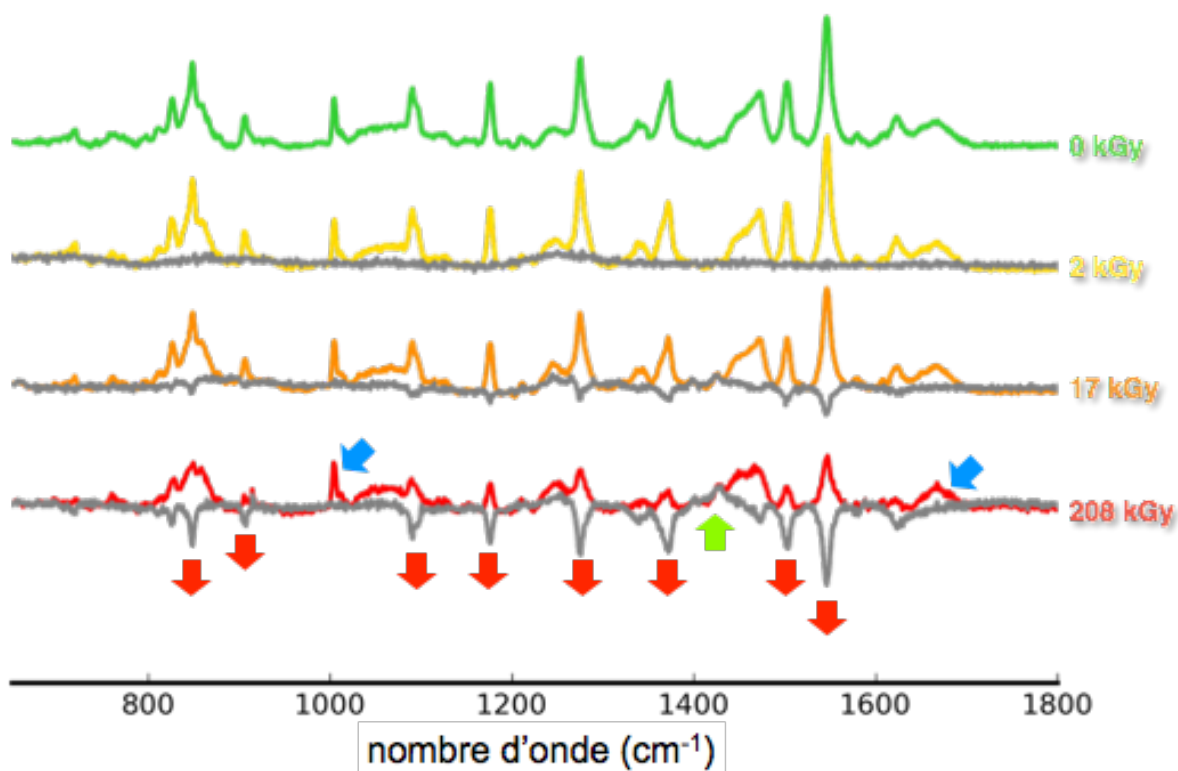


FIGURE IV.20 : Série de spectres Raman enregistrés sur un cristal de mNeonGreen avec une dose croissante de rayons X déposée. Le spectre de différence avec le spectre enregistré à 0 kGy ( $S_{0 \text{ kGy}} - S_{x \text{ kGy}}$ ) est représenté en gris.

Le spectre Raman de mNeonGreen ne montre pas de changement évident à 2 kGy. La modification des pics Raman les plus forts est claire à partir de 17 kGy et s'accroît avec la dose.

L'étude globale des spectres de différence Raman montre que quelques modes de vibrations visibles de la protéine, à  $1005 \text{ cm}^{-1}$  (mode des cycles de la phénylalanine) et  $1667 \text{ cm}^{-1}$  (mode des amides I) ne sont pas modifiés (flèches en bleu de la figure IV.20). La stabilité de ces modes de vibrations au cours de l'expérience suggère que la protéine n'est pas significativement affectée par les rayons X [von Stetten et al., 2015]. Les pics forts dominent ceux de la protéine

sont dus à un phénomène de pré-résonance du chromophore qui résulte de la relative proximité entre la longueur d'onde d'excitation (785 nm) et du maximum d'absorption de mNeonGreen (506 nm). Ainsi, les huit pics les plus intenses qui diminuent en fonction de la dose (flèches en rouge de la figure IV.20) sont tous attribués à des modes de vibration du chromophore (table IV.4).

Description du mode Raman	$\nu$ (cm <sup>-1</sup> )	Différence
Empreinte du chromophore <sup>a</sup>	848	-
Empreinte du chromophore <sup>a</sup>	906	-
Cycle de la phénylalanine <sup>b</sup>	1005	=
Empreinte du chromophore <sup>c</sup>	1090	-
Mode de flexion C-H des cycles phényles <sup>c</sup>	1176	-
Empreinte du chromophore <sup>c</sup>	1275	-
Déformation C-H des cycles phénol <sup>d</sup>	1371	-
Mode de cisaillement des CH <sub>2</sub> <sup>c</sup>	1428	+
Phénol 3, mode d'élongation C-C des cycles tyrosine <sup>a</sup>	1503	-
Mode d'élongation C=C exocyclique du chromophore <sup>a</sup>	1545	-
Amide I <sup>b</sup>	1667	=

Tableau IV.4 : Description des principaux modes de vibration en spectroscopie Raman pour les protéines fluorescentes, [Schellenberg et al., 2001]<sup>a</sup>, [von Stetten et al., 2015]<sup>b</sup>, [Bell et al., 2000]<sup>c</sup>, [He et al., 2002]<sup>d</sup>

### B.3.c Protocole de collecte de données de diffraction

Une série de 23 jeux de données de diffraction ont été collectés sur un cristal de mNeonGreen à pH physiologique. Chaque jeu de données a été collecté de façon quasi-identique avec la collecte de 400 images de diffraction avec une oscillation de  $0,5^\circ$  soit une exposition aux rayons X de 40 s avec  $E = 12,7$  KeV. La transmission du faisceau est le paramètre qui a été modifié au cours de l'expérience. Les dix premiers jeu de données ont été collectés avec une transmission de 0,51% et les six jeux suivants (n°11 à n°16) avec une transmission de 0,99%. A partir du jeu n°17, la transmission a été doublée pour chaque jeu de données jusqu'à atteindre 100% de transmission pour le jeu n°23 (figure IV.21).

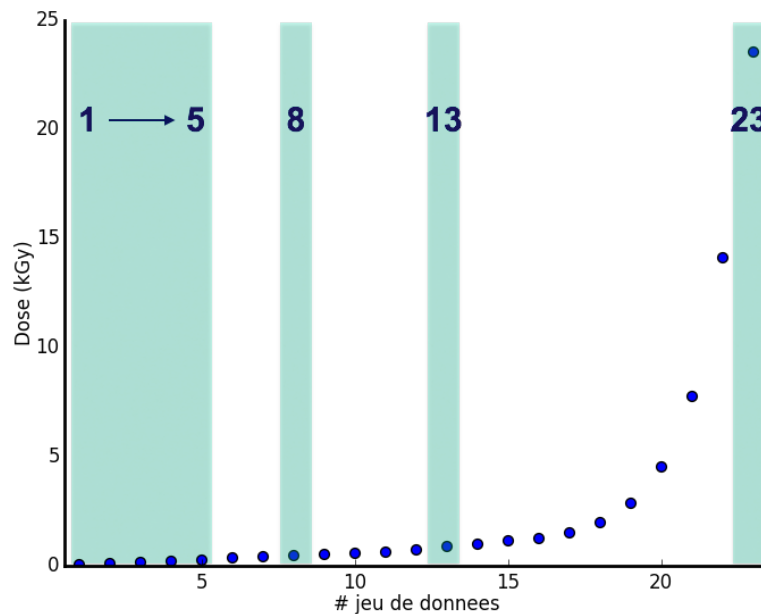


FIGURE IV.21 : Dose de rayons X accumulée en fonction du jeu de donnée. Les rectangles verts représentent les jeux de données exploités.

	1	2	3	4	5	8	13	23
# du jeu de données (n)	50	101	161	221	281	462	895	23556
Dose (kGy)								
<b>Paramètre de maille et statistique des données collectées</b>								
Longueur d'onde (Å)	0,976							
Température (K)	100							
pH	7,5							
Groupe d'espace	P6 <sub>5</sub> 22							
Maille élémentaire								
a, b, c (Å)	72,4 72,4 152,9	72,4 72,4 152,9	72,4 72,4 152,9	72,4 72,4 152,9	72,4 72,4 152,9	72,4 72,4 152,9	72,4 72,4 153,0	72,6 72,6 153,6
$\alpha, \beta, \gamma$ (°)	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0	90,0 90,0 120,0
Résolution (Å)	48,47 - 1,40 (1,44 - 1,40)*	48,47 - 1,40 (1,44 - 1,40)*	48,47 - 1,40 (1,44 - 1,40)*	48,47 - 1,40 (1,44 - 1,40)*	48,48 - 1,42 (1,46 - 1,42)*	48,47 - 1,40 (1,44 - 1,40)*	48,48 - 1,40 (1,44 - 1,40)*	39,69 - 1,44 (1,48 - 1,44)*
Nb. total de réflexions	976 228 (49 137)	976 388 (49 158)	979 900 (50 314)	980 190 (50 388)	950 157 (54 310)	980 014 (50 379)	977 355 (49 954)	889 393 (60 296)
Nb. de réflexions uniques	47 411 (3 424)	47 408 (3 424)	47 411 (3 424)	47 412 (3 424)	45 484 (3 279)	47 408 (3 224)	47 432 (3 418)	44 110 (3 220)
Multiplicité	20,6 (14,4)	20,6 (14,4)	20,7 (14,7)	20,7 (14,7)	20,9 (16,6)	20,7 (15,6)	20,6 (14,6)	20,2 (18,7)
Complétude (%)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)	100,0 (100,0)
I/ $\sigma$ moyen (I)	27,9 (1,9)	27,9 (1,8)	27,5 (1,8)	27,3 (1,7)	28,52 (2,05)	26,85 (1,66)	31,38 (2,13)	30,27 (2,04)
Facteur de Wilson (Å <sup>2</sup> )	22,3	22,5	22,5	22,6	22,8	22,8	23,0	28,7
$R_{meas.}^{\dagger}$	0,072 (1,389)	0,072 (1,431)	0,071 (1,378)	0,071 (1,508)	0,072 (1,363)	0,074 (1,562)	0,061 (1,272)	0,054 (1,585)
CC <sub>1/2</sub> <sup>#</sup>	1,000 (0,703)	1,000 (0,692)	1,000 (0,706)	1,000 (0,681)	1,000 (0,757)	1,000 (0,676)	1,000 (0,754)	1,000 (0,773)

Tableau IV.5 : Table de réduction des données cristallographiques collectées lors de l'expérience de dommages aux radiations menée sur un cristal de mNeonGreen. Les valeurs entre parenthèse sont pour les coquilles de plus haute résolution, (#) la coupure en résolution est basée sur le  $CC_{1/2}$ , (†)  $R_{meas.} = R_{merge} \times [N / (N - 1)]^{1/2}$ , où N est la multiplicité.



Des cartes de différence Fourier ont été calculées entre les jeux de données n°n et n°1 ( $F_{obs(n)} - F_{obs(1)}$ ). Ce calcul a été réalisé en utilisant les facteurs de structures de chacun des jeux de données ( $F_{obs(n)}$  et  $F_{obs(1)}$ ) et les phases calculées du modèle affiné avec le jeu de données n°1 de mNeonGreen ( $\Phi_{calc}$ ).

$$(F_{obs(n)} - F_{obs(1)}) e^{i\Phi_{calc}} \quad (IV.1)$$

L'observation des cartes de différence Fourier  $F_{obs(n)} - F_{obs(1)}$ , qui correspondent à des doses d'irradiation entre 50 et 844 kGy, permettent d'établir un scénario retraçant les effets structuraux des rayons X qui affectent l'environnement direct du chromophore chez mNeonGreen (figure IV.22 et voir VII.G). La hauteur des pics qui sont localisés sur les douze groupements chimiques les plus affectés de la structure de mNeonGreen sont listés dans la table IV.6.

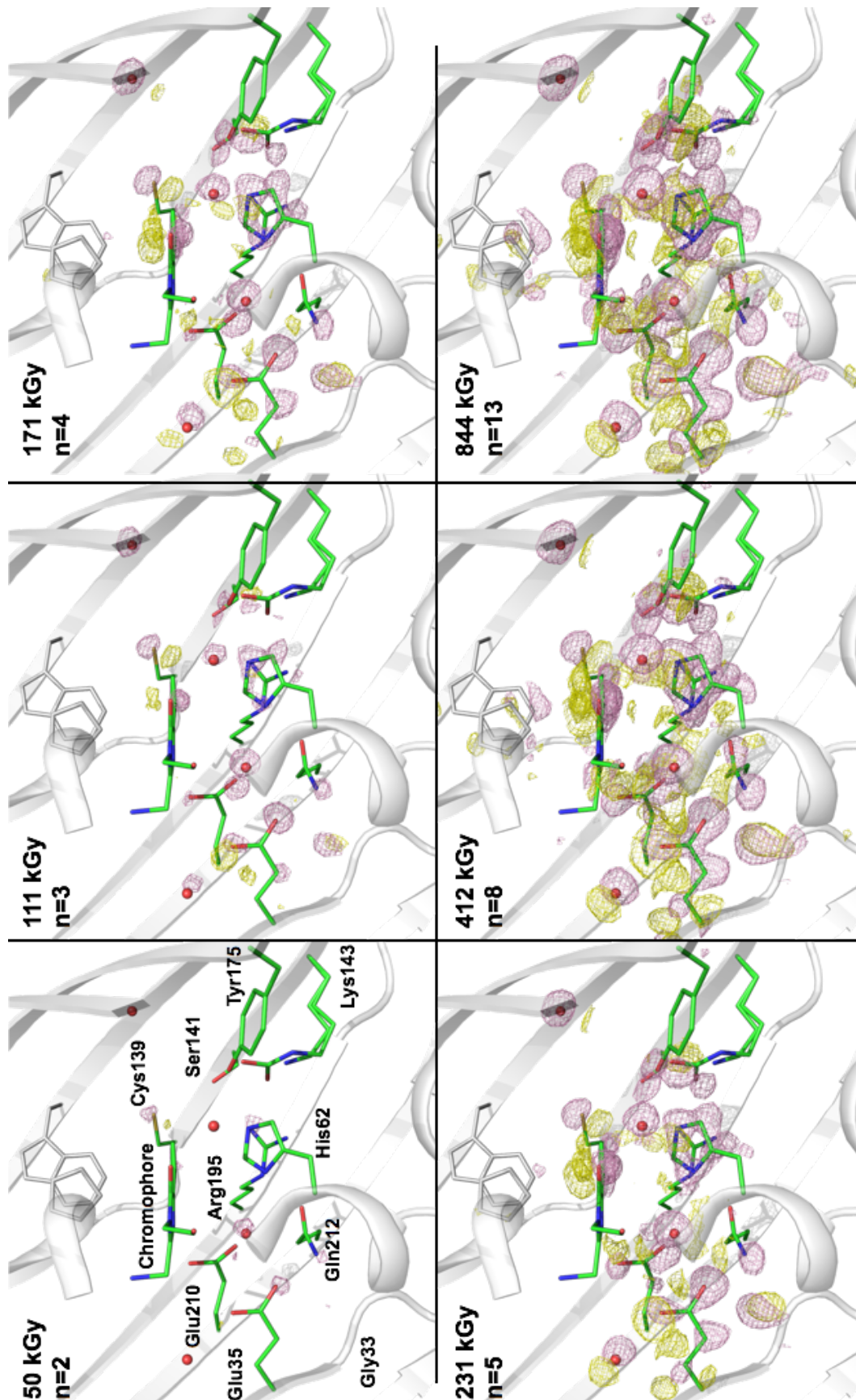


FIGURE IV.22 : Cartes de différence Fourier représentées au voisinage du chromophore entre 50 et 844 kGy. les doses de radiation sont modélisées par un niveau de contour de  $7,0 \times 10^{-2} e^{-} \cdot \text{\AA}^3$  ce qui correspond à  $6,0 \sigma$  pour le jeu n°2 (rose : négatif, jaune : positif).

Carte	$F_{o2}-F_{o1}$	$F_{o3}-F_{o1}$	$F_{o4}-F_{o1}$	$F_{o5}-F_{o1}$	$F_{o8}-F_{o1}$	$F_{o13}-F_{o1}$	$F_{o23}-F_{o1}$
Dose (kGy)	50	111	171	231	412	844	23500
$\sigma$	0,011	0,013	0,015	0,016	0,020	0,021	0,064
Pics négatifs intenses							
Glu35	-0,097	-0,174	-0,225	-0,256	-0,377	-0,464	-0,891
H <sub>2</sub> O ↔[Glu35-Glu210-His62]	-0,114	-0,152	-0,195	-0,255	-0,389	-0,500	-0,879
Glu210	-0,074	-0,134	-0,184	-0,231	-0,312	-0,392	-0,885
Cys139	-0,110	-0,162	-0,171	-0,211	-0,294	-0,373	-0,634
H <sub>2</sub> O ↔[Thr173-Ser153]	-0,071	-0,139	-0,190	-0,208	-0,286	-0,344	-0,492
His62	-0,084	-0,128	-0,161	-0,203	-0,281	-0,354	-0,634
Gly33	-0,071	-0,122	-0,151	-0,198	-0,275	-0,333	-0,653
Gln212	-0,089	-0,120	-0,150	-0,194	-0,248	-0,285	-0,402
Chromophore ↔(phénolate)	-0,077	-0,134	-0,165	-0,194	-0,275	-0,325	-0,589
H <sub>2</sub> O ↔[Chromophore (phénolate)-Tyr175]	-0,060	-0,095	-0,131	-0,176	-0,280	-0,281	-0,582
Lys143	-0,055	-0,086	-0,136	-0,173	-0,242	-0,279	-0,415
Chromophore ↔(imidazolinone)	-0,057	-0,103	-0,111	-0,148	-0,219	-0,294	-0,447
Pics localisés sur le pont méthylène							
Négatif	-0,049	-0,062	-0,108	-0,124	-0,188	-0,221	-0,383
Positif ↔(configuration sp <sup>2</sup> )	0,035	0,058	0,076	0,079	0,120	0,154	0,370
Positif ↔(configuration sp <sup>3</sup> )	0,026	0,050	0,066	0,076	0,099	0,117	0,171

Tableau IV.6 : Intensité des pics observés en  $e^- \cdot \text{\AA}^3$  dans les cartes de différence Fourier où  $n$  est le numéro du jeu de donnée enregistré par diffraction aux rayons X au cours de l'expérience de dommages aux radiations.

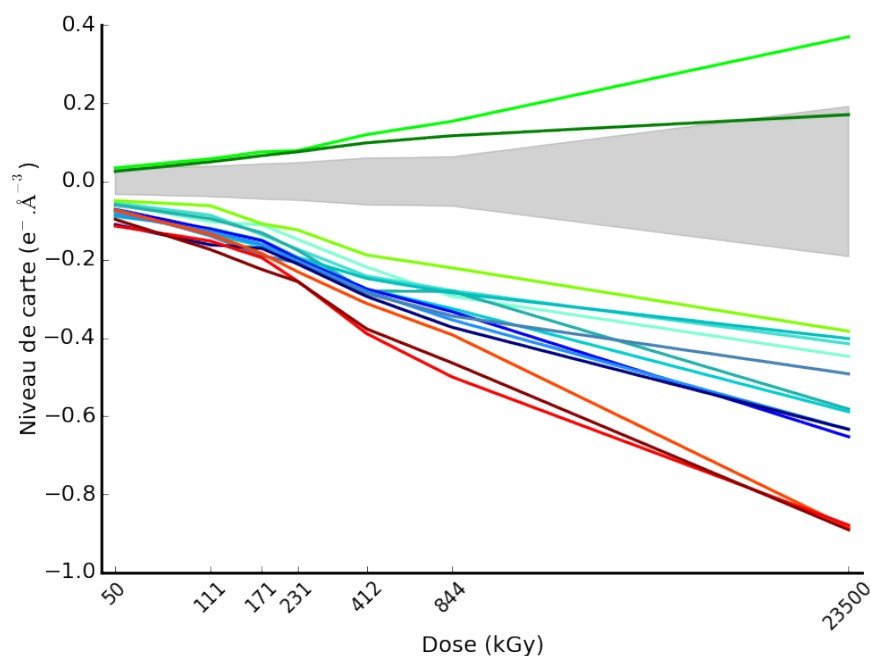


FIGURE IV.23 : Gain et perte d'électrons induit en  $e^- \cdot \text{\AA}^3$  en fonction de la dose de radiations pour différents groupe de la cavité du chromophore : en rouge, les trois pics les plus forts, en vert, pics associés au pont méthylène, en bleu, les autres pics forts.

Les pics de densités les plus forts sont localisés d'un seul côté du chromophore (figure IV.22). L'évolution des trois pics de loin les plus forts sont représentés en nuances de rouge dans la figure IV.23. Ils concernent deux glutamates à proximité immédiate du chromophore Glu35 et Glu210 ainsi qu'une molécule d'eau qui leur est associée par des liaisons hydrogène. Trois pics (un négatif, deux positifs) associés au pont méthylène du chromophore sont représentés en nuances de vert. Le pic négatif ainsi qu'un pic positif se détachent progressivement du bruit, tandis que le second pic positif émerge du bruit pour y retomber à haute dose. Enfin, les autres pics les plus forts de la cavité sont représentés en nuances de bleu. On retrouve dans ce groupe les résidus : Gly33, His62, Cys139, Lys143, Gln212. A cela nous avons ajouté deux molécules d'eau ainsi que les cycles phenol et imidazolinone du chromophore.

### **B.3.d Discussion**

L'examen des pics les plus forts suggèrent que la premiers dommages spécifiques des rayons X consistent en la décarboxylation des glutamates 35 et 210 avec un effet légèrement plus marqué pour Glu35. Cependant, on observe des pics positifs sous le chromophore qui sont compatibles avec des conformations alternées pour Glu35 et Glu210. Cela peut s'interpréter par l'endommagement alterné de l'un des deux glutamates (avec une préférence pour Glu35) qui entraîne la réorientation de la chaîne latérale de l'autre. L'endommagement aux rayons X de cette protéine fluorescente de *Branchiostoma* montre une modification de l'endommagement observé pour les PF de cnidaires, où seul l'équivalent de Glu210 (Glu222 chez AvGFP) est affecté [van Thor et al., 2002, Adam et al., 2009, Royant and Noirclerc-Savoie, 2011].

L'évolution des autres pics forts (représentés en bleu dans la figure IV.23) permet d'appréhender les changements structuraux qui découlent de la décarboxylation. La molécule d'eau fortement affectée près de glutamates est liée à His62, qui logiquement, est déstabilisée entraînant Lys143, Arg195 et Gln212. Le soufre de Cys139, de part son nombre d'électrons élevés, permet de détecter rapidement le déplacement du chromophore. Pendant que la chaîne latérale de Cys139 se réoriente, l'ensemble du chromophore apparaît se déplacer d'un bloc (flèche en bleu de la figure IV.24) vers la zone non affectée par les rayons X.

Cependant, on observe deux pics positifs de part et d'autre du pont méthylène. Le premier (flèche bleu de la figure IV.24) correspond au déplacement en bloc du chromophore, l'autre (flèche en rouge de la figure IV.24) indique un déplacement opposé de l'atome central du pont méthylène. Cela rappelle le résultat décrit pour l'endommagement à faible dose de rayons X

de la protéine Iris FP [Adam et al., 2009] qui montre une perte de planarité du chromophore compatible avec la dehybridation du  $C_\beta$  des orbitales  $sp^2$  en orbitales  $sp^3$  [Roy et al., 2011]. On note que cette modification structurale n'est plus valable à haute dose (figure IV.23) ce qui suggère que l'on a visualisé le même mécanisme dans mNeonGreen, en parallèle de la réorganisation définitive de la cavité du chromophore.

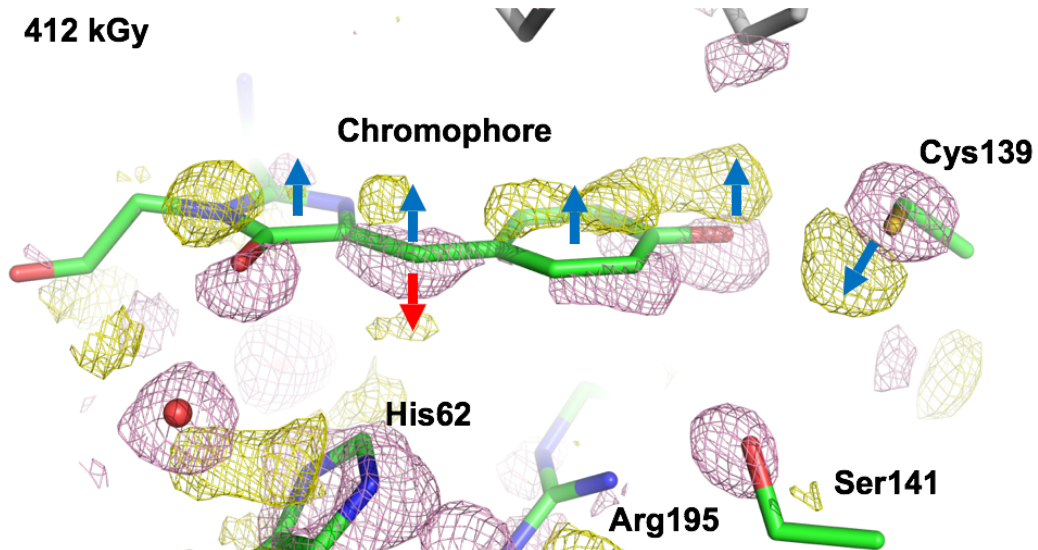


FIGURE IV.24 : Changements structuraux induits par les rayons X dans l'environnement du chromophore de mNeonGreen, avec une modélisation à un niveau de  $4,0 \sigma$  après une irradiation de 412 kGy. Les flèches en bleu indiquent le déplacement du chromophore et du mouvement opposé du résidu Cys139 tandis que la flèche en rouge suggère le passage transitoire de l'hybridation  $sp^2$  à  $sp^3$  du pont méthylène.

## C Analyse de l'environnement du chromophore par calcul d'énergie d'interaction

### C.1 But de l'analyse

Si l'on souhaite améliorer les propriétés de fluorescence d'une PF dont on a résolu la structure, il est nécessaire en premier lieu d'analyser la façon dont le chromophore est stabilisé, ou au contraire déstabilisé, par les résidus environnants. Une structure cristallographique permet d'appréhender relativement aisément les interactions de type liaison hydrogène, empilement  $\pi/\pi$  ou électrostatique, qui peuvent être envisagées comme des interactions ponctuelles. Par contre les interactions de van der Waals (vdW) sont plus difficiles à évaluer car elles sont faibles et réparties sur une large surface d'interaction. La structure cristallographique, qui est statique, n'est pas suffisante pour appréhender avec précision les termes énergétiques d'interaction de vdW entre acides aminés et chromophore [Lelimosin et al., 2009]. Une fois un champ de force entre le chromophore et son environnement protéique disponible, nous pouvons effectuer des simulations par dynamique moléculaire, et déduire des valeurs d'énergie d'interaction moyennées au cours du temps.

### C.2 Principe de l'analyse

J'ai choisi de limiter l'analyse des interactions entre le chromophore et les résidus environnants aux résidus possédant un atome à moins de 5 Å d'un atome du chromophore. Pour ce faire, j'ai utilisé le logiciel de visualisation graphique PyMOL (<https://www.pymol.org>) pour identifier ces résidus dans les structures de *lanYFP* et *mNeonGreen* à pH physiologique. Ces résidus sont au nombre (n) de 26 dans *lanYFP* et 20 dans *mNeonGreen*. Comme il est également important de considérer les interactions entre résidus voisins, le nombre d'interactions à considérer est donc de 325 et 190 (selon  $\frac{n(n-1)}{2}$ ). Pour réduire le nombre conséquent de ces interactions, j'ai réduit la liste des résidus en identifiant les paires de résidus situés à moins de 5 Å l'un de l'autre, ce qui réduit la liste à 111 et 67 interactions.

Dans les simulations par dynamique moléculaire, j'ai utilisé un critère de coupure ( $r_{coupure}$ ) des énergies non-liées (énergies d'interactions de vdW et électrostatiques) ayant pour valeur  $r_{coupure} = 10$  Å, ce qui permet de considérer ces énergies comme nulles au delà de  $r_{coupure}$  (figure IV.25). Cette approximation n'est pas valable pour certaines interactions électrostatiques, ce qui

est corrigé en appliquant une méthode du type somme de réseau (*Particle Mesh Ewald*). Dans le cas présent, je m'intéresse aux interactions pour des distances inférieures à 5 Å où cette approximation se justifie pour l'ensemble des interactions électrostatiques (à noter qu'on utilise parfois un paramètre de lissage linéaire pour que le terme d'énergie soit continu à la coupure).

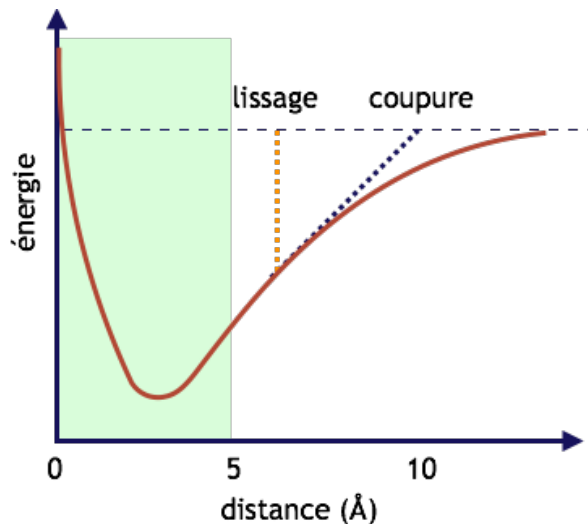


FIGURE IV.25 : Energie d'interaction de van der Waals avec l'effet des paramètres de coupure (ligne pointillé bleue) et de lissage (ligne pointillée orange)

### C.3 Principe d'extraction des énergies d'interaction de van der Waals

Le module *NAMD*Energy disponible dans la suite logiciel *VMD* [Humphrey et al., 1996] permet d'extraire les énergies d'interaction de van der Waals (vdW) pour une sélection de paires de résidus au sein d'une simulation par dynamique moléculaire de 30 ns. L'extraction des énergies d'interaction de vdW est effectuée paire par paire, pour toute la durée de la simulation. Les énergies d'interaction de vdW se présentent sous la forme d'une matrice bidimensionnelle qui représente la carte d'interaction entre paires de résidus. Les matrices peuvent être visualisées sous deux formes. Dans la première, on regarde la moyenne au cours de la dynamique, ce qui permet d'en apprécier leur force durant la simulation (matrice de force de stabilité, figure IV.27). Dans la seconde, on considère la densité de probabilité d'interaction déstabilisante (valeurs positives) ce qui renseigne sur la stabilité de ces interactions (matrice de fréquence de déstabilisation, figure IV.27).

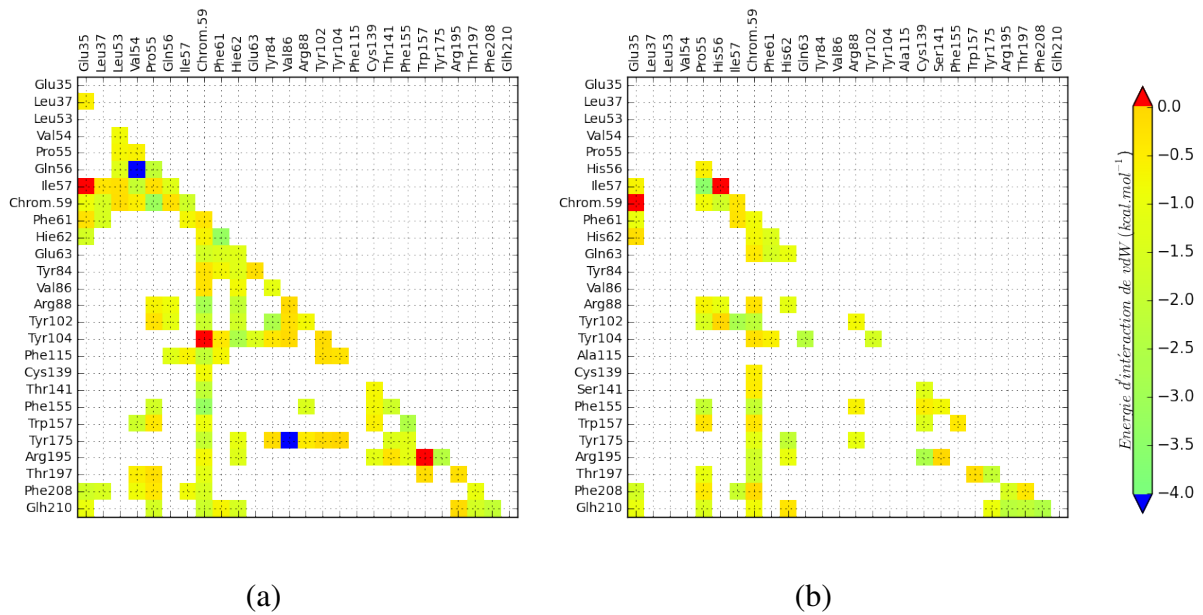


FIGURE IV.26 : Matrice des paires d'interaction selon la valeur moyenne des énergies d'interaction de van der Waals, (a) *lanYFP* et (b) *mNeonGreen*

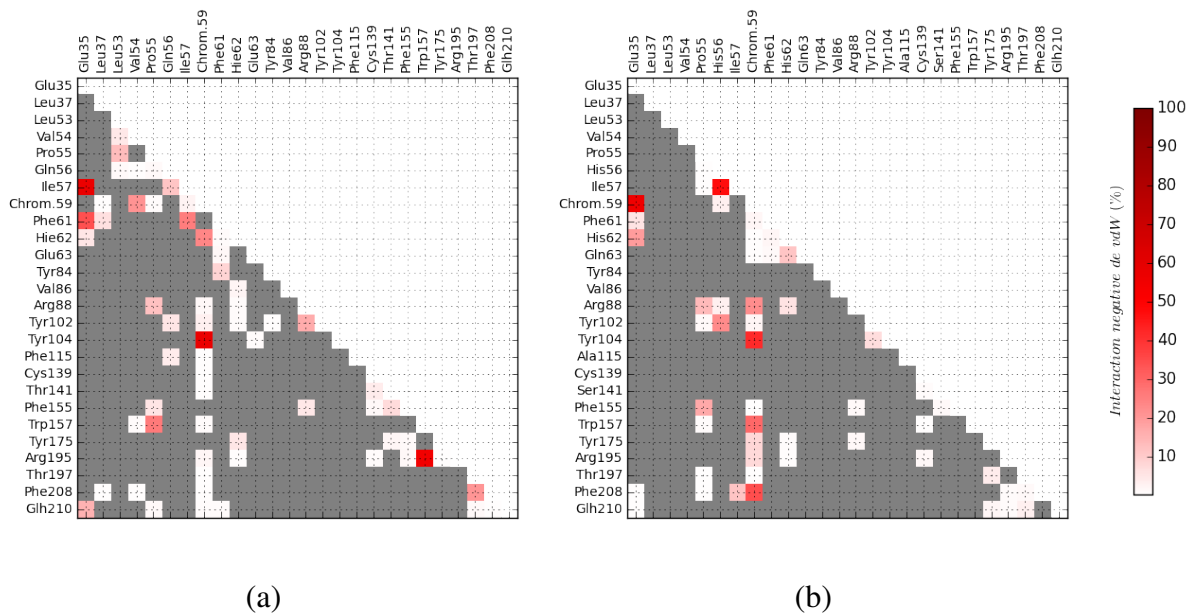


FIGURE IV.27 : Matrice des paires d'interaction selon la densité de probabilité des énergie d'interaction déstabilisantes de van der Waals (a) *lanYFP* et (b) *mNeonGreen*

## C.4 Interprétation

La somme des énergies d'interaction entre le chromophore et les résidus environnants est de  $-16,2 \text{ kcal.mol}^{-1}$  pour *mNeonGreen* et  $-32,0 \text{ kcal.mol}^{-1}$  pour *lanYFP*. Cela était attendu en raison de la meilleure efficacité de fluorescence du chromophore de *lanYFP*. Il y a trois mutations entre *lanYFP* et *mNeonGreen* qui apparaissent dans les matrices : Q56H, F115A et T141S. La muta-



tion Q156H affaiblit une interaction très stabilisante entre le résidu 56 et la valine 54 ainsi qu'une interaction stabilisante avec Pro55, située juste au-dessus du chromophore. La mutation F115A enlève une interaction stabilisante du chromophore dans mNeonGreen par rapport à *lanYFP*. Enfin, la mutation T141S affaiblit une interaction stabilisante avec le chromophore (de -2,1 à -0,5 kcal.mol<sup>-1</sup>). Les matrices de force de stabilisation montrent que trois paires de résidus sont déstabilisantes en moyenne chez *lanYFP* : Ile57/Glu35, Tyr104/Chrom.59 et Arg195/Trp157, et deux particulièrement stabilisantes : Gln46/Val54 et Tyr175/Val86. Dans le cas de mNeonGreen, il n'y a que deux interactions particulièrement déstabilisantes : Chrom.59/Glu35 et Ile57/His56, tandis qu'il n'y a pas d'interaction particulièrement stabilisante. Si on s'en tient aux interactions de van der Waals stabilisantes et déstabilisantes de plus forte amplitude dans *lanYFP*, on s'aperçoit qu'elles ont été réduites dans mNeonGreen. Les matrices de fréquence de déstabilisation montrent que dans *lanYFP*, il y a trois interactions déstabilisant de temps en temps le chromophore, et ce nombre grimpe à sept pour mNeonGreen, ce qui suggère que l'évolution dirigée a significativement affectée la stabilisation du chromophore par les interactions de van der Waals.

## D Conclusion

Au cours de cette partie de thèse, j'ai résolu la structure de des protéines fluorescentes jaune et jaune-vert *lanYFP* et mNeonGreen. Cela m'a permis de rationaliser le processus d'évolution dirigée qui a mené de la protéine tétramérique *lanYFP* à la protéine monomérique mNeonGreen.

Le "verdissement" entre *lanYFP* et mNeonGreen s'explique par un ensemble de mutations nécessaire pour restaurer les propriétés de fluorescence après monomérisation, mais qui ont entraîné un déplacement de quelques résidus sur l'hélice portant le chromophore et modifié significativement une interaction clef avec le nuage électronique délocalisé du chromophore. L'étude structurale de mNeonGreen m'a conduit à m'interroger sur les effets spécifiques des rayons X sur la structure de la protéine. Une étude corrélant diffraction aux rayons X et spectroscopie d'absorption UV-visible et Raman m'ont permis de montrer un réarrangement massif des résidus de la cavité du chromophore et soit un déplacement en bloc soit la perte de la conjugaison du chromophore au niveau du pont méthylène. J'ai mis en évidence un site de fixation par les ions chlorures, qui est partiellement aboli dans mNeonGreen à pH physiologique par une modification post-traductionnelle spontanée, la carboxylation d'une lysine à proximité du chromophore. Enfin, nous avons exploré le paysage énergétique de van der Waals entre le chromophore et un proche environnement afin de comprendre comment ces interactions subtiles étaient modifiées entre les deux protéines.

# V CHAPITRE

## Perspectives

La recherche et la visualisation des réseaux de liaisons hydrogènes est un problème complexe qui nécessite une représentation concise et interactive. L'algorithme d'extraction des données qui a été écrit (langage Fortran) et son traitement de données (langage Python) complexifie la procédure d'utilisation. Un effort supplémentaire est à fournir pour que l'algorithme devienne interactif, simple d'utilisation et que l'ensemble des réseaux de liaisons hydrogène soient étudiés. Dans le cadre d'un développement logiciel, il serait intéressant d'écrire un module pour un logiciel de représentation moléculaire graphique tel que PyMOL [DeLano, 2002], SAM-SOM [Artemova and Redon, 2012] ou VMD [Humphrey et al., 1996]. Les coupes sagittales permettent une représentation originale des réseaux de liaisons hydrogène. Néanmoins cette représentation nécessite d'être perfectionnée avec l'ajout de la densité de probabilité effective de chaque liaison hydrogène (non modulée par le réseau) et de discriminer d'une part, les liaisons hydrogène qui sont impliquées dans la stabilité des structures tertiaires (carbonyle et azote de la chaîne principale) et d'autre part, celles effectivement impliquées dans de potentiels transferts de proton. Enfin, des applications sur des systèmes biologiques comportant des transfert de protons sont à envisager comme c'est le cas dans la bactériorhodopsine, le complexe IV de la chaîne respiratoire mitochondriale et le centre réactionnel bactérien. Par ailleurs, l'étude de la dynamique des réseaux de liaisons hydrogènes appliquée aux protéines fluorescentes jaunes pourrait suggérer des pistes pour leur ingénierie et ainsi améliorer leur propriétés photophysiques.

Le versant structural de cette thèse aura été marqué par la compréhension du processus de monomérisation de *lanYFP*, protéine fluorescente jaune naturelle tétramérique, en mNeonGreen, protéine fluorescente jaune-vert monomérique. Au vu des caractéristiques photophysiques exceptionnelles de *lanYFP* et de son mutant mNeonGreen, il est intéressant d'envisager le transfert

des déterminants de fluorescence vers d'autres protéines peu fluorescentes afin de maximiser leur brillance. Par ailleurs, certaines propriétés biochimiques comme le  $pH_{1/2}$  se sont dégradées au cours de l'ingénierie de mNeonGreen. L'étude de la dynamique des réseaux de LH de *lanYFP* et mNeonGreen pourrait fournir de précieux indices sur la modulation du  $pH_{1/2}$  et de rationaliser sa dégradation et ainsi envisager un mutant de mNeonGreen bien moins sensible aux différents pH intra-cellulaires. L'amélioration du  $pH_{1/2}$  de mNeonGreen permettrait de lui conférer un avantage supplémentaire non négligeable sur l'ensemble des autres protéines fluorescentes jaune-vert et jaune. Les récents développements en séquençage haut débit et les différentes expéditions scientifiques qui ont été entreprises au cours de la dernière décennie (expédition Tara Pacific 2016 sur la grande barrière de corail en cours) permettent d'envisager la découverte d'organismes portant des nouveaux gènes de protéines fluorescentes. La combinaison des techniques de mutagenèse haut débit couplées à une connaissance accrue des espèces marines potentiellement porteuses de nouveaux gènes codant pour des protéines fluorescentes est sans nul doute la voie royale pour identifier de nouvelles protéines fluorescentes à haute brillance qui seront ensuite ingénierées pour les besoins spécifiques de la microscopie (monomérisation,  $pH_{1/2}$ ...).

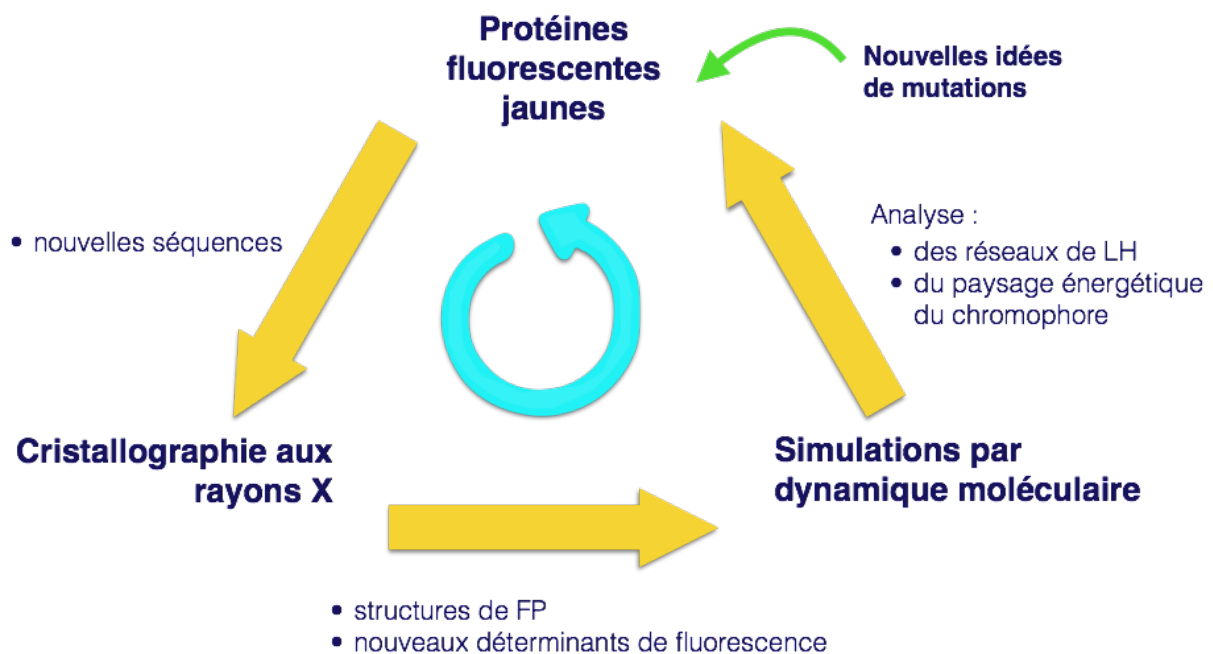


FIGURE V.1 : Schéma synoptique d'amélioration des protéines fluorescentes jaunes intégrant les méthodes discutées au cours de la thèse.

L'ensemble des résultats obtenus au cours de cette thèse permet de proposer un cadre à la fois théorique et expérimental pour la compréhension des défauts des différentes protéines fluorescentes jaunes (brillance, sensibilité au pH, sensibilité aux ions chlorure). Les outils et méthodes développés vont mettre en évidence différentes pistes de mutagenèse pour remédier à ces défauts. C'est sans doute l'intégration de l'ensemble de ces méthodes et leur application à différentes PF jaune qui mènera à l'élaboration de "la" protéine fluorescente étalon dans cette gamme spectrale.

# VI CHAPITRE

## Méthodes

### A Expression et purification

#### A.1 Expression de protéine recombinante

##### A.1.a Description du gène des protéines fluorescentes de *Branchiostoma lanceolatum*

Les constructions de mNeonGreen et *lanYFP* ont été réalisées par Nathan Shaner. Les séquences codantes ont été insérées à la suite d'une série d'étiquettes permettant à la fois de ne pas altérer la maturation des protéines ainsi que d'assurer leur purification. Une étiquette poly-histidine composée de six histidines (HHHHHH, 0,84 kDa) a été insérée à l'extrémité du N-terminal suivi d'un épitope T7 tag (MASMTGGQMG, 1,10 kDa) couramment utilisé pour détecter et purifier les protéines. On trouve une seconde étiquette d'affinité aux anticorps nommé Xpress<sup>TM</sup> (DLYDDDDK, 0,99 kDa). Les extrémités N-terminal (MVSKGEE, 0,78 kDa) et C-terminal (GMDELYK, 0,86 kDa) de EGFP, composées de sept acides aminés chacune, ont été ajoutées à la séquence de *lanYFP* selon les pratiques de biologie moléculaire en vigueur pour l'élaboration de nouvelles PF [Shaner et al., 2004]. Enfin une séquence peptidique de liaison DNMA (0,44 kDa) a été intégrée à la séquence pour sa capacité à augmenter le repliement et la localisation. Le clonage de la séquence codante de *lanYFP* a été effectué dans un vecteur pNCS [Yueh and Goff, 2003] qui comporte une cassette de résistance à l'ampicilline (amp<sup>r</sup>) entre les sites de restriction BamH1 et EcoR1. La construction de mNeonGreen est équivalente à celle de *lanYFP* (figure VI.1) mais elle comporte une délétion de deux acides aminés dans la séquence de liaison entre l'étiquette d'affinité aux anticorps Xpress<sup>TM</sup> et l'extrémité du N-terminal ajoutée de EGFP.

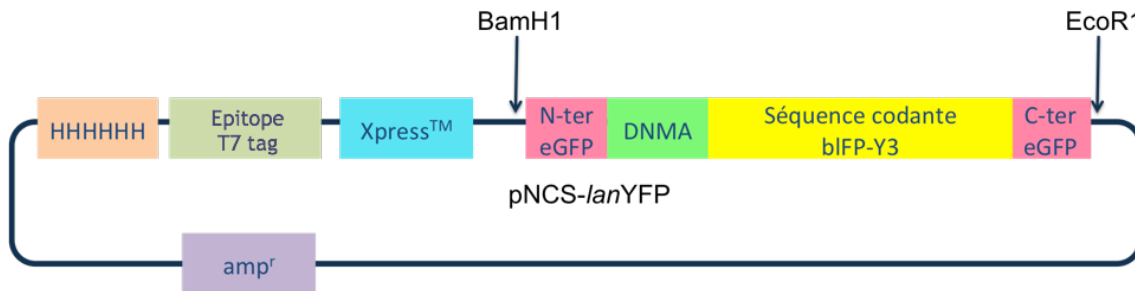


FIGURE VI.1 : Plasmide pNCS-lanYFP codant pour la protéine fluorescente jaune lanYFP.

### A.1.b Expression des protéines recombinantes

Les protocoles de biologie moléculaire précisent que les systèmes d'expression bactériens pour l'amplification des copies de plasmides et de production de protéines recombinantes sont différents. Le vecteur pNCS permet une expression constitutive du gène codant pour une protéine avec la machinerie de *Escherichia coli* [Yueh and Goff, 2003]. Ainsi ce vecteur est particulièrement adapté à l'évolution des PF puisqu'il ne nécessite pas d'induire l'expression avec des molécules exogènes comme l'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). De plus la machinerie utilisée étant celle d'*E. coli*, les rendements obtenus sont importants ( $> 30 \text{ mg}\cdot\text{L}^{-1}$  de culture). L'expression des protéines recombinantes mNeonGreen et lanYFP a été réalisée avec les bactéries DH5 $\alpha^{TM}$  [Woodcock et al., 1989].

La première étape nécessaire à l'amplification de l'ADN et à la production des protéines est la transformation bactérienne. Un aliquot de  $50 \mu\text{L}$  de bactérie compétente DH5 $\alpha^{TM}$  mélangé avec  $1 \mu\text{L}$  de plasmide à  $100 \text{ ng}\cdot\mu\text{L}^{-1}$  est mis sur glace durant 20 min dans un tube stérile de 1,5 mL. Un choc thermique à  $42^\circ\text{C}$  est réalisé dans un bain chaud durant 45 s. L'aliquot est ensuite refroidi sur glace durant 2 min 30 s avant d'être resuspendu avec  $400 \mu\text{L}$  de milieu *Super Optimal Broth* enrichi en glucose (SOC). Ce tube est placé dans un bloc thermique à  $37^\circ\text{C}$  durant 1 h à  $37^\circ\text{C}$  avec une agitation à 200 rpm (Thermomixer Compact, Eppendorf, Hambourg, All.). Le milieu de culture est centrifugé à 1000 g et le surnageant est renouvelé par du milieu SOC frais. Les bactéries sont étalées sur boîte de Petri avec un milieu LB agar / ampicilline et laissées en étuve durant 12 h.

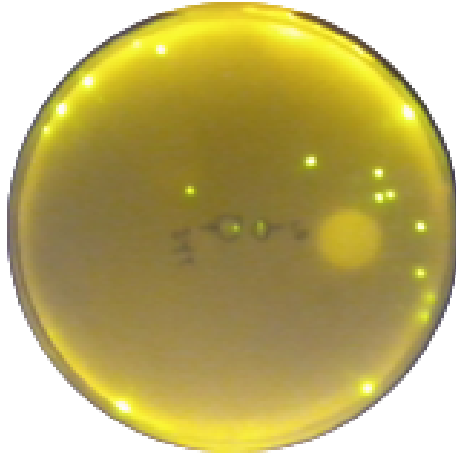


FIGURE VI.2 : image boîte de petri sous lumière bleue

Les précultures sont réalisées avec les bactéries compétente DH5 $\alpha^{TM}$ , en piquant une colonie unique dans 50 mL de milieu 2YT contenant de l'ampicilline (concentration finale 100  $\mu\text{g.mL}^{-1}$ ). Cette phase permettant la sélection dure 12 h dans une étuve à 37°C avec une agitation de 150 rpm.

La production de protéines recombinantes s'effectue avec des erlenmeyers de 2 L à ailettes contenant 500 mL de milieu de culture 2YT et la même concentration finale en ampicilline, estensemencé par 25 mL de préculture (ratio 1 / 21). La culture est généralement conduite sur la nuit (14 - 16 h). Les bactéries sont culotées à 5000 g (rotor à angle fixe JLA-8.1000, Beckman Coulter, Brea, USA) durant 20 min à 4°C. Le culot de bactéries est resuspendu dans une solution de tampon de lyse contenant 20 mM de tris HCl à pH 8,0, 20 mM d'imidazole, 5 mg de lysozyme et 250 mM de chlorure de sodium. Les processus de digestion enzymatique sont ralentis par l'ajout d'une pastille d'inhibiteur de protéase (cOmplete EDTA-free, ROCHE, Bâle, Suisse). Les bactéries sont congelées à -80°C et décongelées à 37°C afin de favoriser l'éclatement des bactéries dont les parois cellulaires ont été entamées par le lysosyme. La lyse des bactéries est finalisée avec l'utilisation d'un broyeur cellulaire (Constant Systems Limited, Daventry, GB) réglé à 950 bar, à 7°C. Les bactéries broyées sont centrifugées à 43000 g (rotor à angle fixe JA-25.50, Beckman Coulter, Brea, USA) durant 25 min à 4°C. Le surnageant obtenu est immédiatement purifié afin d'éviter tous phénomènes de protéolyse non souhaités.

## A.2 Purification de protéine recombinante

L'étape de purification des protéines met en oeuvre deux méthodes de chromatographie successives. La première fait appel à l'étiquette poly-histidines et consiste en une étape de chro-



matographie d'affinité par chélation de métaux immobilisés (IMAC) sur une résine sépharose (HisTrap HP 5 mL, GE Healthcare, Little Chalfont, GB). La seconde consiste en une chromatographie d'exclusion de taille (Superdex 75 10/300 GL, GE Healthcare, Little Chalfont, GB). Toutes les étapes de purification ont été réalisées sur des automates ÄKTA prime et ÄKTA purifier (GE Healthcare, Little Chalfont, GB).

La chromatographie IMAC permet de conserver uniquement les protéines qui ont une étiquette poly-histidine en jouant sur l'affinité de cette étiquette avec les ions  $\text{Ni}^{2+}$  de la phase stationnaire. Les protéines possédant une étiquette poly-histidines sont éluées progressivement de la phase stationnaire en augmentant la concentration d'imidazole dans le tampon d'élution. Les protéines mNeonGreen et *lanYFP* se détachent de la colonne Ni-sépharose à partir de 35% d'imidazole à 4°C avec un flux de 1 mL.min<sup>-1</sup>. Les fractions collectées sont contrôlées par électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium (SDS-PAGE) à 12,5% pour vérifier la qualité de la première étape de purification.

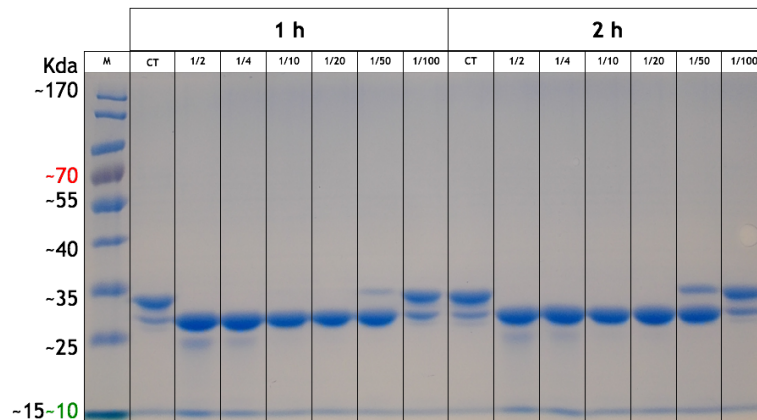
Les fractions purifiées issues de la chromatographie IMAC sont concentrées jusqu'à un volume de 500  $\mu\text{L}$  avant d'être injectées sur une colonne d'exclusion de taille via une boucle de chargement de 500  $\mu\text{L}$ . La chromatographie d'exclusion est réalisée avec un tampon d'élution contenant 20 mM de tris HCl à pH 8,0 avec un flux de 0,80 mL.min<sup>-1</sup>. La pureté des fractions correspondant au pic d'absorption à 280 nm et colorées sont contrôlées avec un gel d'acrylamide à 12,5%.

### A.3 Protocole de digestion à la trypsine

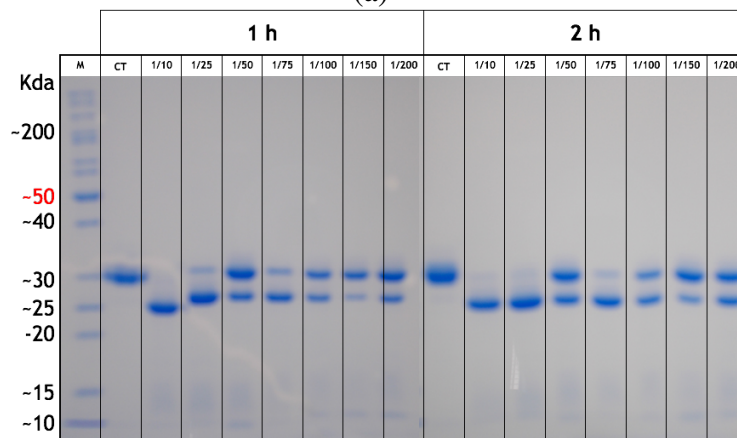
Les étiquettes utilisées au cours de la purification sont largement non-structurées et peuvent constituer une gêne pour la cristallogénèse des protéines. Afin de cliver ces parties flexibles, une étape de digestion aménagée à la trypsine peut être réalisée (Sigma-Aldrich, Saint Louis, USA). En particulier, la cristallogénèse des protéines fluorescentes étudiées dans ce manuscrit peuvent être améliorées par cette étape qui permet l'élimination des zones flexibles et de dégradation des protéines dont le repliement n'est pas optimal. Le protocole de digestion à la trypsine consiste en une phase de criblage permettant de déterminer les conditions optimales pour la digestion. Des ratios massique trypsine / protéine (1/5 à 1/200) sont testés durant 1h et 2 h à température ambiante (25°C). La meilleure condition est déterminée par SDS-PAGE en utilisant un gel acrylamide à 12,5%. Cette condition est pour *lanYFP* et mNeonGreen, une digestion tryptique de 1h à 25°C avec un ratio trypsine / protéine de 1/10. La digestion est ralentie en

plongeant les tubes de réaction dans un bain de glace voir en stockant les échantillons à  $-20^{\circ}\text{C}$ .

Afin de se débarrasser de la trypsine contaminante, la protéine est purifiée par chromatographie d'exclusion de taille (Superdex 75 10/300 GL, GE Healthcare, Little Chalfont, GB). Il est cependant possible, même si la pratique peut paraître peu conventionnelle, d'introduire directement la trypsine dans la goutte de cristallisation. Les deux techniques permettent d'obtenir des cristaux de mNeonGreen (voir figure IV.7.b).



(a)



(b)

FIGURE VI.3 : SDS-PAGE de contrôle de la digestion trypsique pour lanYFP (a) et mNeonGreen (b). CT correspond au puit contrôle avec la protéine sans ajout de trypsine, les ratios trypsine/protéine sont en masse et le temps de coupure est indiqué au dessus des puits.

## B Détermination de la structure de protéines par cristallographie aux rayons X

### B.1 La cristallisation des protéines fluorescentes

#### B.1.a Généralités concernant les cristaux et la cristallisation

L'étymologie du mot cristal provient du grec ancien  $\kappa\rho\upsilon\sigma\tau\alpha\lambda\lambda\omicron\zeta$  signifiant "glace". L'état cristallin est une forme énergiquement favorable de la matière. Les cristaux sont des solides composés d'atomes, d'ions ou de molécules agencés de façon ordonnée et répétée formant une structure microscopique tridimensionnelle nommée réseau de Bravais. Chaque unité indépendante et équivalente constituant un cristal est appelée unité asymétrique. Il est à noter qu'il faut distinguer les monocristaux où chacune des unités sont agencées de façon cohérente au sein de la maille cristalline, les polycristaux qui vont être un mélange de plusieurs mailles et l'empilement amorphe de matériaux tels que le verre (figure VI.4).

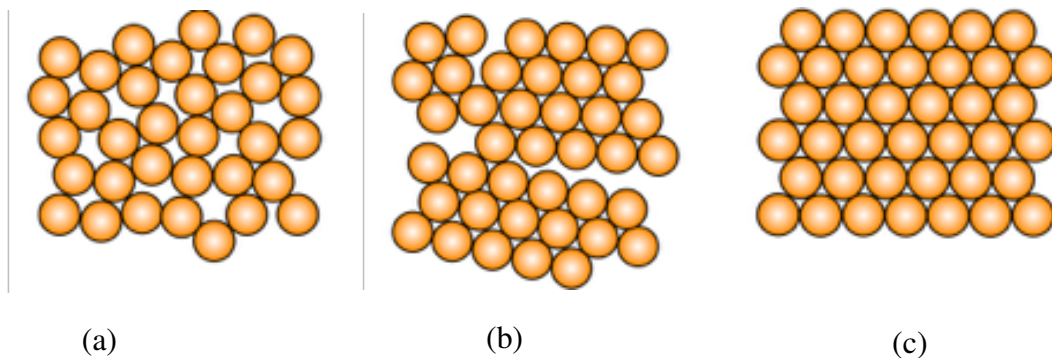


FIGURE VI.4 : Arrangement microscopique des atomes dans un matériau amorphe (a), des polycristaux (b) et un monocristal (c).

La maille d'un cristal est définie par 3 angles  $\alpha$ ,  $\beta$  et  $\gamma$  et 3 vecteurs **a**, **b** et **c** (figure VI.5). Les relations de symétrie qui existent entre les unités asymétriques au sein d'un cristal sont appelées groupes d'espaces. Au total, il existe 230 groupes d'espaces pouvant être classés selon sept systèmes réticulaires mais du fait de la chiralité des macromolécules biologiques, seuls 65 groupes d'espace sont possibles pour des cristaux de protéines. Du fait de cet arrangement ordonné, et surtout de leur périodicité, les cristaux possèdent des propriétés de diffusion de particules (neutrons, électrons, rayons X) menant à la diffraction qui vont pouvoir être exploitées par différentes approches expérimentales regroupées sous le terme de cristallographie. J'ai utilisé la cristallographie aux rayons X au cours de ma thèse (voir chapitre IV).

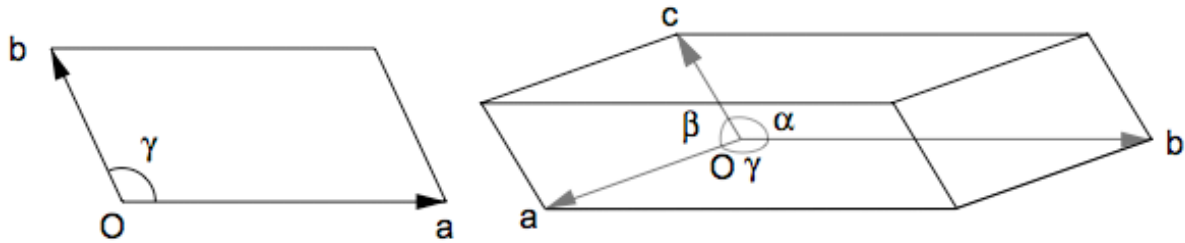


FIGURE VI.5 : Représentation des paramètres de maille

### B.1.b Principe de cristallisation

Bien que pouvant être un processus simple pour de petites molécules ou ions (exemple cristallisation du sel de cuisine), l'obtention de cristaux de molécules plus grosses et plus complexes telles que les macromolécules biologiques est souvent l'étape limitante lors d'un projet d'étude structurale. La cristallisation des protéines repose sur des techniques empiriques dont le principe général est d'explorer des conditions (principalement concentration de la protéine, pH, température, concentration en agent précipitant) afin de trouver celle(s) permettant l'agrégation ordonnée des molécules et non en précipitation amorphe. Ce processus peut être divisé en deux étapes : la nucléation et la croissance. Les deux étapes peuvent être optimisées par différentes méthodes détaillées ci-après. Le diagramme de phase est une représentation simplifiée d'un processus multidimensionnel et qui permet d'appréhender les principales situations qui vont être rencontrées lors de la recherche des conditions de cristallisation (figure VI.6).

La nucléation n'a lieu que si elle est énergiquement favorable, c'est-à-dire si l'énergie libérée par la formation du cristal est supérieure à l'énergie de la tension superficielle de l'interface liquide / solide. Afin de favoriser les interactions entre macromolécules, il est courant d'utiliser des agents chimiques précipitants, souvent des polymères d'éthylène glycol (PEG) ou des sels à haute concentration, qui vont abstraire les molécules d'eau de solvation des molécules. Les nuclei cristallins apparaissent lorsque la concentration en macromolécule et en agent précipitant est telle que celle-ci se trouve en condition de sursaturation (zone labile). Il n'est pas rare d'observer des cristaux se former à partir de protéine précipitée. Une fois les germes cristallins obtenus, la croissance du cristal peut s'initier. La croissance est optimale lorsque la macromolécule se trouve dans la zone métastable de la sursaturation. Lorsque la macromolécule repasse dans la zone de solubilité, la croissance s'arrête. Il est possible de découpler l'optimisation de ces deux phases par des techniques d'ensemencement (voir paragraphe VI.B.1.d).

Au cours de cette thèse, j'ai eu recours à différentes méthodes expérimentales dans le but d'explorer le digramme de phase afin de déterminer des conditions expérimentales permettant la cristallisation des protéines *lanYFP* et *mNeonGreen*. En particulier, la méthode de cristallisation par diffusion de vapeur en goutte assise qui a permis de cribler des centaines de conditions de cristallisation et la méthode de diffusion de vapeur en goutte suspendue utilisée pour optimiser les meilleurs conditions obtenues lors du premier criblage.

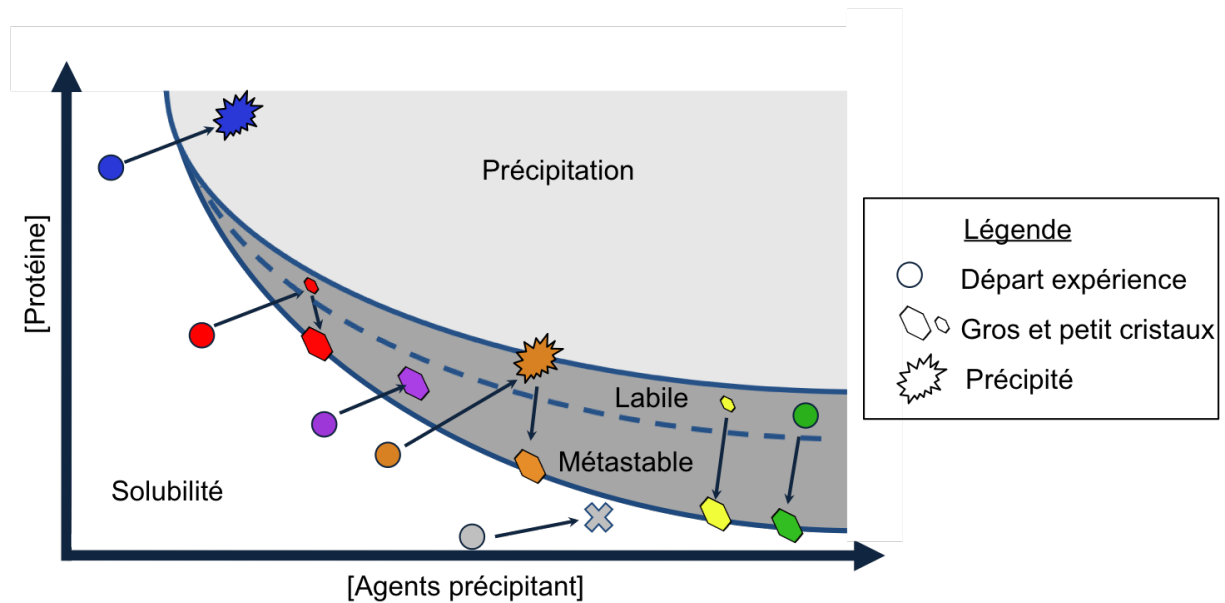


FIGURE VI.6 : Diagramme de phase et les conséquences empiriques de la cristallisation : précipitation (bleu), accroissement cristallin par diffusion en phase vapeur (rouge), dialyse (violet), cristallisation après précipitation partielle (orange), ensemencement (jaune), cristallisation en batch (vert) et pas de cristallisation (gris) [Haas and Drenth, 1999, Bergfors, 2003].

### B.1.c Méthode de diffusion de vapeur

Le principe de cette méthode repose sur l'échange de vapeur d'eau entre un réservoir contenant une combinaison particulière de réactifs chimiques (tampon, agent précipitant) et une goutte de la macromolécule d'intérêt purifiée et concentrée mélangée selon un certain ratio avec la condition du puit. Une fois le puit scellé, l'échange de vapeur permettra l'équilibration osmotique des concentrations d'eau entre la goutte et le puit. Selon la température du système la vitesse de l'échange pourra être plus ou moins rapide. Le criblage de conditions de cristallisation est un processus fastidieux et requérant le test de centaines conditions expérimentales simultanément. Afin de faciliter ce travail, j'ai eu recours à une plateforme de cristallisation à haut débit (HTX), disponible au Laboratoire Européen de Biologie Moléculaire (EMBL), permettant le criblage automatique de conditions. Des kits de cristallisation sont disponibles et contiennent

les conditions statistiquement les plus favorables à la cristallisation des protéines [Dimasi et al., 2007]. Le criblage à haut débit de ces conditions a été réalisé selon la méthode de diffusion en phase vapeur en goutte assise en plaques 96 puits ("Grenier Crystal Quick plates") et en utilisant des volumes compris entre 200 et 300 nL. Le robot de cristallisation de la plateforme HTX (PixSys 4200<sup>TM</sup> Genomic Solutions, Ann Arbor, USA) a été utilisé pour cribler plus de 1000 conditions de cristallisation pour mNeonGreen. L'observation et l'annotation des conditions criblées est également cruciale. Celle-ci est réalisée grâce à un second automate, véritable "ferme à cristaux" où la température est régulée et des photographies des gouttes de cristallisation sont prises à intervalles réguliers. Une fois les premiers "hits" obtenus, par exemple des microcristaux ou des conditions présumées favorables à l'obtention de cristaux, celles-ci sont optimisées manuellement par la technique de diffusion de vapeur en goutte suspendue. Cette optimisation manuelle a pour but de faire varier graduellement la condition "hit" en modifiant principalement deux paramètres qui peuvent par exemple être le pH et la concentration en agent précipitant. Concernant mNeonGreen et lanYFP, cette optimisation a été réalisée avec des plaques de cristallisation 24 ou 48 puits de type "Hampton limbro" ou "Hampton VDX48" (Hampton research, Aliso Viejo, USA). Le dispositif expérimental, schématisé en figure VI.7, est constitué du réservoir contenant la condition de cristallisation, une goutte ( $\sim 2 \mu\text{L}$ ) de la protéine purifiée et concentrée ( $30 \text{ mg.mL}^{-1}$ ) est déposée sur une lamelle de verre et mélangée à la condition selon un ratio prédéterminé et le tout est scellé en utilisant de la graisse silicone.

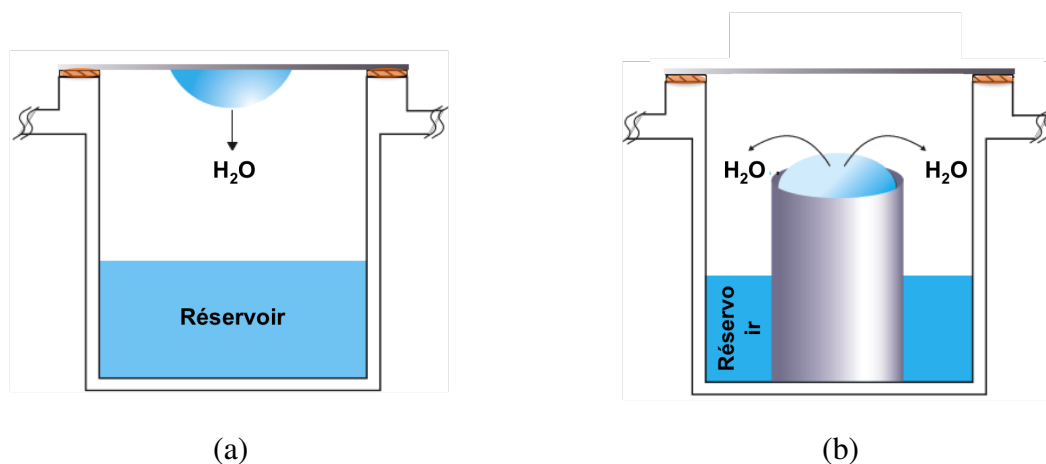


FIGURE VI.7 : Schéma expérimentaux de diffusion en phase vapeur par la technique de goutte suspendue (a) et goutte assise (b). La graisse silicone scellant les dispositifs apparaît en orange.

### **B.1.d Optimisation des cristaux par ensemencement homogène**

L'étape limitante dans le processus de cristallisation des protéines est la nucléation. Si trop de nuclei sont obtenus, la croissance sera limitée et nous parlerons alors de microcristaux. S'il n'y a qu'un seul nucleus dans une goutte de cristallisation, les molécules vont alors pouvoir venir s'agréger à celui-ci qui deviendra plus gros, certains pouvant atteindre parfois plusieurs millimètres. Le contrôle de cette étape est donc crucial pour obtenir des monocristaux de taille importante pouvant être utilisés pour une expérience de diffraction des rayons X.

L'ensemencement homogène à l'aide d'une moustache de chat ou un crin de cheval consiste à frotter un cristal de la protéine d'intérêt. Des nuclei sont arrachés grâce à la rugosité du poil, et déposés par trempage dans les gouttes de cristallisation. En revanche, cette technique ne permet pas de contrôler le nombre de nuclei déposés dans la goutte de cristallisation. Une autre technique consiste à écraser un ou des cristaux à l'aide de billes de verre et d'un vortex ou en utilisant un micropotter. On peut ainsi contrôler le nombre de nuclei utilisés pour l'ensemencement en diluant les germes dans la condition de cristallisation. Les cristaux ainsi dilués sont alors directement mélangés dans les gouttes de cristallisation [Bergfors, 2003]. Ainsi, en combinant la digestion ménagée, l'ensemencement et l'optimisation manuelle des conditions de cristallisation, il m'a été possible d'obtenir des cristaux de mes protéines à différents pH.

## **B.2 Diffraction aux rayons X**

Cette méthode expérimentale repose sur des principes fondamentaux qui sont décrits dans cette partie et sont extraits de la littérature [Drenth, 1999, McPherson, 2011] et des différents cours suivis.

### **B.2.a La loi de Bragg**

Etant donné la nature périodique, discrète et ordonnée de l'arrangement des molécules (voir paragraphe VI.B), les cristaux présentent des propriétés de diffusion particulière appelée diffraction des rayons X. En effet, la longueur d'onde des rayons X ( $\text{\AA}$ ) étant de même ordre de grandeur que la distance entre les plans réticulaires (figure VI.8), ils vont être diffusés par les électrons des atomes présents au sein de la maille. Dans certaines directions, les ondes réfléchies par différents plans réticulaires sont en phase et les interactions onde/matière sont constructives, on parle alors de pics de Bragg (nœud). La loi de Bragg qui régit les conditions de diffraction

(équation VI.1), permet de décrire la relation entre la longueur d'onde expérimentale utilisée ( $\lambda$ ), la distance inter-réticulaire ( $d$ ) et l'angle de diffraction ( $\theta$ ).

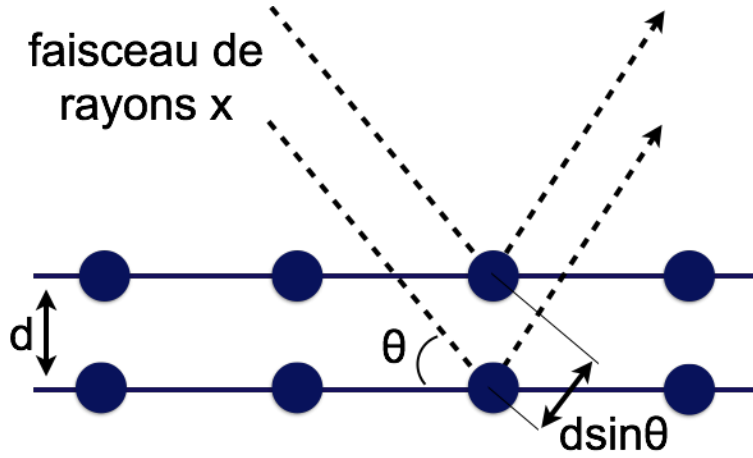


FIGURE VI.8 : Illustration de la loi de Bragg avec l'interaction du faisceau de rayons X et les plans réticulaires d'un cristal.

$$2 \cdot d_{hkl} \cdot \sin(\theta) = n\lambda \quad (\text{VI.1})$$

### B.2.b L'espace direct et l'espace réciproque

L'espace direct peut être défini par des vecteurs entre les différents atomes du cristal, à ce réseau est associé un réseau réciproque par transformée de Fourier (équation VI.2). Ainsi, chaque point de l'espace réel  $xyz$  s'exprime dans l'espace réciproque  $hkl$  comme une somme de fonctions de la distribution de la densité électronique grâce aux facteurs de structure  $F_{hkl}$  de module  $|F_{hkl}|$  et de phase  $\phi_{hkl}$  :

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \cdot e^{-2\pi \cdot i(hx+ky+lz)+i \cdot \phi_{hkl}}, \quad \text{où } F_{hkl} = |F_{hkl}| \cdot e^{i\phi_{hkl}} \quad (\text{VI.2})$$

Si  $V$  est le volume de la maille élémentaire, les facteurs de structures  $F_{hkl}$  sont obtenus en appliquant une transformée de Fourier inverse sur l'ensemble des facteurs de diffusion atomique  $f_j$  des atomes  $j = 1 \dots N$ .

$$F_{hkl} = \sum_{j=1}^N f_j(hkl) \cdot e^{-2\pi \cdot i(hx_j+ky_j+lz_j)+i \cdot \phi_{hkl}} \rightarrow \int_V \rho(hkl) \cdot e^{-2\pi \cdot i(hx+ky+lz)+i \cdot \phi_{hkl}} dV \quad (\text{VI.3})$$



A chaque réflexion est associé un facteur de structure possédant une amplitude et une phase. Si l'on est capable de mesurer l'intensité et la phase de tous les pics de Bragg, cette information permet de déterminer la densité électronique au sein du cristal.

### B.2.c Représentation de la sphère d'Ewald

Les expériences de diffraction classiquement réalisées au synchrotron utilisent des faisceaux de rayons X monochromatiques. C'est pourquoi, afin de mesurer l'intensité de tous les pics de Bragg, il est donc nécessaire de faire varier l'angle incident en faisant tourner le cristal positionné dans l'axe du faisceau de rayons X incident. Une représentation d'une expérience de diffraction est la sphère d'Ewald. Dans cette représentation (figure VI.9), la sphère a et le rayon de cette sphère est inversement proportionnel à la longueur d'onde expérimentale ( $1/\lambda$ ). Ainsi, lorsque le réseau réciproque intersecte la sphère d'Ewald, On définit également une seconde sphère ayant pour le centre de l'espace réciproque et dont l'intersection avec permet de définir la limite de diffraction du cristal. Son rayon est proportionnel à la limite de diffraction ( $1/d$ ).

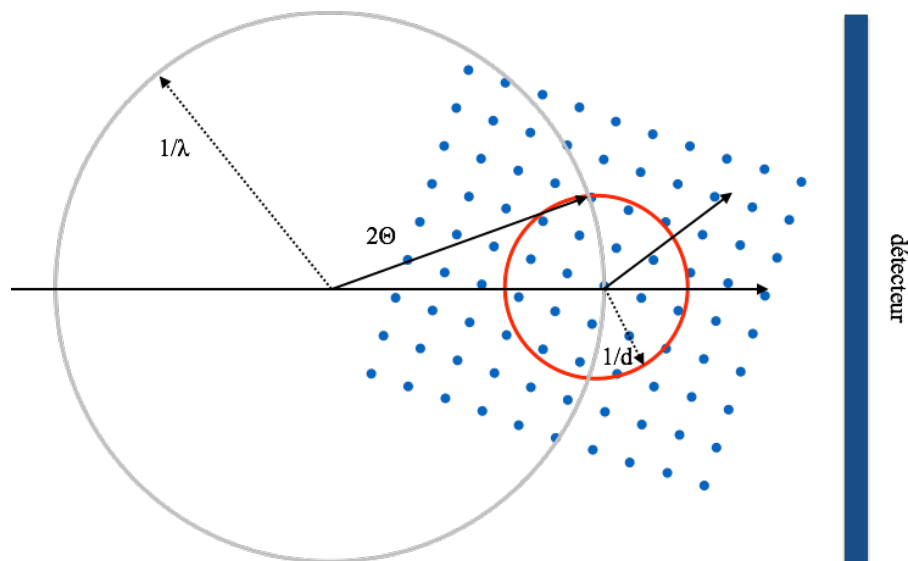


FIGURE VI.9 : Représentation de la construction d'Ewald. A chaque intersection entre la sphère d'Ewald de rayon  $1/\lambda$  (cercle gris) et un point du réseau réciproque contenu dans la sphère de résolution de rayon  $1/d$  (cercle rouge), une interférence constructive est visible par diffraction

### B.2.d Collecte des données de diffraction aux rayons X

Afin de mesurer l'intensité des pics de diffraction, on utilise un détecteur de rayons X. Les récents développements dans ce domaine ont abouti à l'obtention de détecteurs à pixels permettant de compter le nombre de photons atteignant celui-ci et ont une fréquence de rafraîchissement

élevé. Alors qu'il fallait jusqu'à plusieurs minutes pour obtenir un cliché de diffraction il y a tout juste une décennie, les images de diffraction ne prennent désormais que quelques dizaines de millisecondes. Les données de diffraction ont été collectées par la méthode dite du *fine slicing*, qui permet d'échantillonner les pics de diffraction par pas de 0,1 degrés.

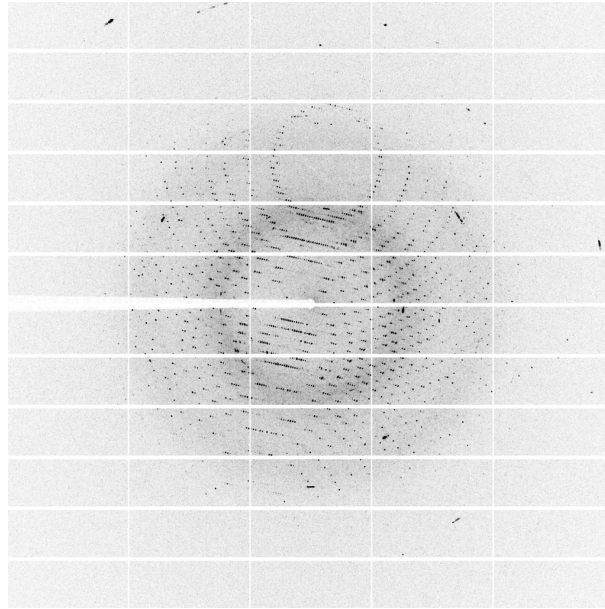


FIGURE VI.10 : cliché expérimental de diffraction par rayons X obtenu pour un cristal de *mNeonGreen*. L'ombre sur le cliché correspond à la position du capteur de faisceau.

### B.2.e Traitement des données de diffraction aux rayons X

Le traitement des données de diffraction est réalisé par des étapes séquentielles avec l'indexation, l'intégration, la mise à l'échelle et le calcul des cartes de densité électronique. L'indexation des tâches de diffraction permet de lier chaque tâche de diffraction à un point de l'espace réciproque et de déterminer la taille de la maille. L'intégration mesure l'intensité de chaque tâche de diffraction en effectuant la somme des intensités de chaque pixels constituant la tâche de diffraction. L'opération de mise à l'échelle combine les indices et les intensités de chaque tâche de diffraction en prenant en compte l'intensité des réflexions équivalentes. Le paramètre de  $R_{sym}$ , indicateur de la qualité du cristal, est obtenu à cette étape en calculant l'écart-type de la valeur moyenne des intensités mesurées pour les réflexions équivalentes. A partir de ces trois opérations, les facteurs de structures  $F_{hkl}$  de chaque réflexions sont calculés en utilisant la proportionnalité entre la racine carré de l'intensité des tâches de diffraction  $I_{hkl}$  et la valeur absolue des facteurs de structures  $F_{hkl}$ .

## **B.2.f Calcul de la carte de densité électronique et affinement de structure**

### **B.2.f i Le problème de la phase**

Les détecteurs de rayons X utilisés ne permettent que la mesure de l'intensité des réflexions. Or, d'après la théorie de la diffraction, il faut l'amplitude et la phase pour obtenir les facteurs de structure. La phase peut être obtenue soit par des méthodes expérimentales utilisant des atomes lourds, soit, et c'est la technique que j'ai utilisée, par remplacement moléculaire à partir d'un modèle préexistant [Rossmann, 1990].

### **B.2.f ii Le remplacement moléculaire**

Le principe de base de cette méthode est la réutilisation des phases obtenues pour une protéine précédemment résolue et qui possède une similarité de séquence ou de structure supérieure à 60% [Rossmann, 1990]. Cette technique est devenue incontournable en cristallographie puisque environ 70% des structures déposées sur la PDB ont été résolues avec cette méthode. On considère que les faibles différences de séquence n'induisent que des différences subtiles au niveau de la structure tridimensionnelle des protéines [Chothia and Lesk, 1986]. En dépit d'une divergence pouvant être importante au sein des protéines fluorescentes, leur structure tertiaire est globalement très bien conservée, facilitant ainsi l'obtention des phases par cette méthode.

### **B.2.f iii Affinement de structure et facteurs R**

L'affinement d'une structure cristallographique est un procédé itératif dont le but est de proposer un modèle structural moléculaire expliquant le plus fidèlement possible les données de diffraction. En d'autres termes, l'affinement doit permettre la convergence des facteurs de structure calculés à partir du modèle de travail ( $F_{calc}$  ou  $F_c$ ) et ceux expérimentaux provenant des données de diffraction aux rayons X ( $F_{obs}$  ou  $F_o$ ).

Des cartes de densité électronique sont calculées par transformée de Fourier en utilisant les amplitudes de facteurs de structures observés et calculés et des phases des facteurs de structure calculés générant des cartes dite  $2F_{obs}-F_{calc}$  autour de l'intégralité du modèle en construction (acides aminés, molécules d'eau, ligand). Néanmoins, les cartes  $2F_{obs}-F_{calc}$  sont biaisées par les phases du modèle référent. Afin de surmonter le biais introduit par le remplacement moléculaire, un second type de carte est calculé, appelé carte de Fourier différence  $F_{obs}-F_{calc}$  permettant de visualiser les différences entre le modèle en construction et les données expérimentales de diffraction. Les densités négatives (rouges) signalent des atomes positionnés de façon incorrecte

et les densités positives (vertes) invitent à la modélisation de nouveaux atomes.

La séquence itérative de l'affinement est composée par la construction dans l'espace réel d'un modèle avec COOT [Emsley and Cowtan, 2004] suivi d'un affinement automatique dans l'espace réciproque afin de diminuer les différences entre les facteurs de structure calculés, avec le modèle, et observé avec REFMAC5 [Murshudov et al., 1997]. Une fonction de minimisation  $L(p)$  peut être décrite par la formule :

$$L(p) = w \cdot L_{cris.}(p) + L_{geom.}(p) \quad (VI.4)$$

Où  $L_{cris.}(p)$  décrit la partie expérimentale issue de la diffraction et décrivant l'accord avec le modèle,  $L_{geom.}(p)$  prend en compte la chimie des acides aminés des protéines, comme les longueurs de liaisons covalentes, les angles de liaisons, les angles de torsion. Une pondération  $w$  est appliquée aux facteurs cristallographiques et géométriques en fonction de la résolution des données de diffraction aux rayons X. Le calcul d'un facteur de corrélation appelé facteur  $R_{cris.}(\%)$  permet de suivre l'évolution de l'affinement et de la qualité du modèle en construction.

$$R_{cris.} = \frac{\sum_{hkl} |F_{obs} - F_{calc}|}{\sum_{hkl} |F_{obs}|} \quad (VI.5)$$

Plus le facteur  $R_{cris.}(\%)$  est bas, plus l'accord entre les facteurs de structure observés et calculés est fort et par conséquent la confiance accordée dans le modèle proposé sera élevée. Afin de d'éviter des problèmes de sur-affinement, un second facteur de corrélation dit  $R_{free}$  a été introduit pour réaliser une validation avec un groupe de réflexions exclues lors de la procédure d'affinement (5 ou 10 % des réflexions totales) [Briinger, 1992].

$$R_{free} = \frac{\sum_{hkl \in T} |F_{obs} - F_{calc}|}{\sum_{hkl \in T} |F_{obs}|} \quad (VI.6)$$

La diminution conjointe des facteurs  $R_{cris.}$  et  $R_{free}$  est un indicateur de la justesse de l'affinement. La valeur de  $R_{free}$  est toujours supérieure à celle de  $R_{cris.}$  avec une différence typique de 3 - 5%. La procédure de validation inclue la confrontation des paramètres finaux du modèle avec les valeurs théoriques attendues avec des programmes comme *PROCHECK* [Laskowski et al., 1996] ou *MolProbity* [Read et al., 2011].

### **B.2.g Logiciels de cristallographie des protéines**

Pour l'ensemble des structures résolues et présentées, les clichés de diffraction ont été intégrés avec *XDS* et mis à l'échelle avec *XSCALE* [Kabsch, 2010]. Les structures cristallines ont été résolues par remplacement moléculaire en utilisant *Phaser MR* [McCoy et al., 2007, McCoy, 2007]. L'affinement des structures cristallographiques a été réalisé avec *REFMAC5* [Murshudov et al., 1997] et la reconstruction des modèles a été conduite grâce à *COOT* [Emsley and Cowtan, 2004]. Le calcul des doses de radiation a été réalisé avec le logiciel *RADDOSE-3D* [Zeldin et al., 2013]. L'ensemble des images de structures cristallographiques a été réalisé avec le logiciel de représentation *PyMOL* [DeLano, 2002].

## C Spectroscopie optique

Cette section est dédiée aux méthodes et instruments de spectroscopies optiques employés dont les principes fondamentaux, qui sont exposés dans cette section, sont extraits des ouvrages de Bernard Valeur et Joseph Lakowicz [[Valeur and Berberan-Santos, 2012](#), [Lakowicz, 2013](#)].

### C.1 Généralités

L'état fondamental  $S_0$  décrit l'état des molécules à leur niveau d'énergie la plus basse. Le passage au premier état excité  $S_1$  ou supérieur  $S_n$  s'effectue après absorption d'énergie, par exemple sous forme de photon lumineux. La vitesse du phénomène d'absorption est de l'ordre de la femtoseconde. La fluorescence est un phénomène physique qui peut se produire lors de la relaxation de la molécule à l'état excité  $S_1$  vers l'état fondamental  $S_0$  avec l'émission d'un photon lumineux. Ce phénomène radiatif s'effectue sur une échelle de temps de la nanoseconde ( $10^{-10}$  et  $10^{-7}$  s). D'autres voies de désexcitation des molécules rentrent en compétition avec l'émission de fluorescence. Les processus non radiatifs en compétition sont la conversion interne (de  $10^{-11}$  s à  $10^{-9}$  s) et le croisement inter-système (de  $10^{-10}$  s à  $10^{-8}$  s). La phosphorescence est un processus radiatif de désexcitation, postérieur au croisement inter-système, s'effectuant sur des échelles de temps entre une fraction de microseconde et la seconde ( $10^{-8}$  et  $10^0$  s).

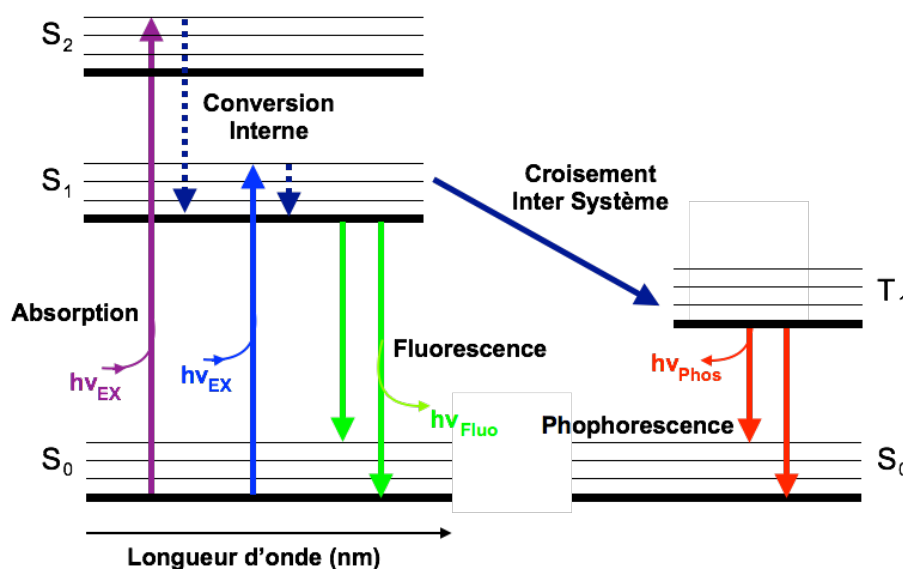


FIGURE VI.11 : Diagramme de Perrin-Jablonski représentant les processus de désexcitation moléculaire radiatifs et non radiatifs [[Perrin, 1926](#), [Jabłoński, 1935](#), [Valeur and Berberan-Santos, 2012](#)]

## C.2 Spectroscopie d'émission de fluorescence

Les spectres d'émission de fluorescence ont été enregistrés avec un spectrofluorimètre FP-8500 series (Jasco, Hachioji, Japon) à température ambiante. La largeur de la bande passante pour l'excitation et l'émission a été fixée à 1 nm. Les spectres d'émission de fluorescence ont été mesurés avec des concentrations  $C_{max} < 0,05$  pour 1 cm de trajet optique afin d'éviter les déformations des spectres liés à l'effet de filtre interne [Valeur and Berberan-Santos, 2012, Lakowicz, 2013]. Le rendement quantique d'émission de fluorescence ( $\Phi_{fl}$ ) rend compte de l'efficacité du phénomène d'émission de fluorescence en représentant la probabilité d'émission d'un photon de fluorescence par le rapport entre le nombre de photons émis ( $N_{em.}$ ) et ceux absorbés ( $N_{abs.}$ ). Le rendement quantique d'émission de fluorescence peut s'exprimer en fonction des constantes de vitesse des processus de désexcitation moléculaire radiatifs ( $k_r$ ) et non radiatifs ( $k_{nr}$ ) à l'état excité  $S_1$ .

$$\Phi_{fl} = \frac{N_{em.}}{N_{abs.}} = \frac{k_r}{k_r + k_{nr}} \quad (\text{VI.7})$$

## C.3 Spectroscopie Raman

La spectroscopie Raman permet d'identifier des modes de vibrations moléculaires qui sert à déterminer la composition d'échantillon [Raman and Krishnan, 1928]. Cette technique utilise une source lumineuse monochromatique (par exemple un laser) sur un échantillon et détecte la lumière diffusée. La majorité de la lumière diffusée est de même fréquence que la lumière d'excitation ( $\lambda_{ex.}$ ), ce phénomène est connu sous le nom de diffusion élastique ou diffusion de Rayleigh. Une très faible partie de la lumière diffusée correspondant à  $10^{-6}$  de l'intensité lumineuse totale est décalée en énergie par rapport à la fréquence du laser ( $\lambda_{dif.}$ ).

$$\Delta\omega = \left( \frac{1}{\lambda_{ex.}} - \frac{1}{\lambda_{dif.}} \right) \cdot 10^7 \quad (\text{VI.8})$$

Ce phénomène est dû aux interactions qui s'établissent entre l'onde électromagnétique incidente et les états vibrationnels des molécules de l'échantillon. Un spectre Raman est la représentation graphique de l'intensité de lumière en fonction du décalage en énergie (équation VI.8). La position de chaque bande de fréquence correspond aux niveaux d'énergie vibrationnelle de différents groupements chimiques, de la même manière que dans les spectres d'absorption infrarouge (IR). Ces positions sont indépendantes de la longueur d'onde incidente. La spec-

troscopie Raman possède une série d'avantages non négligeable par rapport à la spectroscopie IR : le signal de l'eau est faible, la possibilité de mesurer des bandes correspondant à des modes symétriques qui apparaissent moins bien en spectroscopie IR telles que les liaisons di-sulfures, doubles liaisons carbone-carbone ou les liaisons simples carbone-souffre. A contrario, la spectroscopie Raman voit mal les modes asymétriques en tous cas pas ceux des molécules centro-symétriques.

La spectroscopie Raman a été utilisée pour étudier les phénomènes de photo-blanchiment et de protonation du chromophore des protéines fluorescentes [Violot et al., 2009, Adam et al., 2009, McGeehan et al., 2011]. L'attribution des fréquences aux groupements chimiques d'après les spectres Raman a été faite en se basant sur la littérature (voir table IV.4)

## C.4 Protocoles expérimentaux de détermination de constantes spectroscopiques

### C.4.a Détermination du coefficient d'extinction molaire

Le coefficient d'extinction molaire ( $\epsilon$  du chromophore de protéine fluorescente) est mesuré selon le protocole de dénaturation basique avec une solution d'hydroxyde de sodium concentrée à 250 mM qui expose le chromophore au solvant en dénaturant la protéine [Ward et al., 1981, Ward and Bokman, 1982, Ward, 2005]. Le coefficient d'absorption molaire du chromophore de type XYG dénaturé étant égal à  $44000 \text{ M}^{-1} \cdot \text{cm}^{-1}$  à 447 nm pour un pH supérieur à 10 [Ward et al., 1981], la mesure de l'absorption du chromophore à cette longueur d'onde permet de remonter à  $\epsilon$  de la PF.

### C.4.b Détermination du rendement quantique par comparaison

Les rendements quantiques d'émission de fluorescence ont été déterminé en prenant comme référence mNeonGreen avec  $\Phi^{mNG} = 0,80$  [Shaner et al., 2013]. Les échantillons doivent avoir des densités optiques (DO) proches et inférieur à 0,05 pour un trajet optique de 1 cm à la longueur d'onde d'excitation ( $\lambda = 480 \text{ nm}$ ). Le rendement quantique est calculé avec l'équation VI.9. Les intensités de fluorescence  $I_{fl}$  ont été calculées en intégrant l'aire sous les spectres d'émission de fluorescence.

$$\Phi = \frac{I_{fl}}{I_{fl}^{ref}} \cdot \frac{1 - 10^{-A_{ref}}}{1 - 10^{-A}} \cdot \Phi^{ref} \quad , \text{ où } \Phi^{ref} = \Phi^{mNG} \quad (\text{VI.9})$$



### C.4.c Détermination de la brillance

La brillance est le produit entre le coefficient d'extinction molaire ( $M^{-1}.cm^{-1}$ ) et le rendement quantique (ua). La valeur de la brillance ( $M^{-1}.cm^{-1}$ ) caractérise les performances photophysiques des protéines fluorescentes. L'utilisation de la brillance permet la comparaison entre protéines fluorescentes de gammes spectrales différentes.

## C.5 Spectroscopie optique *in crystallo*

Des instruments pour l'application de spectroscopies optiques sur cristal de protéines ont été développés dans le cadre d'une collaboration entre l'IBS et l'ESRF. Ils sont rassemblés au sein de la plateforme ID29S-Cryobench et permettent l'enregistrement de spectres d'absorbance, d'émission de fluorescence et Raman de protéines en général colorées, à la fois sur (*online*) et hors (*offline*) des lignes de lumière [von Stetten et al., 2015].

La spectroscopie optique est réalisée à concentration constante (celle déterminée par l'empilement cristallin) et très forte (effet de filtre interne très important). Les mesures de fluorescence sont réalisées par réflexion en s'adaptant à la géométrie de l'échantillon qui varie suivant le cristal et le support de l'échantillon généralement sur un boucle nylon.

La température d'acquisition est contrôlée par une canne cryogénique à écoulement ouvert à températures réglable (100 à 270 K) (Cryostream Series 700, Oxford Cryosystems, Oxford, England). Des spectres à température ambiante et sous atmosphère contrôlée peuvent être acquis grâce à un déshumidificateur HC1 (ESRF/EMBL) [Sanchez-Weatherby et al., 2009]. Trois types de spectroscopie optiques sont disponibles au *cryobench* :

- **Absorption UV-visible** qui est mesurée avec deux objectifs opposés à  $180^\circ$  (mode transmission, figure VI.12.a), une lampe halogène au deutérium (Mikropack DH2000-BAL, Ocean Optics, Dunedin, USA) et un spectrophotomètre à réseau fixe (model HR2000+ ou QE65Pro, Ocean Optics, Dunedin, USA).
- **Emission de fluorescence** qui est enregistrée avec deux objectifs orientés à  $90^\circ$  (mode réflexion, figure VI.12.b). L'enregistrement des données est synchronisé électroniquement afin de minimiser le temps d'exposition au laser (laser disponible entre 266 nm et 671 nm).
- **Raman** qui est réalisée avec un spectrophotomètre Renishaw InVia [Carpentier et al., 2007] et les signaux sont traités avec le logiciel Wire3 (mode rétro-diffusion, figure

VI.12.c).

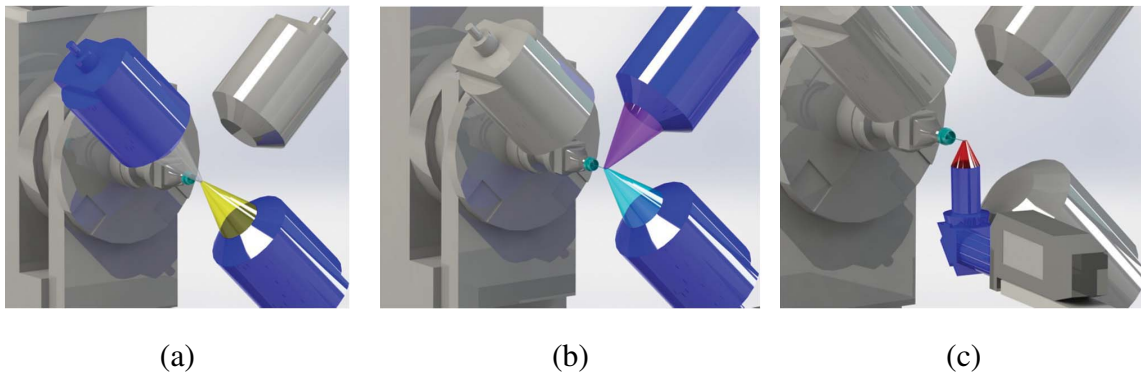


FIGURE VI.12 : Montage expérimental cryobench (ESRF-ID29S) et les configurations des objectifs utilisés pour (a) l'absorption UV-visible avec le mode transmission (géométrie  $180^\circ$ ), (b) l'émission de fluorescence avec le mode réflexion (géométrie  $90^\circ$ ) et (c) la spectroscopie Raman avec le mode rétro-diffusion (géométrie  $0^\circ$ ) [von Stetten et al., 2015].

Un montage équivalent, nommé microspectrophotomètre, est installable directement sur les lignes de lumière de l'ESRF, BM30A et ID30A-3. Ce montage est relié par une fibre de 50 m au diffractomètre de la ligne de ID29S afin de collecter des données de spectroscopie "online" Raman (figure VI.13).

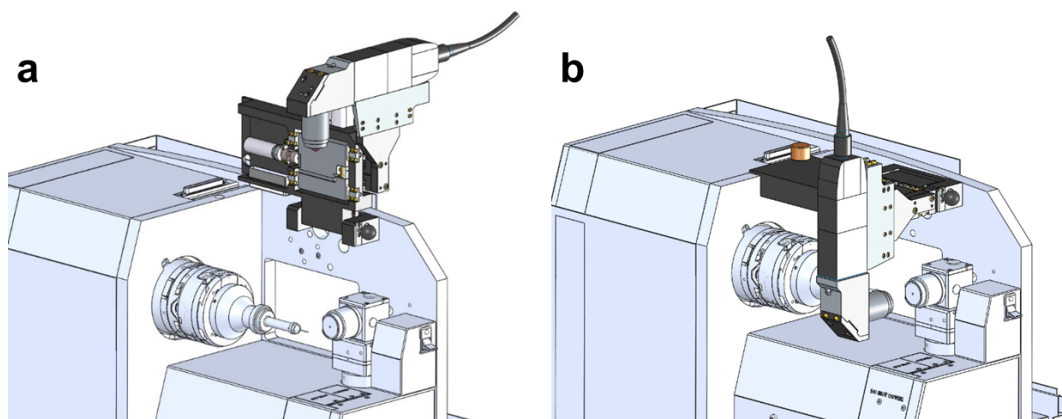


FIGURE VI.13 : Montage de la sonde Raman sur le diffractomètre de la ligne de lumière ID29. (a) Position de la sonde hors du faisceau de rayons X pour collecte de données de diffraction. (b) Position de la sonde en face de l'échantillon pour collecte de données de spectroscopie Raman.

## D Simulations par dynamique moléculaire

Les simulations de dynamique moléculaire tout atome visent à modéliser les propriétés structurales et dynamiques à l'échelle atomique de l'ensemble des atomes d'une macromolécule dans des conditions expérimentales réalistes de température et de pression. Cela implique de représenter au mieux le système étudié et en particulier les différentes interactions intra- et inter-moléculaires subies par chacun de ces atomes. Cette partie vise à présenter les principes généraux des simulations de dynamique moléculaire classique qui sont effectuées avec la suite de programmes *AMBER 14*. Les paramètres d'utilisation des algorithmes tels que *tLEAP* ou *SHAKE* ainsi que les champs de forces spécifiques au chromophore seront introduits dans cette partie

### D.1 Principe

#### D.1.a Equations du mouvement de Newton

La modélisation du mouvement des atomes est accessible par l'application des lois du mouvement de Newton sur un ensemble de points, représentant les  $N$  atomes de la macromolécule. Chaque point ou atome est caractérisé par une position  $(x,y,z)$ , une charge, une masse et une accélération. Cette description peut se résumer à la deuxième loi de Newton :

$$\vec{F}_i = m_i \cdot \vec{a}_i \quad (\text{VI.10})$$

où la somme des forces appliquées sur un atome  $i$  de masse  $m_i$  est représentée par  $F_i$ . L'accélération subie par cet atome  $i$  est décrit par  $a_i$ . La force agissant sur l'ensemble des atomes est définie par le gradient de l'énergie potentielle  $U(r_1, \dots, r_N)$  qui représente la surface d'énergie potentielle de  $N$  atomes qui interagissent en fonction de leur position  $r_i$ .

$$\vec{F}_i = -\vec{\nabla} \cdot U(r_i) \quad (\text{VI.11})$$

La combinaison des deux équations VI.10 et VI.11 permet d'établir une relation entre la dérivée de l'énergie potentielle et le déplacement des coordonnées  $r_i$  des atomes en fonction du temps.

$$-\frac{\partial U(r_i)}{\partial \vec{r}} = m_i \cdot \frac{\partial^2 \vec{r}_i(t)}{\partial t^2} \quad (\text{VI.12})$$

Le couplage de tous les atomes qui composent le système rend difficile la résolution analytique de ces équations. Par conséquent, les trajectoires de dynamique moléculaire sont obtenues itérativement par intégration numérique des équations du mouvement de Newton. Plusieurs approches d'intégration existent. L'une d'entre elles consiste à utiliser un développement en série de Taylor appelé méthode des différences finies. Cette méthode, appliquée par la suite de programmes *AMBER* avec l'algorithme *Leapfrog Verlet* (mélange de *Leapfrog* et de *velocity Verlet* pour plus de précision), intègre les équations du mouvement sur un temps court  $t + \frac{1}{2}\partial t$  durant lequel les forces agissant sur chaque atome du système sont considérées comme constantes (équation VI.13). Les coordonnées  $r_i$  de chaque atome sont calculées pour un temps égal à  $t + \partial t$  (équation VI.14).

$$v\left(t + \frac{1}{2}\partial t\right) = v\left(t + \frac{1}{2}\partial t\right) + a(t)\partial t \quad (\text{VI.13})$$

$$r(t + \partial t) = r(t) + v\left(t + \frac{1}{2}\partial t\right)\partial t \quad (\text{VI.14})$$

Les coordonnées initiales de l'ensemble des atomes découlent de la structure cristallographique alors que leurs vitesses initiales sont fixées aléatoirement et vérifient la loi de distribution de vitesses de Maxwell-Boltzmann. Par ailleurs, le choix du temps d'intégration  $\partial t$  est un paramètre essentiel de la dynamique. L'ordre de grandeur du temps d'intégration  $\partial t$  doit être environ dix fois inférieur à la période de vibration interatomique la plus rapide. Dans le cas des systèmes biologiques, cette vibration est celle correspondant à l'élongation des liaisons X-H dont la période est de l'ordre de 10 fs ce qui correspondrait à un temps d'intégration de 1 fs au maximum. Pour réduire le temps calcul, les liaisons X-H sont figés en utilisant le programme *SHAKE* [Ryckaert et al., 1977, Miyamoto and Kollman, 1992]. Ceci permet d'augmenter le pas de calcul à 2 fs.

### D.1.b Conditions aux limites périodiques

Les calculs de dynamique moléculaires ont été réalisés dans une boîte de simulation contrôlée par des conditions aux limites périodiques. Ces conditions ont pour objectif de reproduire les conditions expérimentales. Par conséquent, les contraintes spatiales définissent la forme, la

taille ainsi qu'un maillage infini. Dans nos dynamiques, la boîte de simulation est de forme parallélépipédique et un modèle de solvant explicite décrit les molécules d'eau (VI.16). La taille de la boîte de simulation est en accord avec la distance limite des interactions à longue distance (*cutoff*). La périodicité infinie du lattice est imposée dans les simulations isotherme-isobare (NPT) permettant l'utilisation de la méthode de Particle Mesh Ewald (PME) pour le calcul des interactions électrostatiques tout en évitant les problèmes de taille finie.

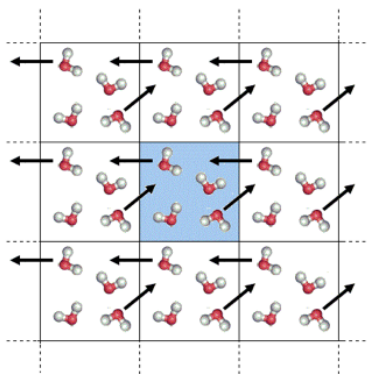


FIGURE VI.14 : Vue bidimensionnelle d'une cellule de simulation répliquée dans les trois dimensions de l'espace. Lors de l'utilisation des conditions aux limites périodiques, si une molécule quitte la boîte centrale ses images dans les cellules fantômes voisines se déplacent de manière similaire.

### D.1.c Contrôle de température et de pression

L'utilisation d'un thermostat permet le calcul de trajectoires qui se rapprochent au plus près des conditions expérimentales avec  $T = 300$  K et  $P = 1 \times 10^5$  pascal (1 bar). Les objectifs du contrôle de la température sont multiples comme l'étude de la dépendance en température du système, la recherche conformationnelle ou pour éviter d'éventuelles dérives de l'énergie suite à l'accumulation d'erreurs de calcul. La modulation de la température dans nos dynamiques est assurée par les équations dites de Langevin préalablement implantées dans la suite *AMBER* [Pastor et al., 1988, Loncharich et al., 1992]. La dynamique de Langevin permet le couplage de tous les atomes d'un système et d'un bain chaud au travers d'une force de frottement fluide du type  $f = -k \cdot v_i$  et une force complémentaire représentant le choc aléatoire des molécules de solvant. La seconde équation de Newton peut être remplacée par l'équation de Langevin (équation VI.15).

$$\vec{F}_i = m_i \cdot \vec{a}_i = -\vec{\nabla}V_i - k \cdot v_i + R(t) \quad (\text{VI.15})$$

où  $k$  représente le coefficient de frottement visqueux et  $R(t)$  est une force variable traduisant les

chocs aléatoires du solvant sur l'atome  $i$ . La gestion de la pression est réalisée par un barostat qui module le volume de la boîte de simulation par des petits ajustements. Cette méthode permet de faire converger la pression et la température du système vers des valeurs de référence équivalente à 1 bar et 298 K.

### D.1.d Champs de force

#### D.1.d i Définition

Dans la suite *AMBER*, le champ de force donnant l'expression analytique de l'énergie potentielle  $U_{tot}$  du système qui régit les interactions intra- et inter-moléculaires est défini selon l'expression suivante :

$$\begin{aligned}
 U_{tot} = & \sum_{\substack{\text{liaisons} \\ \text{covalentes}}} k_r (r - r_{eq})^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{eq})^2 + \\
 & \sum_{\substack{\text{angles} \\ \text{dièdres}}} k_\omega (1 + \cos[n\omega - \gamma]) + \sum_{\substack{\text{angles} \\ \text{dièdres} \\ \text{impropres}}} k_\phi (\phi - \phi_{eq})^2 + \\
 & \sum_{i < j}^{\text{atomes}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i < j}^{\text{atomes}} \epsilon_m \left[ \left( \frac{r_m}{r_{ij}} \right)^{12} - 2 \left( \frac{r_m}{r_{ij}} \right)^6 \right]
 \end{aligned} \tag{VI.16}$$

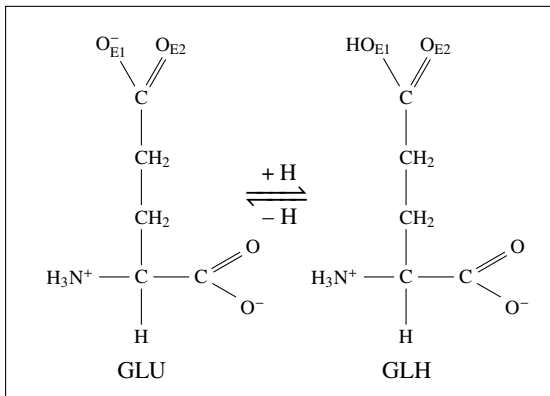
Les quatre premiers termes prennent en charge les interactions intra-moléculaires ou liées. Le premier terme représente l'énergie potentielle d'élongation d'une liaison. Ce terme est décrit par un oscillateur harmonique avec une constante de force  $k_r$  ( $\text{kcal.mol}^{-1}.\text{Å}^{-2}$ ) et une distance d'équilibre  $r_{eq}$  (Å) qui perd en précision avec l'éloignement des atomes. Le second terme, également décrit par un oscillateur harmonique, correspond à l'énergie de déformation des angles de valence. La constante de force est notée  $k_\theta$  ( $\text{kcal.mol}^{-1}.\text{radian}^{-2}$ ) et l'angle d'équilibre  $\theta_{eq}$  (radian). La déformation des angles dièdres permet de représenter l'énergie contenue dans la torsion d'une liaison covalente à cause de l'ordre de liaison et des liaisons environnantes qui dépendent des paires libres d'électrons. Elle est décrite par  $k_\omega$  qui est la valeur de mi hauteur de la barrière de torsion ( $\text{kcal.mol}^{-1}$ ) autour de l'axe de la torsion  $\omega$ ,  $n$  et  $\gamma$  représentent respectivement la multiplicité de la liaison et la phase (radian). Le dernier terme d'interactions liées correspond à la déformation des angles dièdres impropres décrit par un oscillateur harmonique avec une constante de force  $k_\phi$  ( $\text{kcal.mol}^{-1}.\text{radian}^{-2}$ ) et l'angle d'équilibre  $\phi_{eq}$  (radian). Le potentiel  $U_{tot}$  modélise également les interactions dites non liées avec la description des interactions

électrostatiques et de van der Waals. Les interactions électrostatiques sont décrites par la loi de Coulomb où  $q$  désigne la charge atomique,  $r$  la distance inter-atomique des noyaux et  $\epsilon_0$  la permittivité du vide. Le dernier terme décrit un potentiel de Lennard-Jones entre une paire d'atomes distant d'une longueur  $r$ . Seules les interactions inter-atomiques entre atomes distants de moins de 8 Å sont prises en compte.

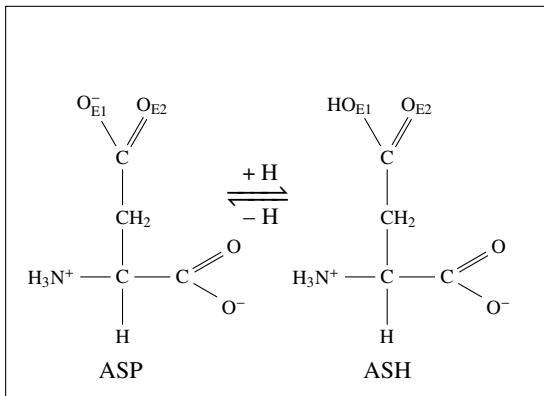
#### D.1.d ii Champ de force *ff99* de la suite de programmes *AMBER* : résidus standards

Lors de simulation par dynamique moléculaire, le champ de force utilisé pour décrire les acides aminés standards est *ff99* [Wang et al., 2000] de la suite de programmes *AMBER*. Le champ de force *ff99* fait appel à une librairie (`all_amino94.lib`) pour l'ensemble des acides aminés qui contient les valeurs des paramètres intervenant dans l'équation VI.16 pour chacun d'entre eux [Cornell et al., 1995]. Chaque acide aminé possède un patron de référence (qui définit sa composition chimique et sa topologie) correspondant à la forme neutre décrit par un code à trois lettres (e.g. ALA pour Alanine). Les acides aminés acido-basiques ont plusieurs patrons qui permettent de les décrire suivant leur état de protonation. En particulier, l'histidine possède trois patrons différents correspondant à trois états de protonation de l'imidazole. Une histidine portant un proton sur l'azote  $\epsilon$  ou  $\delta$  est noté respectivement HIE ou HID. Lorsque l'histidine est doublement protonée, le patron associé est nommé HIP (figure VI.15). L'acide glutamique, quant à lui, est décrit par deux patrons GLU et GLH correspondant à l'état déprotonné et protoné du groupe carboxyle de sa chaîne latérale. Les mêmes considérations sont à prendre en compte dans le cas de l'acide aspartique avec son patron déprotonné, ASP, et protoné ASH. En outre, dans le cas de GLH et de ASH, le proton est invariablement porté par l'atome OE1.

## Acide glutamique



## Acide aspartique



## Histidines

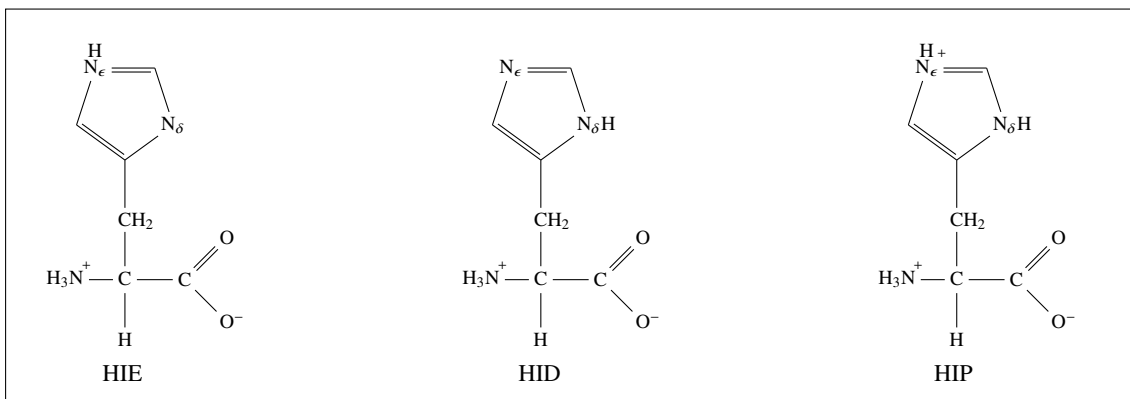


FIGURE VI.15 : Structures chimiques des différents états de protonation des acides aminés acide glutamique (GLU, GLH), acide aspartique (ASP, ASH) et histidine (HID, HIE, HIP). Le nom indiqué sous chaque structure correspond au code utilisé dans la suite AMBER

### D.1.d iii Champ de force spécifique au chromophore : un résidu non-standard

Le chromophore des protéines fluorescentes ne fait pas partie des acides aminés standards. Il n'est donc pas défini par le champ de force *ff99* de la suite *AMBER*. Nos travaux ont nécessité de faire appel à des champs de force permettant de décrire les chromophores de type GYG, SYG et TYG [Vallverdu et al., 2010, Jonasson et al., 2011]. Pour ce faire, nous avons utilisé les champs de force à l'état fondamental ( $S_0$ ) précédemment mis au point par l'équipe ThéoSim du LCP. La différence entre les champs de force des différents chromophores GYG, SYG et TYG se situe au niveau de la chaîne latérale du résidu 65. La topologie du chromophore est modifiée et les charges partielles sont réparties sur les atomes adjacents. Pour chaque champ de force, la somme des charges des atomes du chromophore est nulle.



Topologie du chromophore	Etat de protonation	
	Protoné	Déprotoné
GYG	$ff\_GYG\_prot$	$ff\_GYG\_deprot$
SYG	$ff\_SYG\_prot$	–
TYG	$ff\_TYG\_prot$	–

Tableau VI.1 : Nomenclature des champs de force décrivant les différents types de chromophore utilisés. Le champ de force utilisé dépend de l'état de protonation du groupe phénol du chromophore

#### D.1.d iv Le modèle de molécule d'eau

L'ajout d'une boîte d'eau autour de la protéine permet de réaliser des calculs de dynamique moléculaire en solution. Le modèle des molécules d'eau utilisé pour la solvation est de type TIP3P [Jorgensen et al., 1983]. Il décrit un modèle à trois sites d'interactions qui correspondent aux trois atomes de la molécule d'eau. Une charge partielle est assignée à chaque atome avec pour l'atome d'oxygène  $q(O) = -0,834 C$  et pour les atomes d'hydrogène  $q(H) = 0,417 C$ , seul l'atome d'oxygène est muni d'un potentiel de Lennard-Jones. Le type TIP3P est un modèle de molécule d'eau adapté aux simulations des systèmes biologiques [Stumpe et al., 2010].

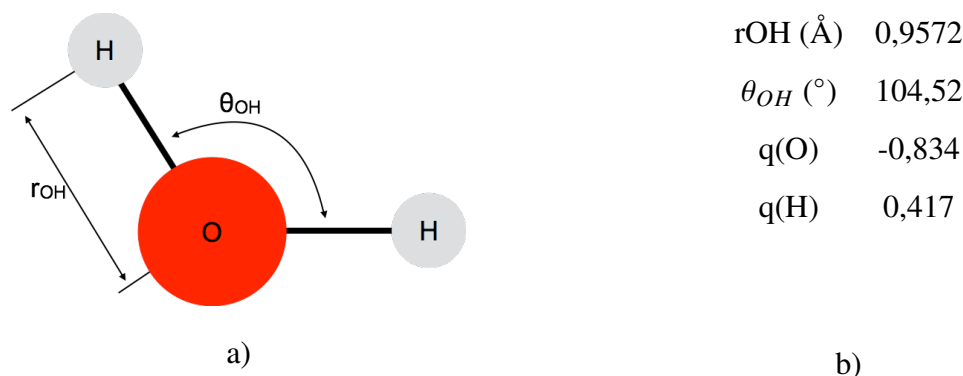


FIGURE VI.16 : modèle TIP3P à trois sites d'interaction pour les molécules d'eau. (a) Description géométrique, (b) paramètres physiques

## D.2 De la structure cristallographique à la dynamique moléculaire

### D.2.a Préparation du système

La simulation par dynamique moléculaire appliquée aux protéines requiert les coordonnées initiales (x,y,z) des différents atomes de la protéine. Ces informations sont accessibles grâce

aux techniques de diffractions aux rayons X (voir chapitre B). Ces structures contiennent des lacunes par exemple au niveau du squelette, des chaînes latérales ou des atomes d'hydrogène, dont le nombre dépend de la qualité des cartes de densité électronique et de leur modélisation par le cristallographe. A cela s'ajoute une modélisation de la structure non compatible avec la nomenclature de la suite *AMBER* (voir paragraphe VI.D.2.a i) ou encore la présence de différentes conformations locales dont il faut choisir la plus probable. Par conséquent, une première étape est indispensable pour obtenir des coordonnées initiales de tous les atomes nécessaire pour une bonne description du système. Cela implique de définir au départ l'ensemble des états de protonation des résidus de la protéine.

### **D.2.a i Etat de protonation des résidus acido-basiques de la protéine**

L'état de protonation des résidus portant une chaîne latérale constituée d'une fonction acido-basique est déterminant car il régit les interactions entre les différents résidus ce qui peut être crucial pour la fonction, l'activité et la stabilité de la protéine. Dans la recherche des réseaux de LH, il gouverne l'établissement des LH et leur orientation pour établir un réseau susceptible de sous-tendre une translocation de proton du donneur D vers l'accepteur A de proton. Cet état de protonation a en effet une incidence sur le caractère D, A ou encore D et A des atomes protonables. Au cours d'une dynamique moléculaire, la nature des atomes n'est pas modifiable, nous devons donc définir dès le départ l'ensemble des états de protonation les plus probables de tous les résidus protonables (Asp, Glu, His, Lys, Arg, Cys, Tyr) de la protéine. A titre d'exemple, dans AvGFP, il y a 99 résidus protonables. L'état de protonation de ces résidus dépend du pH auquel a été obtenue la structure cristallographique et du pKa de chacun de ces résidus. Dans la protéine, celui-ci peut fortement diverger de celui connu en solvant, dépendant de la nature électrostatique de l'environnement dans lequel le résidu se trouve, induite par la présence d'autres résidus protonables ou d'ions, notamment. Nous avons utilisé le serveur de calcul  $H^{++}$  [Gordon et al., 2005, Anandakrishnan et al., 2012] qui calcule les courbes de titration et détermine la valeur du  $pK_{1/2}$  de chaque résidu acido-basique de la structure.

Pour la grande majorité des résidus, nous avons sélectionné la forme basique des résidus protonables dès lors que le pH cristallographique était supérieur au  $pK_{1/2}$  estimé de ces résidus et la forme acide dans le cas contraire [Shinobu et al., 2010]. L'état de protonation de certains résidus, notamment proche du chromophore, ayant un comportement particulier a été établi avec attention. Les choix pour ces résidus seront discutés dans la partie résultat (voir paragraphe

III.C.2.a).

### D.2.a ii Préparation du système pour la dynamique moléculaire

Le programme *iLEAP* de la suite *AMBER* permet de compléter de façon automatisée des structures cristallographiques. Cela nécessite tout d'abord d'utiliser la librairie de patrons d'acides aminés pour ajouter les atomes manquant aux structures, tenant compte des états de protonation préalablement déterminés. Dans un second temps la protéine est solvatée grâce à l'ajout de molécules d'eau dans la boîte de simulation. Enfin, les simulations sont menées sur des systèmes globalement neutres. Cette charge globale nulle est assurée grâce à l'ajout d'un nombre nécessaire de contre-ions  $Na^+$  ou  $Cl^-$  qui se substituent alors à un nombre équivalent de molécules d'eau. A l'issue cette étape, le système est prêt pour la phase de minimisation et de mise à l'équilibre. Les paramètres des champs de force et la topologie du système sont compilés dans des fichiers d'entrée nécessaires au lancement des dynamiques.

### D.2.b Minimisation de l'énergie du système et mise à l'équilibre

Les coordonnées de départ des simulations par dynamique moléculaire sont celles issues de la structure cristallographie obtenue à température cryogénique (100 K). Le système doit progressivement être amené dans les conditions de référence afin d'atteindre une température de 300 K et une pression de 1 bar. Ce processus est réalisé au cours de deux grandes étapes avec les procédures de minimisation de l'énergie du système et sa mise à l'équilibre.

La phase de préparation du système consistant globalement à l'ajout d'atomes d'hydrogène, de molécules d'eau pour représenter le solvant et de contre-ions pour assurer la neutralité globale du système peut générer des conflits stériques s'accompagnant d'une libération d'énergie pouvant conduire jusqu'à des réarrangements structuraux non physiques. Afin de relaxer ces répulsions, un ensemble d'étapes de minimisation de l'énergie du système reposant sur l'application de contraintes fortes sur les coordonnées de départ des atomes et à leur retrait progressif sont menées. La minimisation de l'énergie est un processus divisé en trois étapes où la position des hydrogènes, des molécules d'eau et enfin de l'ensemble des atomes du système est successivement optimisée. Le niveau de contraintes est de 5000 kcal/mol pour l'optimisation des angles des LH et pour les molécules d'eau. Ce traitement permet d'améliorer les interactions entre la protéine et le solvant aqueux. La minimisation de l'énergie du système est réalisée avec une contrainte plus faible de l'ordre de 5 kcal/mol permettant une optimisation des coordonnées.

La mise à l'équilibre du système est un processus se déroulant en deux temps, le chauffage et la mise sous pression. Le chauffage du système est réalisé progressivement entre 100 K et 300 K en conservant un volume constant. Cette dynamique est menée avec l'aide d'un thermostat durant 200 ps avec un paramètre de fréquence de collision fixé à  $1 \text{ ps}^{-1}$ . Lors de cette étape, les contraintes de 5 kcal/mol sont maintenues sur la protéine seule. La mise sous pression est réalisée sous pression constante avec une dynamique isotherme-isobare (NPT). A ce stade de la mise à l'équilibre, les contraintes appliquée à la protéine sont relâchées petit à petit jusqu'à la libération complète du système. La diminution des contraintes sur la protéine sont effectués dans l'ordre suivant, 5 kcal/mol, 2 kcal/mol, 1 kcal/mol, 0,5 kcal/mol et 0,1 kcal/mol avec des durées de 200 ps pour la première étape et de 100 ps pour les suivantes. Cette phase d'équilibration dure 900 ps. Une série de 40 structures avec des coordonnées et des vitesses propres sont échantillonnées au cours de la dernière étape d'équilibration, la moins contrainte (0,1 kcal/mol). Cette série de structures sert de points de départ à la phase de production de la simulation de dynamique moléculaire à l'état fondamental ( $S_0$ ).

### D.2.c Phase de production

Le système est prêt à entrer en phase de production dès la fin de la phase d'équilibration. La phase de production correspond à une série de 5 simulations courtes de la protéine d'une durée de 8 ns chacune et permet d'explorer le paysage conformationel à l'état fondamental de la protéine. Le pas d'écriture choisi est de 2 ps soit l'écriture d'une série de 500 fichiers de coordonnées et vitesses par nanoseconde de simulation.

Le contrôle approfondi de l'agitation de la structure protéique et notamment du squelette (atomes N, C, O et  $C_\alpha$ ) des structures secondaires via le calcul de RMSd et RMSf est nécessaire pour vérifier que le système tend vers un équilibre.

$$RMSd(t) = \sqrt{\frac{1}{M_{totale}} \sum_{i=1}^{N_{atomes}} m_i (r_i(t) - r_i^{ref})^2} \quad (\text{VI.17})$$

Le RMSd décrit l'évolution en fonction du temps de la moyenne du déplacement carré moyen des atomes de la protéine. La configuration  $r_i(t)$  est extraite de la simulation pour être comparée à  $r_i^{ref}$ . L'équation du RMSd comprend les termes  $M_{totale}$ ,  $N_{atomes}$  et  $m_i$  qui sont la masse totale, le nombre des atomes ainsi que la masse de l'atome considéré  $i$ . Dans le cas du calcul du RMSd, la configuration de référence  $r_i^{ref}$  est la structure cristalline. Le RMSF est

quant à lui décrit comme l'évolution au cours d'un intervalle de temps  $t$  du déplacement carré moyen des atomes de la protéine. La configuration  $r_i(t_j)$  est extraite de la simulation pour être comparée à  $r_i^{ref}$  qui est moyenné sur un intervalle de temps  $T$  de 1 ps.

$$RMSf_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (r_i(t_j) - r_i^{ref})^2} \quad \text{où } t_j = 2ps \quad \text{avec } r_i^{ref} = \langle r_i \rangle \quad (\text{VI.18})$$

La différence entre ces deux observables réside dans leur mesure, d'une part contre une structure cristalline pour le RMSd et d'autre part contre une structure dont la moyenne des coordonnées des atomes a été calculée pour le RMSf. La valeur RMSf permet le calcul du facteur de température  $B$  pour chaque résidu  $i$  traduisant l'agitation thermique.

$$B_i = \frac{8}{3} \pi^2 (RMSf_i)^2 \quad (\text{VI.19})$$

# Bibliographie

- [Adam et al., 2009] Adam, V., Carpentier, P., Violot, S., Lelimousin, M., Darnault, C., Nienhaus, G. U., and Bourgeois, D. (2009). Structural basis of x-ray-induced transient photo-bleaching in a photoactivatable green fluorescent protein. *Journal of the American Chemical Society*, 131(50) :18063–18065.
- [Agmon, 2005] Agmon, N. (2005). Proton pathways in green fluorescence protein. *Biophysical journal*, 88(4) :2452–2461.
- [Alieva et al., 2008] Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., Miller, D. J., Wiedenmann, J., Salih, A., and Matz, M. V. (2008). Diversity and evolution of coral fluorescent proteins. *PLoS one*, 3(7) :e2680.
- [Anandakrishnan et al., 2012] Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0 : automating pk prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research*, 40(W1) :W537–W541.
- [Anderson and Parrish, 1981] Anderson, R. R. and Parrish, J. A. (1981). The optics of human skin. *Journal of Investigative Dermatology*, 77(1) :13–19.
- [Artemova and Redon, 2012] Artemova, S. and Redon, S. (2012). Adaptively restrained particle simulations. *Physical review letters*, 109(19) :190201.
- [Ausaf Ali et al., 2014] Ausaf Ali, S., Hassan, I., Islam, A., Ahmad, F., et al. (2014). A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. *Current Protein and Peptide Science*, 15(5) :456–476.
- [Baker and Hubbard, 1984] Baker, E. and Hubbard, R. (1984). Hydrogen bonding in globular proteins. *Progress in biophysics and molecular biology*, 44(2) :97–179.
- [Baumann et al., 2008] Baumann, D., Cook, M., Ma, L., Mushegian, A., Sanders, E., Schwartz, J., and Yu, C. R. (2008). A family of gfp-like proteins with different spectral properties in lancelet branchiostoma floridae. *Biology Direct*, 3 :28–28.
- [Bell et al., 2000] Bell, A. F., He, X., Wachter, R. M., and Tonge, P. J. (2000). Probing the ground state structure of the green fluorescent protein chromophore using raman spectroscopy. *Biochemistry*, 39(15) :4423–4431.
- [Bergfors, 2003] Bergfors, T. (2003). Seeds to crystals. *Journal of structural biology*, 142(1) :66–76.
- [Bizzarri et al., 2007] Bizzarri, R., Nifosì, R., Abbruzzetti, S., Rocchia, W., Guidi, S., Arosio, D., Garau, G., Campanini, B., Grandi, E., Ricci, F., et al. (2007). Green fluorescent protein ground states : the influence of a second protonation site near the chromophore. *Biochemistry*, 46(18) :5494–5504.
- [Bomati et al., 2014] Bomati, E. K., Haley, J. E., Noel, J. P., and Deheyn, D. D. (2014). Spectral and structural comparison between bright and dim green fluorescent proteins in amphioxus. *Sci Rep*, 4 :5469.
- [Bondar et al., 2004] Bondar, A.-N., Elstner, M., Suhai, S., Smith, J. C., and Fischer, S. (2004). Mechanism of primary proton transfer in bacteriorhodopsin. *Structure*, 12(7) :1281–1288.

- [Bork et al., 2015] Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., and Wincker, P. (2015). Tara oceans studies plankton at planetary scale. *Science*, 348(6237) :873–873.
- [Bregestovski et al., 2009] Bregestovski, P., Waseem, T., and Mukhtarov, M. (2009). Genetically encoded optical sensors for monitoring of intracellular chloride and chloride-selective channel activity. *Front. Mol. Neurosci*, 2(15).
- [Brejc et al., 1997] Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J. (1997). Structural basis for dual excitation and photoisomerization of the aequorea victoria green fluorescent protein. *Proceedings of the National Academy of Sciences*, 94(6) :2306–2311.
- [Briinger, 1992] Briinger, A. T. (1992). Free r value : a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355 :472–475.
- [Carpentier et al., 2007] Carpentier, P., Royant, A., Ohana, J., and Bourgeois, D. (2007). Advances in spectroscopic methods for biological crystals. 2. raman spectroscopy. *Journal of Applied Crystallography*, 40(6) :1113–1122.
- [Carugo and Bordo, 1999] Carugo, O. and Bordo, D. (1999). How many water molecules can be detected by protein crystallography ? *Acta Crystallographica Section D : Biological Crystallography*, 55(2) :479–483.
- [Casey et al., 2010] Casey, J. R., Grinstein, S., and Orlowski, J. (2010). Sensors and regulators of intracellular ph. *Nature reviews Molecular cell biology*, 11(1) :50–61.
- [Chalfie, 2009] Chalfie, M. (2009). Gfp : Lighting up life (nobel lecture). *Angewandte Chemie International Edition*, 48(31) :5603–5611.
- [Chalfie et al., 1994] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., and Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. *Science*, 263(5148) :802–805.
- [Chattoraj et al., 1996] Chattoraj, M., King, B. A., Bublitz, G. U., and Boxer, S. G. (1996). Ultra-fast excited state dynamics in green fluorescent protein : multiple states and proton transfer. *Proceedings of the National Academy of Sciences*, 93(16) :8362–8367.
- [Chaudhury and Gray, 2008] Chaudhury, S. and Gray, J. J. (2008). Conformer selection and induced fit in flexible backbone protein–protein docking using computational and nmr ensembles. *Journal of molecular biology*, 381(4) :1068–1087.
- [Chothia and Lesk, 1986] Chothia, C. and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, 5(4) :823.
- [Conway et al., 2014] Conway, J. R., Carragher, N. O., and Timpson, P. (2014). Developments in preclinical cancer imaging : innovating the discovery of therapeutics. *Nature Reviews Cancer*, 14(5) :314–328.
- [Cornell et al., 1995] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19) :5179–5197.
- [Cubitt et al., 1995] Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends in biochemical sciences*, 20(11) :448–455.
- [Day and Davidson, 2014] Day, R. N. and Davidson, M. W. (2014). *The Fluorescent Protein Revolution*. CRC Press.

- [de Sanctis et al., 2012] de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., et al. (2012). Id29 : a high-intensity highly automated esrf beamline for macromolecular crystallography experiments exploiting anomalous scattering. *Journal of synchrotron radiation*, 19(3) :455–461.
- [Deheyn et al., 2007] Deheyn, D. D., Kubokawa, K., McCarthy, J. K., Murakami, A., Porrachia, M., Rouse, G. W., and Holland, N. D. (2007). Endogenous green fluorescent protein (gfp) in amphioxus. *The Biological Bulletin*, 213(2) :95–100.
- [Deheyn and Latz, 2007] Deheyn, D. D. and Latz, M. I. (2007). Bioluminescence characteristics of a tropical terrestrial fungus (basidiomycetes). *Luminescence*, 22(5) :462–467.
- [DeLano, 2002] DeLano, W. L. (2002). The pymol molecular graphics system.
- [Didenko, 2001] Didenko, V. V. (2001). Dna probes using fluorescence resonance energy transfer (fret) : designs and applications. *Biotechniques*, 31(5) :1106.
- [Dimasi et al., 2007] Dimasi, N., Flot, D., Dupeux, F., and Márquez, J. A. (2007). Expression, crystallization and x-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells irem-1. *Acta Crystallographica Section F : Structural Biology and Crystallization Communications*, 63(3) :204–208.
- [Dolinsky et al., 2004] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). Pdb2pqr : an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research*, 32(suppl 2) :W665–W667.
- [Drenth, 1999] Drenth, J. (1999). Crystallizing a protein. In *Principles of Protein X-ray Crystallography*, pages 1–21. Springer.
- [Drobizhev et al., 2015] Drobizhev, M., Callis, P., Nifosì, R., Wicks, G., Stoltzfus, C., Barnett, L., Hughes, T., Sullivan, P., and Rebane, A. (2015). Long-and short-range electrostatic fields in gfp mutants : Implications for spectral tuning. *Scientific reports*, 5.
- [Dunn et al., 2011] Dunn, K. W., Kamocka, M. M., and McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, 300(4) :C723–C742.
- [Ellenbroek and van Rheenen, 2014] Ellenbroek, S. I. and van Rheenen, J. (2014). Imaging hallmarks of cancer in living mice. *Nature Reviews Cancer*, 14(6) :406–418.
- [Elslinger et al., 1999] Elslinger, M.-A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999). Structural and spectral response of green fluorescent protein variants to changes in ph. *Biochemistry*, 38(17) :5296–5301.
- [Emsley and Cowtan, 2004] Emsley, P. and Cowtan, K. (2004). Coot : model-building tools for molecular graphics. *Acta Crystallographica Section D : Biological Crystallography*, 60(12) :2126–2132.
- [Erard et al., 2013] Erard, M., Fredj, A., Pasquier, H., Beltolngar, D.-B., Bousmah, Y., Derrien, V., Vincent, P., and Merola, F. (2013). Minimum set of mutations needed to optimize cyan fluorescent proteins for live cell imaging. *Molecular BioSystems*, 9(2) :258–267.
- [Erez et al., 2011] Erez, Y., Gepshtein, R., Presiado, I., Trujillo, K., Kallio, K., Remington, S. J., and Huppert, D. (2011). Structure and excited-state proton transfer in the gfp s205a mutant. *The Journal of Physical Chemistry B*, 115(41) :11776–11785.
- [Evdokimov et al., 2006] Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkorporov, A. N., Sander, I., Lukyanov, K. A., and Chudakov, D. M. (2006). Structural basis for the fast maturation of arthropoda green fluorescent protein. *EMBO reports*, 7(10) :1006–1012.



- [Fredj, 2012] Fredj, A. (2012). *Elaboration de protéines fluorescentes ayant un fort potentiel en imagerie*. PhD thesis, Université Paris Sud-Paris XI.
- [Galiotta et al., 2001] Galiotta, L. J., Haggie, P. M., and Verkman, A. (2001). Green fluorescent protein-based halide indicators with improved chloride and iodide affinities. *FEBS letters*, 499(3) :220–224.
- [Geer et al., 2009] Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S. H. (2009). The ncbi biosystems database. *Nucleic acids research*, page gkp858.
- [Gerschel, 2012] Gerschel, A. (2012). *Liaisons intermoléculaires-Les forces en jeu dans la matière condensée*. EDP Sciences.
- [Goedhart et al., 2012] Goedhart, J., von Stetten, D., Noirclerc-Savoye, M., Lelimosin, M., Joosen, L., Hink, M. A., van Weeren, L., Gadella Jr, T. W., and Royant, A. (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications*, 3 :751.
- [Gordon et al., 2005] Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++ : a server for estimating pkas and adding missing hydrogens to macromolecules. *Nucleic acids research*, 33(suppl 2) :W368–W371.
- [Griesbeck et al., 2001] Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A., and Tsien, R. Y. (2001). Reducing the environmental sensitivity of yellow fluorescent protein mechanism and applications. *Journal of Biological Chemistry*, 276(31) :29188–29194.
- [Haas and Drenth, 1999] Haas, C. and Drenth, J. (1999). Understanding protein crystallization on the basis of the phase diagram. *Journal of crystal growth*, 196(2) :388–394.
- [Haddock, 2016] Haddock, S. H. (2016). The bioluminescence web page.
- [He et al., 2002] He, X., Bell, A. F., and Tonge, P. J. (2002). Isotopic labeling and normal-mode analysis of a model green fluorescent protein chromophore. *The Journal of Physical Chemistry B*, 106(23) :6056–6066.
- [Heim et al., 1994] Heim, R., Prasher, D. C., and Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences*, 91(26) :12501–12504.
- [Heim and Tsien, 1996] Heim, R. and Tsien, R. Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology*, 6(2) :178–182.
- [Hubbard and Kamran Haider, 2010] Hubbard, R. E. and Kamran Haider, M. (2010). Hydrogen bonds in proteins : role and strength. *eLS*.
- [Hubbs, 1922] Hubbs, C. L. (1922). A list of the lancelets of the world with diagnosis of five new species of branchiostoma.
- [Humphrey et al., 1996] Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd : visual molecular dynamics. *Journal of molecular graphics*, 14(1) :33–38.
- [Hunt et al., 2010] Hunt, M. E., Scherrer, M. P., Ferrari, F. D., and Matz, M. V. (2010). Very bright green fluorescent proteins from the pontellid copepod pontella mimocerami. *PloS one*, 5(7) :e11517.
- [Ippolito et al., 1990] Ippolito, J. A., Alexander, R. S., and Christianson, D. W. (1990). Hydrogen bond stereochemistry in protein structure and function. *Journal of molecular biology*, 215(3) :457–471.

- [Israelsson, 2013] Israelsson, O. (2013). Fluorescent proteins and genes encoding them. US Patent 8,609,393.
- [Jabłoński, 1935] Jabłoński, A. (1935). Über den mechanismus der photolumineszenz von farbstoffphosphoren. *Zeitschrift für Physik*, 94(1-2) :38–46.
- [Jares-Erijman and Jovin, 2003] Jares-Erijman, E. A. and Jovin, T. M. (2003). Fret imaging. *Nature biotechnology*, 21(11) :1387–1395.
- [Jares-Erijman and Jovin, 2006] Jares-Erijman, E. A. and Jovin, T. M. (2006). Imaging molecular interactions in living cells by fret microscopy. *Current opinion in chemical biology*, 10(5) :409–416.
- [Jimenez-Morales et al., 2014] Jimenez-Morales, D., Adamian, L., Shi, D., and Liang, J. (2014). Lysine carboxylation : unveiling a spontaneous post-translational modification. *Acta Crystallographica Section D : Biological Crystallography*, 70(1) :48–57.
- [Jonasson et al., 2011] Jonasson, G., Teuler, J.-M., Vallverdu, G., Mérola, F., Ridard, J., Lévy, B., and Demachy, I. (2011). Excited state dynamics of the green fluorescent protein on the nanosecond time scale. *Journal of chemical theory and computation*, 7(6) :1990–1997.
- [Jorgensen et al., 1983] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2) :926–935.
- [Kabsch, 2010] Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D : Biological Crystallography*, 66(2) :133–144.
- [Kabsch and Sander, 1983] Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure : pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12) :2577–2637.
- [Kanda et al., 1998] Kanda, T., Sullivan, K. F., and Wahl, G. M. (1998). Histone–gfp fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Current biology*, 8(7) :377–385.
- [Kandori, 2000] Kandori, H. (2000). Role of internal water molecules in bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1460(1) :177–191.
- [Kennis et al., 2004] Kennis, J. T., Larsen, D. S., van Stokkum, I. H., Vengris, M., van Thor, J. J., and van Grondelle, R. (2004). Uncovering the hidden ground state of green fluorescent protein. *Proceedings of the National Academy of Sciences*, 101(52) :17988–17993.
- [Kobayashi et al., 2009] Kobayashi, H., Ogawa, M., Alford, R., Choyke, P. L., and Urano, Y. (2009). New strategies for fluorescent probe design in medical diagnostic imaging. *Chemical reviews*, 110(5) :2620–2640.
- [Kremers et al., 2011] Kremers, G.-J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W., and Piston, D. W. (2011). Fluorescent proteins at a glance. *J Cell Sci*, 124(2) :157–160.
- [Kremers et al., 2006] Kremers, G.-J., Goedhart, J., van Munster, E. B., and Gadella, T. W. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and fret förster radius. *Biochemistry*, 45(21) :6570–6580.
- [Krissinel and Henrick, 2007] Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, 372(3) :774–797.
- [Lakowicz, 2006] Lakowicz, J. R. (2006). Principles of fluorescence spectroscopy.
- [Lakowicz, 2013] Lakowicz, J. R. (2013). *Principles of fluorescence spectroscopy*. Springer Science & Business Media.

- [Langhojer et al., 2009] Langhojer, F., Dimler, F., Jung, G., and Brixner, T. (2009). Ultrafast photoconversion of the green fluorescent protein studied by accumulative femtosecond spectroscopy. *Biophysical journal*, 96(7) :2763–2770.
- [Laskowski et al., 1996] Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). Aqua and procheck-nmr : programs for checking the quality of protein structures solved by nmr. *Journal of biomolecular NMR*, 8(4) :477–486.
- [Ledford, 2013] Ledford, H. (2013). Bioengineers look beyond patents. *Nature*, 499(7456) :16–17.
- [Lee and Richards, 1971] Lee, B. and Richards, F. M. (1971). The interpretation of protein structures : estimation of static accessibility. *Journal of molecular biology*, 55(3) :379–IN4.
- [Lelimosin et al., 2009] Lelimosin, M., Noirclerc-Savoie, M., Lazareno-Saez, C., Paetzold, B., Le Vot, S., Chazal, R., Macheboeuf, P., Field, M. J., Bourgeois, D., and Royant, A. (2009). Intrinsic dynamics in ecfp and cerulean control fluorescence quantum yield. *Biochemistry*, 48(42) :10038–10046.
- [Lill and Helms, 2002] Lill, M. A. and Helms, V. (2002). Proton shuttle in green fluorescent protein studied by dynamic simulations. *Proceedings of the National Academy of Sciences*, 99(5) :2778–2781.
- [Loncharich et al., 1992] Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992). Langevin dynamics of peptides : The frictional dependence of isomerization rates of n-acetylalanyl-n- $\alpha$ -methylamide. *Biopolymers*, 32(5) :523–535.
- [Lovell et al., 2000] Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000). The penultimate rotamer library. *Proteins : Structure, Function, and Bioinformatics*, 40(3) :389–408.
- [Lukyanov, 2013] Lukyanov, S. A. (2013). Bioimaging studies in situ and in vivo.
- [Masuda et al., 2006] Masuda, H., Takenaka, Y., Yamaguchi, A., Nishikawa, S., and Mizuno, H. (2006). A novel yellowish-green fluorescent protein from the marine copepod, *chiridius poppei*, and its use as a reporter protein in hela cells. *Gene*, 372 :18–25.
- [Matz et al., 1999] Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999). Fluorescent proteins from nonbioluminescent anthozoa species. *Nat Biotech*, 17(10) :969–973.
- [McCoy, 2007] McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with phaser. *Acta Crystallographica Section D : Biological Crystallography*, 63(1) :32–41.
- [McCoy et al., 2007] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography*, 40(4) :658–674.
- [McGeehan et al., 2011] McGeehan, J. E., Bourgeois, D., Royant, A., and Carpentier, P. (2011). Raman-assisted crystallography of biomolecules at the synchrotron : Instrumentation, methods and applications. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(6) :750–759.
- [McPherson, 2011] McPherson, A. (2011). *Introduction to macromolecular crystallography*. John Wiley & Sons.
- [Medvedev et al., 2014] Medvedev, K. E., Alemasov, N. A., Vorobjev, Y. N., Boldyreva, E. V., Kolchanov, N. A., and Afonnikov, D. A. (2014). Molecular dynamics simulations of the

- nip7 proteins from the marine deep-and shallow-water pyrococcus species. *BMC structural biology*, 14(1) :1.
- [Merola et al., 2010] Merola, F., Levy, B., Demachy, I., and Pasquier, H. (2010). Photophysics and spectroscopy of fluorophores in the green fluorescent protein family. In *Advanced Fluorescence Reporters in Chemistry and Biology I*, pages 347–383. Springer.
- [Mishin et al., 2008] Mishin, A. S., Subach, F. V., Yampolsky, I. V., King, W., Lukyanov, K. A., and Verkhusha, V. V. (2008). The first mutant of the aequeorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry*, 47(16) :4666–4673.
- [Miyamoto and Kollman, 1992] Miyamoto, S. and Kollman, P. A. (1992). Settle : an analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry*, 13(8) :952–962.
- [Miyawaki et al., 1999] Miyawaki, A., Griesbeck, O., Heim, R., and Tsien, R. Y. (1999). Dynamic and quantitative ca<sup>2+</sup> measurements using improved cameleons. *Proceedings of the National Academy of Sciences*, 96(5) :2135–2140.
- [Miyawaki et al., 1997] Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997). Fluorescent indicators for ca<sup>2+</sup> ; based on green fluorescent proteins and calmodulin. *Nature*, 388(6645) :882–887.
- [Miyawaki et al., 2003] Miyawaki, A., Nagai, T., and Mizuno, H. (2003). Mechanisms of protein fluorophore formation and engineering. *Current opinion in chemical biology*, 7(5) :557–562.
- [Murbach and Shearer, 1902] Murbach, L. and Shearer, C. (1902). Xvi.—preliminary report on a collection of medusæ from the coast of british columbia and alaska. *Journal of Natural History*, 9(49) :71–73.
- [Murshudov et al., 1997] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D : Biological Crystallography*, 53(3) :240–255.
- [Mushegian et al., 2014] Mushegian, A., Yu, C. R., Schwartz, J., Ma, L., and Cook, M. (2014). Branchiostoma derived fluorescent proteins. US Patent 8,680,235.
- [Nagai et al., 2002] Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology*, 20(1) :87–90.
- [Niwa et al., 1996] Niwa, H., Inouye, S., Hirano, T., Matsuno, T., Kojima, S., Kubota, M., Ohashi, M., and Tsuji, F. I. (1996). Chemical nature of the light emitter of the aequeorea green fluorescent protein. *Proceedings of the National Academy of Sciences*, 93(24) :13617–13622.
- [Ormö et al., 1996] Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996). Crystal structure of the aequeorea victoria green fluorescent protein. *Science*, 273(5280) :1392–1395.
- [Pace et al., 2014] Pace, C. N., Fu, H., Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., et al. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, 23(5) :652–661.
- [Paddock et al., 2003] Paddock, M., Feher, G., and Okamura, M. (2003). Proton transfer pathways and mechanism in bacterial reaction centers. *FEBS letters*, 555(1) :45–50.
- [Pallas, 1774] Pallas, P. (1774). Spicilegia zoologica, quibus novae imprimis et obscurae animalium species iconibus, descriptionibus atque commentariis illustrantur. *Fasc X, Berolini*.

- [Palm et al., 1997] Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997). The structural basis for spectral variations in green fluorescent protein. *Nature Structural & Molecular Biology*, 4(5) :361–365.
- [Pastor et al., 1988] Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of langevin and molecular dynamics algorithms. *Molecular Physics*, 65(6) :1409–1419.
- [Pavelka et al., 2016] Pavelka, A., Sebestova, E., Kozlikova, B., Brezovsky, J., Sochor, J., and Damborsky, J. (2016). Caver : Algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(3) :505–517.
- [Peng and Voth, 2012] Peng, Y. and Voth, G. A. (2012). Expanding the view of proton pumping in cytochrome c oxidase through computer simulation. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1817(4) :518–525.
- [Perrin, 1926] Perrin, F. (1926). Polarisation de la lumière de fluorescence. vie moyenne des molécules dans l'état excité. *J. phys. radium*, 7(12) :390–401.
- [Phair and Misteli, 2000] Phair, R. D. and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778) :604–609.
- [Pitman et al., 2015] Pitman, D. J., Banerjee, S., Macari, S. J., Castaldi, C. A., Crone, D. E., and Bystroff, C. (2015). Exploring the folding pathway of green fluorescent protein through disulfide engineering. *Protein Science*, 24(3) :341–353.
- [Pletnev et al., 2013] Pletnev, V. Z., Pletneva, N. V., Lukyanov, K. A., Souslova, E. A., Fradkov, A. F., Chudakov, D. M., Chepurnykh, T., Yampolsky, I. V., Wlodawer, A., Dauter, Z., and Pletnev, S. (2013). Structure of the red fluorescent protein from a lancelet (branchiostoma lanceolatum) : a novel gyg chromophore covalently bound to a nearby tyrosine. *Acta Crystallogr D Biol Crystallogr*, 69(Pt 9) :1850–60.
- [Pletneva et al., 2013] Pletneva, N. V., Pletnev, V. Z., Souslova, E., Chudakov, D. M., Lukyanov, S., Martynov, V. I., Arhipova, S., Artemyev, I., Wlodawer, A., Dauter, Z., et al. (2013). Yellow fluorescent protein phiyfpv (phialidium) : structure and structure-based mutagenesis. *Acta Crystallographica Section D : Biological Crystallography*, 69(6) :1005–1012.
- [Polenzani et al., 1991] Polenzani, L., Woodward, R., and Miledi, R. (1991). Expression of mammalian gamma-aminobutyric acid receptors with distinct pharmacology in xenopus oocytes. *Proceedings of the National Academy of Sciences*, 88(10) :4318–4322.
- [Poll, 1947] Poll, M. (1947). *Poissons marins*. Patrimoine du Musée royal d'histoire naturelle de Belgique.
- [Prasher et al., 1992] Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992). Primary structure of the aequorea victoria green-fluorescent protein. *Gene*, 111(2) :229–33.
- [Putnam et al., 2008] Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 453(7198) :1064–1071.
- [Ramachandran et al., 1963] Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of molecular biology*, 7(1) :95–99.

- [Raman and Krishnan, 1928] Raman, C. V. and Krishnan, K. S. (1928). A new type of secondary radiation. *Nature*, 121 :501–502.
- [Read et al., 2011] Read, R. J., Adams, P. D., Arendall, W. B., Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütke, T., Otwinowski, Z., et al. (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure*, 10(19) :1395–1412.
- [Rizzo et al., 2004] Rizzo, M. A., Springer, G. H., Granada, B., and Piston, D. W. (2004). An improved cyan fluorescent protein variant useful for fret. *Nature biotechnology*, 22(4) :445–449.
- [Robert and Gouet, 2014] Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new endscript server. *Nucleic acids research*, 42(W1) :W320–W324.
- [Rossmann, 1990] Rossmann, M. G. (1990). The molecular replacement method. *Acta Crystallographica Section A : Foundations of Crystallography*, 46(2) :73–82.
- [Roy et al., 2011] Roy, A., Field, M. J., Adam, V., and Bourgeois, D. (2011). The nature of transient dark states in a photoactivatable fluorescent protein. *Journal of the American Chemical Society*, 133(46) :18586–18589.
- [Royant and Noirclerc-Savoie, 2011] Royant, A. and Noirclerc-Savoie, M. (2011). Stabilizing role of glutamic acid 222 in the structure of enhanced green fluorescent protein. *Journal of structural biology*, 174(2) :385–390.
- [Ryckaert et al., 1977] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3) :327–341.
- [Sanchez-Weatherby et al., 2009] Sanchez-Weatherby, J., Bowler, M. W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R. B., and Cipriani, F. (2009). Improving diffraction by humidity control : a novel device compatible with x-ray beamlines. *Acta Crystallographica Section D : Biological Crystallography*, 65(12) :1237–1246.
- [Schellenberg et al., 2001] Schellenberg, P., Johnson, E., Esposito, A. P., Reid, P. J., and Parson, W. W. (2001). Resonance raman scattering by the green fluorescent protein and an analogue of its chromophore. *The Journal of Physical Chemistry B*, 105(22) :5316–5322.
- [Shagin et al., 2004] 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., et al. (2004). Gfp-like proteins as ubiquitous metazoan superfamily : evolution of functional features and structural complexity. *Molecular biology and evolution*, 21(5) :841–850.
- [Shaner et al., 2004] Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp. red fluorescent protein. *Nat Biotech*, 22(12) :1567–1572.
- [Shaner et al., 2013] Shaner, N. C., Lambert, G. G., Chamma, A., Ni, Y., Cranfill, P. J., Baird, M. A., Sell, B. R., Allen, J. R., Day, R. N., Israelsson, M., Davidson, M. W., and Wang, J. (2013). A bright monomeric green fluorescent protein derived from branchiostoma lanceolatum. *Nat Methods*, 10(5) :407–9.
- [Shaner et al., 2011] Shaner, N. C., Patterson, G. H., and Davidson, M. W. (2011). Advances in fluorescent protein technology. *Journal of Cell Science*, 124(13) :2321–2321.
- [Shen et al., 2015] Shen, Y., Lai, T., and Campbell, R. E. (2015). Red fluorescent proteins (rfps) and rfp-based biosensors for neuronal imaging applications. *Neurophotonics*, 2(3) :031203–031203.

- [Shimomura, 1979] Shimomura, O. (1979). Structure of the chromophore of aequorea green fluorescent protein. *FEBS Letters*, 104(2) :220 – 222.
- [Shimomura, 2009] Shimomura, O. (2009). Discovery of green fluorescent protein (gfp) (nobel lecture). *Angewandte Chemie International Edition*, 48(31) :5590–5602.
- [Shimomura and Johnson, 1969] Shimomura, O. and Johnson, F. H. (1969). Properties of the bioluminescent protein aequorin. *Biochemistry*, 8(10) :3991–7.
- [Shimomura and Johnson, 1970] Shimomura, O. and Johnson, F. H. (1970). Calcium binding, quantum yield, and emitting molecule in aequorin bioluminescence. *Nature*, 227(5265) :1356–7.
- [Shimomura and Johnson, 1972] Shimomura, O. and Johnson, F. H. (1972). Structure of the light-emitting moiety of aequorin. *Biochemistry*, 11(9) :1602–8.
- [Shimomura and Johnson, 1975] Shimomura, O. and Johnson, F. H. (1975). Regeneration of the photoprotein aequorin.
- [Shimomura et al., 1962] Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydro-medusan, aequorea. *J Cell Comp Physiol*, 59 :223–39.
- [Shimomura et al., 1963] Shimomura, O., Johnson, F. H., and Saiga, Y. (1963). Microdetermination of calcium by aequorin luminescence. *Science*, 140(3573) :1339–40.
- [Shinobu and Agmon, 2015] Shinobu, A. and Agmon, N. (2015). The hole in the barrel : Water exchange at the gfp chromophore. *The Journal of Physical Chemistry B*, 119(8) :3464–3478.
- [Shinobu et al., 2010] Shinobu, A., Palm, G. J., Schierbeek, A. J., and Agmon, N. (2010). Visualizing proton antenna in a high-resolution green fluorescent protein structure. *Journal of the American Chemical Society*, 132(32) :11093–11102.
- [Shu et al., 2007] Shu, X., Leiderman, P., Gepshtein, R., Smith, N. R., Kallio, K., Huppert, D., and Remington, S. J. (2007). An alternative excited-state proton transfer pathway in green fluorescent protein variant s205v. *Protein Science*, 16(12) :2703–2710.
- [Sievers and Higgins, 2014] Sievers, F. and Higgins, D. G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. *Multiple sequence alignment methods*, pages 105–116.
- [Simkovitch et al., 2013] Simkovitch, R., Huppert, A., Huppert, D., Remington, S. J., and Miller, Y. (2013). Proton transfer in wild-type gfp and s205v mutant is reduced by conformational changes of residues in the proton wire. *The Journal of Physical Chemistry B*, 117(40) :11921–11931.
- [Sniegowski et al., 2005] Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005). Base catalysis of chromophore formation in arg96 and glu222 variants of green fluorescent protein. *Journal of Biological Chemistry*, 280(28) :26248–26255.
- [Stoner-Ma et al., 2005] Stoner-Ma, D., Jaye, A. A., Matousek, P., Towrie, M., Meech, S. R., and Tonge, P. J. (2005). Observation of excited-state proton transfer in green fluorescent protein using ultrafast vibrational spectroscopy. *Journal of the American Chemical Society*, 127(9) :2864–2865.
- [Stoner-Ma et al., 2008] Stoner-Ma, D., Jaye, A. A., Ronayne, K. L., Nappa, J., Meech, S. R., and Tonge, P. J. (2008). An alternate proton acceptor for excited-state proton transfer in green fluorescent protein : rewiring gfp. *Journal of the American Chemical Society*, 130(4) :1227–1235.

- [Striker et al., 1999] Striker, G., Subramaniam, V., Seidel, C. A., and Volkmer, A. (1999). Photochromicity and fluorescence lifetimes of green fluorescent protein. *The Journal of Physical Chemistry B*, 103(40) :8612–8617.
- [Stumpe et al., 2010] Stumpe, M. C., Blinov, N., Wishart, D., Kovalenko, A., and Pande, V. S. (2010). Calculation of local water densities in biological systems : a comparison of molecular dynamics simulations and the 3d-rism-kh molecular theory of solvation. *The journal of physical chemistry B*, 115(2) :319–328.
- [Suto et al., 2009] Suto, K., Masuda, H., Takenaka, Y., Tsuji, F. I., and Mizuno, H. (2009). Structural basis for red-shifted emission of a gfp-like protein from the marine copepod *chiridius poppei*. *Genes to Cells*, 14(6) :727–737.
- [Taraphder and Hummer, 2003] Taraphder, S. and Hummer, G. (2003). Protein side-chain motion and hydration in proton-transfer pathways. results for cytochrome p450cam. *Journal of the American Chemical Society*, 125(13) :3931–3940.
- [Tolbert and Solntsev, 2002] Tolbert, L. M. and Solntsev, K. M. (2002). Excited-state proton transfer : from constrained systems to “super” photoacids to superfast proton transfer. *Accounts of chemical research*, 35(1) :19–27.
- [Tsien, 1998] Tsien, R. Y. (1998). The green fluorescent protein. *Annual review of biochemistry*, 67(1) :509–544.
- [Tsien, 1999] Tsien, R. Y. (1999). Rosy dawn for fluorescent proteins. *Nature biotechnology*, 17(10) :956–957.
- [Tsien, 2009a] Tsien, R. Y. (2009a). Constructing and exploiting the fluorescent protein paint-box (nobel lecture). *Angewandte Chemie International Edition*, 48(31) :5612–5626.
- [Tsien, 2009b] Tsien, R. Y. (2009b). Constructing and exploiting the fluorescent protein paint-box (nobel lecture). *Angewandte Chemie International Edition*, 48(31) :5612–5626.
- [Valeur and Berberan-Santos, 2012] Valeur, B. and Berberan-Santos, M. N. (2012). *Molecular fluorescence : principles and applications*. John Wiley & Sons.
- [Vallverdu et al., 2010] Vallverdu, G., Demachy, I., Mérola, F., Pasquier, H., Ridard, J., and Lévy, B. (2010). Relation between ph, structure, and absorption spectrum of cerulean : A study by molecular dynamics and td dft calculations. *Proteins : Structure, Function, and Bioinformatics*, 78(4) :1040–1054.
- [van Thor et al., 2002] van Thor, J. J., Gensch, T., Hellingwerf, K. J., and Johnson, L. N. (2002). Phototransformation of green fluorescent protein with uv and visible light leads to decarboxylation of glutamate 222. *Nature Structural & Molecular Biology*, 9(1) :37–41.
- [van Thor et al., 1998] van Thor, J. J., Pierik, A. J., Nugteren-Roodzant, I., Xie, A., and Hellingwerf, K. J. (1998). Characterization of the photoconversion of green fluorescent protein with ftir spectroscopy. *Biochemistry*, 37(48) :16915–16921.
- [van Thor et al., 2006] van Thor, J. J., Towrie, M., Ronayne, K., Georgiev, G. Y., and Sage, J. T. (2006). Ultrafast and low-barrier motions in the photoreactions of the green fluorescent protein. In *Biomedical Optics 2006*, pages 609806–609806. International Society for Optics and Photonics.
- [Villoing et al., 2008] Villoing, A., Ridhoir, M., Cinquin, B., Erard, M., Alvarez, L., Vallverdu, G., Pernot, P., Grailhe, R., Mérola, F., and Pasquier, H. (2008). Complex fluorescence of the cyan fluorescent protein : Comparisons with the h148d variant and consequences for quantitative cell imaging. *Biochemistry*, 47(47) :12483–12492.



- [Violot et al., 2009] Violot, S., Carpentier, P., Blanchoin, L., and Bourgeois, D. (2009). Reverse pH-dependence of chromophore protonation explains the large Stokes shift of the red fluorescent protein mKeima. *Journal of the American Chemical Society*, 131(30) :10356–10357.
- [von Stetten et al., 2015] von Stetten, D., Giraud, T., Carpentier, P., Sever, F., Terrien, M., Dobias, F., Juers, D. H., Flot, D., Mueller-Dieckmann, C., Leonard, G. A., et al. (2015). In crystallo optical spectroscopy (icos) as a complementary tool on the macromolecular crystallography beamlines of the esrf. *Acta Crystallographica Section D : Biological Crystallography*, 71(1) :15–26.
- [Wachter et al., 1998] Wachter, R. M., Elsliger, M.-A., Kallio, K., Hanson, G. T., and Remington, S. J. (1998). Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure*, 6(10) :1267–1277.
- [Wachter and Remington, 1999] Wachter, R. M. and Remington, S. J. (1999). Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. *Current Biology*, 9(17) :R628–R629.
- [Wang et al., 2000] Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry*, 21(12) :1049–1074.
- [Wang et al., 2004] Wang, L., Jackson, W. C., Steinbach, P. A., and Tsien, R. Y. (2004). Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48) :16745–16749.
- [Wang and Hazelrigg, 1994] Wang, S. and Hazelrigg, T. (1994). Implications for bcd mRNA localization from spatial distribution of exu protein in drosophila oogenesis. *Nature*, 369(6479) :400–403.
- [Ward and Moss, 2001] Ward, B. M. and Moss, B. (2001). Visualization of intracellular movement of vaccinia virus virions containing a green fluorescent protein-b5r membrane protein chimera. *Journal of Virology*, 75(10) :4802–4813.
- [Ward, 2005] Ward, W. W. (2005). Biochemical and physical properties of green fluorescent protein. *Green fluorescent protein : properties, applications and protocols. 2nd ed. Hoboken, USA : Wiley and Sons*, pages 39–65.
- [Ward and Bokman, 1982] Ward, W. W. and Bokman, S. H. (1982). Reversible denaturation of aequorea green-fluorescent protein : physical separation and characterization of the renatured protein. *Biochemistry*, 21(19) :4535–4540.
- [Ward et al., 1981] Ward, W. W. et al. (1981). Properties of the coelenterate green-fluorescent proteins. *Bioluminescence and Chemiluminescence : Basic Chemistry and Analytical applications. Eds. DeLuca, M. & McElroy, WD (Academic Press, Inc., NY)*, pages 235–242.
- [Wilmann et al., 2006] Wilmann, P. G., Battad, J., Petersen, J., Wilce, M. C., Dove, S., Devnish, R. J., Prescott, M., and Rossjohn, J. (2006). The 2.1 Å crystal structure of copgfp, a representative member of the copepod clade within the green fluorescent protein superfamily. *Journal of molecular biology*, 359(4) :890–900.
- [Wineman-Fisher et al., 2014] Wineman-Fisher, V., Simkovitch, R., Shomer, S., Gepshtein, R., Huppert, D., Saif, M., Kallio, K., Remington, S. J., and Miller, Y. (2014). Insight into the structure and the mechanism of the slow proton transfer in the GFP double mutant t203v/s205a. *Physical Chemistry Chemical Physics*, 16(23) :11211–11223.
- [Woodcock et al., 1989] Woodcock, D., Crowther, P., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S., Michael, M., and Graham, M. (1989). Quantitative evaluation

- of escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic acids research*, 17(9) :3469–3478.
- [Xu et al., 1999] Xu, Y., Piston, D. W., and Johnson, C. H. (1999). A bioluminescence resonance energy transfer (bret) system : application to interacting circadian clock proteins. *Proceedings of the National Academy of Sciences*, 96(1) :151–156.
- [Yang et al., 1996] Yang, F., Moss, L. G., and Phillips, G. N. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology*, 14(10) :1246–1251.
- [Yang et al., 2015] Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The i-tasser suite : protein structure and function prediction. *Nature methods*, 12(1) :7–8.
- [Yarbrough et al., 2001] Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001). Refined crystal structure of dsred, a red fluorescent protein from coral, at 2.0-Å resolution. *Proceedings of the National Academy of Sciences*, 98(2) :462–467.
- [Yue et al., 2016] Yue, J.-X., Holland, N. D., Holland, L. Z., and Deheyn, D. D. (2016). The evolution of genes encoding for green fluorescent proteins : insights from cephalochordates (amphioxus). *Scientific Reports*, 6.
- [Yueh and Goff, 2003] Yueh, A. and Goff, S. P. (2003). Phosphorylated serine residues and an arginine-rich domain of the moloney murine leukemia virus p12 protein are required for early events of viral infection. *Journal of virology*, 77(3) :1820–1829.
- [Zacharias and Tsien, 2006] Zacharias, D. A. and Tsien, R. Y. (2006). Molecular biology and mutation of green fluorescent protein. *Green Fluorescent Protein : Properties, Applications, and Protocols*, 47 :83.
- [Zacharias et al., 2002] Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002). Partitioning of lipid-modified monomeric gfps into membrane microdomains of live cells. *Science*, 296(5569) :913–916.
- [Zagranichny et al., 2004] Zagranichny, V. E., Rudenko, N. V., Gorokhovatsky, A. Y., Zakharov, M. V., Shenkarev, Z. O., Balashova, T. A., and Arseniev, A. S. (2004). zfp538, a yellow fluorescent protein from coral, belongs to the dsred subfamily of gfp-like proteins but possesses the unexpected site of fragmentation. *Biochemistry*, 43(16) :4764–4772.
- [Zeldin et al., 2013] Zeldin, O. B., Gerstel, M., and Garman, E. F. (2013). Raddose-3d : Time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of applied crystallography*, 46(4) :1225–1230.
- [Zimmer, 2002] Zimmer, M. (2002). Green fluorescent protein (gfp) : applications, structure, and related photophysical behavior. *Chemical reviews*, 102(3) :759–782.
- [Zimmer, 2015] Zimmer, M. (2015). *Illuminating Disease : An Introduction to Green Fluorescent Proteins*. Oxford University Press.

# VII CHAPITRE

## Annexes

### A Longueur d'onde associée aux couleurs

Couleur	Longueur d'onde (nm)
rouge	800 - 650
orange	640 - 590
jaune	580 - 550
vert	540 - 490
bleu	480 - 440
violet	430 - 390

### B Compartiments cellulaires et pH chez un archétype de cellule de mammifère

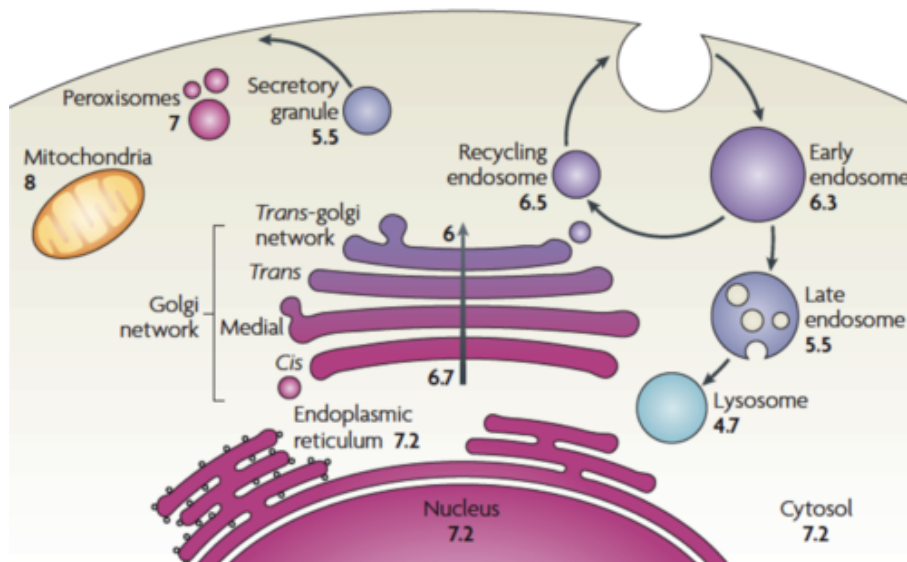
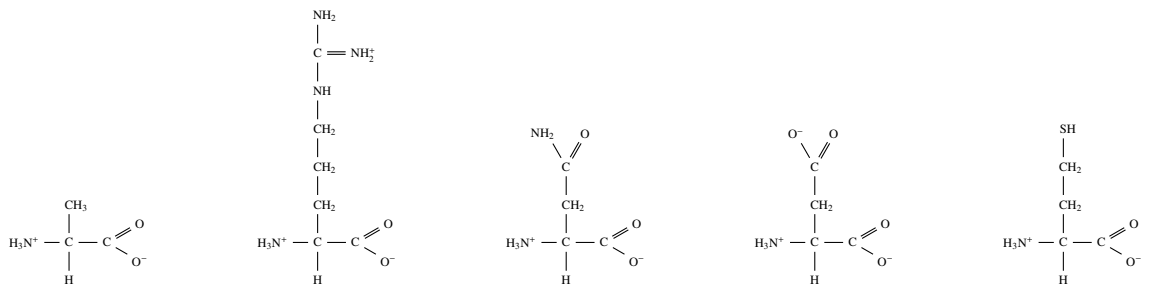


FIGURE VII.1 : pH des différents compartiments cellulaires chez un archétype de cellule de mammifère [Casey et al., 2010]

## C Table des acides aminés



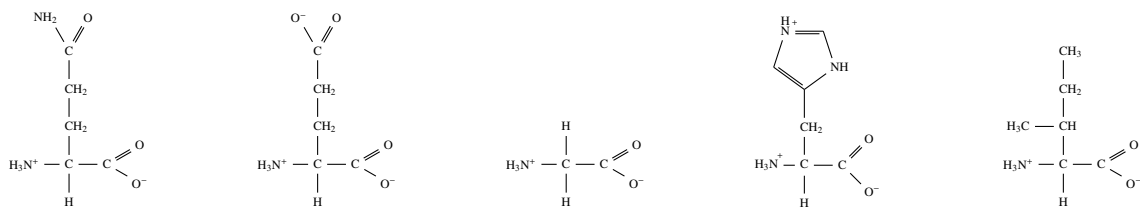
alanine  
Ala - A

arginine  
Arg - R

asparagine  
Asn - N

acide aspartique  
Asp - D

cystéine  
Cys - C



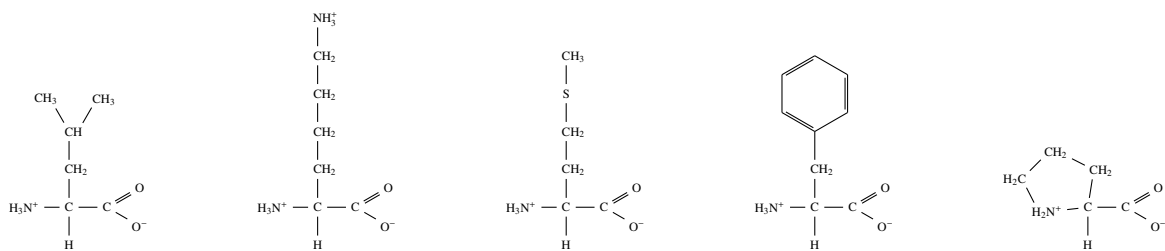
glutamine  
Gln - Q

acide glutamique  
Glu - E

glycine  
Gly - G

histidine  
His - H

isoleucine  
Ile - I



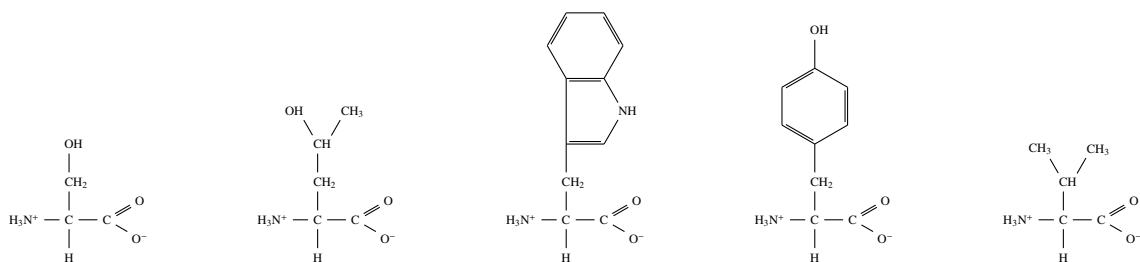
leucine  
Leu - L

lysine  
Lys - K

méthionine  
Met - M

phénylalanine  
Phe - F

proline  
Pro - P



sérine  
Ser - S

thréonine  
Thr - T

tryptophane  
Trp - W

tyrosine  
Tyr - Y

valine  
Val - V

## D Groupe d'espace

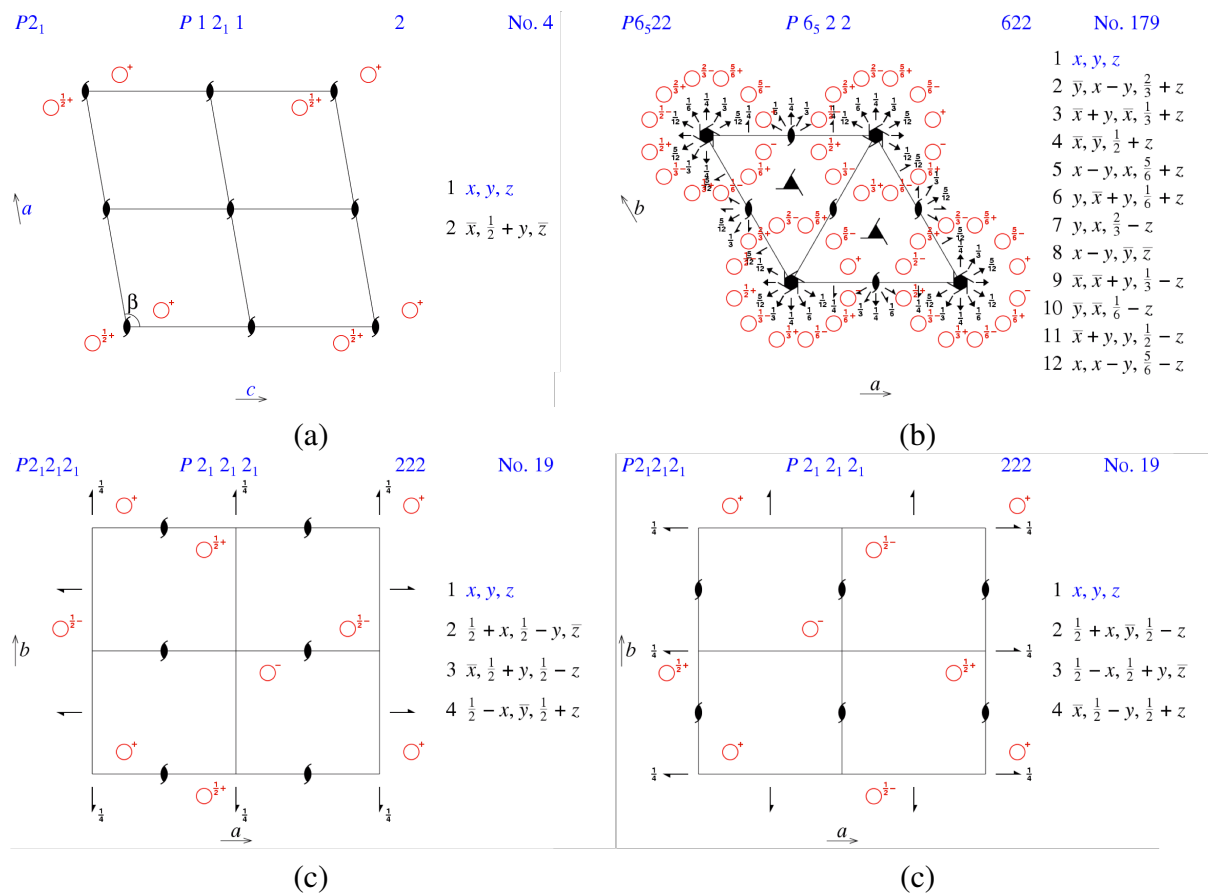


FIGURE VII.2 : Diagramme des groupes d'espace utilisés pour décrire la maille élémentaire des protéines fluorescentes mNeonGreen et lanYFP, (a)  $P12_11$ , (b)  $P6_522$ , (c)  $P2_12_12_1$ , (d)  $P2_12_12_1$



## F Etude des interfaces des structures cristallines

- mNeonGreen à pH acide

Monomère #1	Monomère #2	Opération de symétrie	Aire (Å <sup>2</sup> )	Δ G (kcal.mol <sup>-1</sup> )	N <sub>LH</sub>	N <sub>PtSal.</sub>	N <sub>PtdiSul.</sub>
F	D	X,Y,Z	726.7	-7.4	3	2	0
C	B	X,Y,Z	682.4	-3.8	5	2	0
E	D	X,Y,Z	667.9	-7.5	3	0	0
F	E	X,Y,Z	622.1	-4.9	5	2	0
C	A	X,Y,Z	596.6	-5.5	4	1	0
B	A	X,Y,Z	286.4	-2.7	0	0	0
A	F	-X+1/2,-Y,Z-1/2	546.1	-0.6	7	1	0
E	C	X,Y,Z	481.5	-3.0	3	1	0
F	A	X,Y,Z	174.9	-2.0	0	0	0
D	B	X,Y,Z	481.5	-1.5	5	1	0
E	A	X,Y,Z	91.1	0.5	1	0	0
D	C	X,Y,Z	44.9	0.9	0	0	0
F	B	X,Y,Z	12.1	0.3	0	0	0
E	B	X,Y,Z	4.1	-0.0	0	0	0
...	...	...	...	...	...	...	...

Tableau VII.2 : Interfaces dans le dimère de trimère de mNeonGreen à pH acide [Krissinel and Henrick, 2007]

- mNeonGreen à pH physiologique

Monomère #1	Monomère #2	Opération de symétrie	Aire (Å <sup>2</sup> )	Δ G (kcal.mol <sup>-1</sup> )	N <sub>LH</sub>	N <sub>PtSal.</sub>	N <sub>PtdiSul.</sub>
A	A	-Y+1,-X+1,-Z+1/6	742.0	-5.0	2	10	0
A	A	X-Y,X,Z-1/6	561.9	-5.9	7	0	0
Chromophore A	A	X,Y,Z	289.5	1.7	0	0	0
A	A	-X+Y,Y,-Z+1/2	217.7	-1.8	0	0	0
A	A	-X+1,-X+Y,-Z+1/3	97.9	-0.6	0	0	0
A	A	X-Y,-Y,-Z	3.9	0.1	0	0	0

Tableau VII.3 : Interfaces de mNeonGreen avec ses symétriques à pH physiologique [Krissinel and Henrick, 2007]

• *lanYFP* à pH physiologique

Monomère #1	Monomère #2	Opération de symétrie	Aire (Å <sup>2</sup> )	$\Delta G$ (kcal.mol <sup>-1</sup> )	$N_{LH}$	$N_{PtSal.}$	$N_{PtdiSul.}$
H	F	X,Y,Z	1186.2	-14.9	8	6	0
L	J	X,Y,Z	1155.6	-15.6	7	6	0
K	I	X,Y,Z	1062.6	-15.1	6	4	0
P	N	X,Y,Z	1055.1	-16.3	6	4	0
D	B	X,Y,Z	1047.2	-14.6	5	6	0
G	E	X,Y,Z	1045.0	-15.5	7	6	0
C	A	X,Y,Z	1034.4	-14.3	5	6	0
O	M	X,Y,Z	796.1	-4.4	6	2	0
H	G	X,Y,Z	896.2	-4.9	9	0	0
B	A	X,Y,Z	883.1	-5.6	9	0	0
J	I	X,Y,Z	870.2	-4.2	7	0	0
N	M	X,Y,Z	869.3	-4.5	6	0	0
P	O	X,Y,Z	866.7	-6.5	11	0	0
L	K	X,Y,Z	857.1	-5.3	15	0	0
F	E	X,Y,Z	855.2	-5.2	7	0	0
D	C	X,Y,Z	853.5	-5.7	9	0	0
A	D	X-1,Y,Z	638.1	-1.3	10	0	0
E	J	X-1,Y,Z	444.9	-1.2	8	0	0
F	I	X-1,Y,Z+1	404.9	-3.0	1	0	0
L	C	X,Y,Z	317.7	0.3	1	2	0
...	...	...	...	...	...	...	...

Tableau VII.4 : Interfaces dans le tetramère de tetramère de *lanYFP2* à pH physiologique [Krissinel and Henrick, 2007]



## G Considérations sur l'exploitation des cartes de différence Fourier

L'étude des cartes de différence Fourier a été menée en étudiant l'emplacement et l'intensité des pics de densité de chaque carte. La lecture de l'intensité des pics est menée avec le logiciel de visualisation de structure et de carte de densité COOT de deux façon : l'une, relative, en multiples du niveau de  $\sigma$  de la carte ( $\sigma$  étant la moyenne quadratique de la densité calculée sur tout le volume de la maille) ; l'autre, absolue, en fractions d'électrons par unité de volume. La première est très utilisée dans la littérature, mais si l'on observe la figure VII.3.a, on s'aperçoit que pour un phénomène à priori monotone (l'endommagement croissant d'une protéine) on passe par un maximum de valeurs des pics exprimés en  $\sigma$ . J'ai remarqué que si l'on utilisait la seconde représentation, le phénomène monotone était bien représenté par des courbes monotones. Cela s'explique par le fait qu'au fur et à mesure de l'endommagement aux rayons X, les jeux de données sont de moins en moins isomorphes avec le premier de la série et donc que le niveau de  $\sigma$ , équivalent à un niveau de bruit, augmente. La pertinence de la seconde représentation provient du fait que chaque jeu est mis à l'échelle par rapport au premier jeu, et donc que le nombre d'électrons dans la maille est considéré comme constant au cours du temps et égal à  $F_{000}$  du premier jeu.

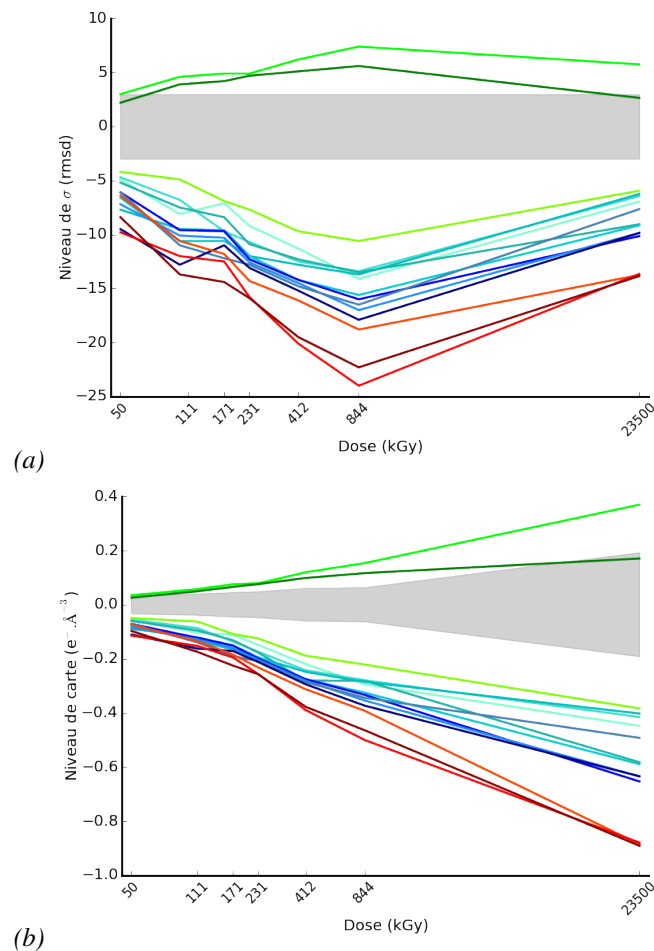


FIGURE VII.3 : Intensité des pics de densité provenant des cartes de différence Fourier en fonction (a) du niveau de  $\sigma$  et (b) de la fraction d'électrons par unité de volume ( $e^- \cdot \text{\AA}^{-3}$ )

## **H Article *Acta Crystallographica section D: Structural Biology***

### Analyse structurale de la protéine fluorescente jaune-vert monomérique et brillante mNeonGreen obtenue par évolution dirigée

Damien Clavel, Guillaume Gotthard, David von Stetten, Daniele de Sanctis, Hélène Pasquier, Gerard G. Lambert, Nathan C. Shaner et Antoine Royant

**Résumé** : Les gènes codant pour des PF sont généralement identifiés chez des animaux marins issus des embranchements Cnidaire et Arthropode. Récemment, de nouveaux gènes ont été identifiés dans l'embranchement des Chordés, notamment une protéine fluorescente jaune, oligomérique et particulièrement brillante, *lanYFP* du vers marin *Branchiostoma lanceolatum*. Une étape d'évolution dirigée de *lanYFP* a permis le développement d'une protéine fluorescente monomérique jaune-vert appelée mNeonGreen avec une brillance élevée. La décoloration remarquable des cristaux de mNeonGreen après la collecte de données de diffraction aux rayons X, nous a conduit à étudier les effets spécifiques des dommages aux radiations sur la cavité du chromophore par la combinaison des techniques de cristallographie aux rayons X, spectroscopie UV-visible et Raman. Nous avons pu montrer que les rayons X conduisent rapidement à la protonation de l'oxygène du phénolate du chromophore et à la rupture de la planarité du pont méthylène.

**Mots clefs** : protéines fluorescentes, *Branchiostoma lanceolatum*, dommages spécifiques aux radiations, spectroscopie optique *in crystallo*, spectroscopie Raman sur ligne de lumière.



# Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution

Damien Clavel,<sup>a,b,c,d,‡</sup> Guillaume Gotthard,<sup>e,‡</sup> David von Stetten,<sup>e</sup> Daniele De Sanctis,<sup>e</sup> Hélène Pasquier,<sup>d</sup> Gerard G. Lambert,<sup>f</sup> Nathan C. Shaner<sup>f</sup> and Antoine Royant<sup>a,b,c,e,\*</sup>

Received 23 September 2016

Accepted 21 November 2016

Edited by M. Czjzek, Station Biologique de Roscoff, France

‡ These authors contributed equally to this work.

**Keywords:** fluorescent protein; *Branchiostoma lanceolatum*; specific radiation damage; *in crystallo* optical spectroscopy; online Raman spectroscopy.

**PDB references:** mNeonGreen, pH 4.5, 5ltq; mNeonGreen, pH 8.0, 5ltr; lanYFP, pH 7.5, 5ltq

**Supporting information:** this article has supporting information at journals.iucr.org/d

<sup>a</sup>Université Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38044 Grenoble, France, <sup>b</sup>CNRS, IBS, F-38044 Grenoble, France, <sup>c</sup>CEA, IBS, F-38044 Grenoble, France, <sup>d</sup>CNRS, Laboratoire de Chimie Physique, UMR 8000 and Université Paris-Sud, F-91405 Orsay, France, <sup>e</sup>European Synchrotron Radiation Facility, F-38043 Grenoble, France, and <sup>f</sup>Department of Photobiology and Bioimaging, The Scintillon Institute, San Diego, California, USA. \*Correspondence e-mail: antoine.royant@ibs.fr

Until recently, genes coding for homologues of the autofluorescent protein GFP had only been identified in marine organisms from the phyla Cnidaria and Arthropoda. New fluorescent-protein genes have now been found in the phylum Chordata, coding for particularly bright oligomeric fluorescent proteins such as the tetrameric yellow fluorescent protein lanYFP from *Branchiostoma lanceolatum*. A successful monomerization attempt led to the development of the bright yellow-green fluorescent protein mNeonGreen. The structures of lanYFP and mNeonGreen have been determined and compared in order to rationalize the directed evolution process leading from a bright, tetrameric to a still bright, monomeric fluorescent protein. An unusual discolouration of crystals of mNeonGreen was observed after X-ray data collection, which was investigated using a combination of X-ray crystallography and UV–visible absorption and Raman spectroscopies, revealing the effects of specific radiation damage in the chromophore cavity. It is shown that X-rays rapidly lead to the protonation of the phenolate O atom of the chromophore and to the loss of its planarity at the methylene bridge.

## 1. Introduction

The discovery and study of green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* led researchers to give it a prime role in live cell biology as a gene reporter (Tsien, 1998). A visible light-absorbing group, the chromophore, autocatalytically forms during protein folding from three consecutive amino-acid residues (Ser, Tyr and Gly in *Aequorea* GFP) in the core of the protein. The chromophore is fluorescent as a result of the favourable environment provided by the rigid 11-strand  $\beta$ -barrel structure of the protein.

Fluorescent proteins (FPs) homologous to GFP were subsequently found in Anthozoa species such as sea anemones and reef-building corals, which belong to the same phylum, Cnidaria, as jellyfish species (Matz *et al.*, 1999). The vast majority of these homologues form oligomers, with the tetramer as the preferred arrangement in Anthozoa (Alieva *et al.*, 2008). Tetrameric FPs are classically described as four protein chains, *A*, *B*, *C* and *D*, forming two pairs of interaction interfaces: a hydrophilic one between monomers *A* and *B* or *C* and *D* ('*A–B* interface'), and a hydrophobic one between monomers *A* and *C* or *B* and *D* ('*A–C* interface'). As oligomer-forming gene reporters would result in obvious artefacts in cell-imaging experiments, it is of paramount



importance to engineer monomeric variants of wild-type fluorescent proteins that retain the original fluorescence properties (Kremers *et al.*, 2011).

Recently, homologues of GFP have been identified in the marine organism lancelet, which belongs to the distinct phylum Chordata (Deheyn *et al.*, 2007). Some of these fluorescent proteins display very high values of fluorescence quantum efficiency (quantum yield; QY) and molar extinction coefficient (EC), resulting in particularly high levels of brightness. Notably, nearly all chordate fluorescent proteins have a GYG chromophore tripeptide sequence. Of the 21 GFP-like genes present in the lancelet *Branchiostoma lanceolatum*, the yellow fluorescent protein *bIFP-Y3* (GenBank accession No. EU482389) attracted particular interest owing to its near-perfect QY (>0.95) and very high EC ( $150\,000\text{ M}^{-1}\text{ cm}^{-1}$ ), and could be successfully evolved into the monomeric green fluorescent protein mNeonGreen, which retains 65% of the brightness of the parent protein (Shaner *et al.*, 2013). Monomerization of *lanYFP* was achieved in a series of rounds of mutagenesis based on rational design and directed evolution. The rational design approach was used to model *lanYFP* based on the known structure of TurboGFP, a solubility-improved variant of the closest primary-sequence homologue pfluGFP2 from *Pontellina plumata* (Evdokimov *et al.*, 2006). In an initial setup to ensure optimal folding and high expression yield, the first and last seven residues of enhanced green fluorescent protein (EGFP) were added to the sequence, with a four-residue linker before the N-terminus of the original protein. The next step consisted of breaking the A–B interface with the I118K mutation, which negatively affected the fluorescence properties. Addition of the folding mutation N174D, and introduction of the additional mutations A45D, S165N and V171A after several rounds of directed evolution, resulted in a dimeric FP with only slightly decreased QY and EC values (0.90 and  $125\,000\text{ M}^{-1}\text{ cm}^{-1}$ , respectively), which was dubbed *dlanYFP*. Finally, 16 mutations were introduced in a series of alternate A–C interface-perturbation and fluorescence-restoration rounds: V140R, I144T, D156K, T158S, Q168R and F192Y at the A–C interface and F151, R25Q, Q56H, F67H, K79V, S100V, F115A, T141S, M143K and I185Y elsewhere. The resulting FP, mNeonGreen, exhibits a small but significant blue shift of its excitation and emission maxima (hence the ‘green’ colour) and, once more, slightly decreased QY and EC values (0.80 and  $116\,000\text{ M}^{-1}\text{ cm}^{-1}$ , respectively). In brief, mNeonGreen is a monomeric version of *lanYFP*, at the expense of a 35% loss in fluorescence, but with a particularly fast maturation speed, a probable gain resulting from the evolution process. Determining and comparing the structures of these proteins should help to pinpoint the various roles of the introduced mutations.

Determining the structure of proteins using high X-ray fluxes at synchrotron beamlines inevitably poses the question of the integrity of the structure, in particular at the active-site location (Weik *et al.*, 2000; Ravelli & McSweeney, 2000; Burmeister, 2000). The so-called ‘specific’ radiation damage notably affects the chromophore site in coloured proteins such as bacteriorhodopsin (Matsui *et al.*, 2002) or photoactive

yellow protein (Kort *et al.*, 2004), but also that of FPs such as EGFP (Royant & Noirclerc-Savoye, 2011) or IrisFP (Adam *et al.*, 2009). The observed changes include the decarboxylation of a conserved glutamate residue appearing on the dose scale of a classical X-ray data collection ( $\sim 1\text{ MGy}$ ) and the loss of electron conjugation on the chromophore, which appears well below the classical dose scale ( $\sim 0.1\text{ MGy}$ ). These results strongly suggest that a radiation-damage study should follow the structure elucidation of any new FP.

We have crystallized *lanYFP* at near-physiological pH and mNeonGreen at both acidic and near-physiological pH values, and obtained their structures at 2.05, 1.70 and 1.21 Å resolution, respectively. We compared the various structures in order to understand the evolution process of the bright variant mNeonGreen and the structural reasons for the loss of fluorescence at acidic pH. Finally, we performed a radiation-damage study of mNeonGreen using complementary optical spectroscopy.

## 2. Methods

### 2.1. Protein expression and purification

Genes coding for *lanYFP* and mNeonGreen were cloned into a pNCST plasmid, a derivative of the constitutive expression plasmid described in Shaner *et al.* (2013). Over-expression of each protein in *Escherichia coli* was performed by growing bacterial cultures at 37°C for 14 h in 2×YT microbial medium. Bacterial cells were pelleted by centrifugation at 5000g and 4°C for 20 min and resuspended in a lysis buffer composed of 20 mM Tris–HCl pH 8.0, 250 mM NaCl, 20 mM imidazole. Cells were disrupted with a high-pressure TS Series cell disruptor (Constant Systems Ltd, Daventry, England) at 7°C and 95 MPa. Cell debris was removed by centrifugation at 43 000g and 4°C for 30 min. The clarified lysate was purified by the successive use of nickel-affinity (HisTrap HP 5 ml, GE Healthcare, Little Chalfont, England) and size-exclusion chromatography columns (Superdex 200 10/300 GL, GE Healthcare). The purified protein was concentrated to 20 mg ml<sup>-1</sup> in 20 mM Tris–HCl pH 8.0. Prior to crystallization, the purified proteins were submitted to tryptic digestion for 1 h at room temperature at a 1:10 trypsin:protein ratio.

### 2.2. Protein crystallization

All crystals were obtained using the vapour-diffusion technique at 20°C. *lanYFP* was crystallized in 0.19 mM CYMAL-7, 100 mM HEPES pH 7.5, 40% (v/v) PEG 400. The acidic form of mNeonGreen was crystallized as long needles in 6.8 mM CYMAL-7, 100 mM sodium citrate tribasic dihydrate pH 4.5, 14% PEG 20 000. The near-physiological pH form of mNeonGreen was crystallized as bipyramids in 100 mM HEPES pH 8.0, 20% PEG 8000.

### 2.3. X-ray data collection and reduction

Only the acidic crystalline form of mNeonGreen required cryoprotection before flash-cooling, which was performed by

**Table 1**  
Data collection and processing.

Values in parentheses are for the outer shell.

Protein	<i>lan</i> YFP (pH 7.5)	mNeonGreen (pH 4.5)	mNeonGreen (pH 8.0)
PDB code	5ltq	5ltp	5ltr
Data collection			
Wavelength (Å)	0.976	0.976	0.954
Temperature (K)	100	100	100
pH	7.5	4.5	8.0
Space group	<i>P</i> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 6 <sub>5</sub> 22
Unit-cell parameters			
<i>a</i> (Å)	78.7	75.9	72.3
<i>b</i> (Å)	197.2	127.6	72.3
<i>c</i> (Å)	115.1	146.9	154.0
$\alpha$ (°)	90.0	90.0	90.0
$\beta$ (°)	90.5	90.0	90.0
$\gamma$ (°)	90.0	90.0	120.0
Resolution range† (Å)	49.29–2.05 (2.12–2.05)	46.34–1.70 (1.76–1.70)	48.56–1.21 (1.25–1.21)
Total reflections	835157 (86173)	670077 (65962)	1330789 (131417)
Unique reflections	217268 (21713)	155051 (15390)	72944 (7142)
Multiplicity	3.8 (4.0)	4.3 (4.3)	18.2 (18.4)
Completeness (%)	99.38 (99.55)	98.87 (99.05)	99.86 (99.92)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	13.5 (2.5)	19.4 (1.9)	18.8 (2.1)
Wilson <i>B</i> factor (Å <sup>2</sup> )	28.5	22.4	12.9
<i>R</i> <sub>meas</sub> ‡	0.088 (0.659)	0.054 (0.844)	0.088 (1.705)
<i>CC</i> <sub>1/2</sub>	0.998 (0.730)	1.000 (0.678)	1.000 (0.729)
Structure refinement			
No. of molecules in the asymmetric unit	16	6	1
<i>R</i> <sub>work</sub>	0.216 (0.296)	0.181 (0.275)	0.158 (0.316)
<i>R</i> <sub>free</sub>	0.244 (0.321)	0.205 (0.301)	0.184 (0.346)
No. of atoms			
Protein	27491	10879	1876
Chromophore	304	114	19
Water	553	911	230
<i>B</i> factors (Å <sup>2</sup> )			
Protein	38.4	27.5	15.1
Chromophore	34.2	18.8	9.5
Water	27.8	32.5	25.8
R.m.s. deviations			
Bond lengths (Å)	0.006	0.008	0.009
Bond angles (°)	1.29	1.50	1.59

† The resolution cutoff is based on *CC*<sub>1/2</sub>. ‡ *R*<sub>meas</sub> = *R*<sub>merge</sub> × [*N*(*N* – 1)]<sup>1/2</sup>, where *N* is the data multiplicity.

the addition of 20% (v/v) glycerol. All data sets were collected on beamline ID29 at the ESRF (de Sanctis *et al.*, 2012). The X-ray damage study on mNeonGreen was also performed on ID29. X-ray diffraction intensities were integrated, scaled and merged using the *XDS* package (Kabsch, 2010). Data-collection and reduction statistics are shown in Table 1.

#### 2.4. Structure phasing and refinement

The structure of acidic mNeonGreen was solved by the molecular-replacement method with *Phaser* (McCoy *et al.*, 2007) using the structure of *lan*RFP, which has 69% sequence identity to mNeonGreen, as a search model (PDB entry 4jf9; Pletnev *et al.*, 2013). The structures of *lan*YFP and near-physiological pH mNeonGreen were both solved by molecular replacement in *Phaser* using the structure of acidic mNeonGreen. Models were manually built in *Coot* (Emsley *et al.*, 2010) and refined with *REFMAC5* (Murshudov *et al.*, 2011). Structure-refinement statistics are shown in Table 1. The structures of *lan*YFP, acidic mNeonGreen and near-physiological pH mNeonGreen have been deposited in the Protein Data Bank with PDB codes 5ltq, 5ltp and 5ltr, respectively.

#### 2.5. *In crystallo* UV–visible absorption spectroscopy

UV–Vis absorption spectra were recorded at the ID29S-Cryobench laboratory of the ESRF (von Stetten *et al.*, 2015) using a high-sensitivity fixed-grating QE65Pro spectrophotometer with a back-thinned CCD detector (Ocean Optics, Dunedin, Florida, USA) and a balanced deuterium–halogen DH2000-BAL light source (Ocean Optics). Spectra were averaged from ten 400 ms acquisitions from ~50 μm thick crystals flash-cooled at 100 K.

#### 2.6. Online Raman spectroscopy

Online Raman spectroscopy was performed on beamline ID29 as described previously (Bui *et al.*, 2014) using a setup specifically designed for the collection of X-ray and Raman data in an interleaved manner (von Stetten *et al.*, unpublished work). In brief, Raman spectra were recorded using an inVia Raman instrument (Renishaw PLC, Wotton-under-Edge, England) equipped with a near-infrared (785 nm) 300 mW diode laser source. Raman spectra were measured from the X-ray-exposed region of a static mNeonGreen crystal with a composite acquisition time of 10 × 10 s for the 300–1800 cm<sup>-1</sup> spectral window. Spectra were corrected for background using the *WiRE* software v.3.4 (Renishaw PLC). X-ray burn

cycles were performed in between Raman data sets, but no diffraction data were recorded.

### 3. Results and discussion

#### 3.1. Structures of *lan*YFP and mNeonGreen at near-physiological pH

We solved the structure of *lan*YFP, modified with the two folding mutations V171A and N174T, at pH 7.5 at 2.05 Å resolution and that of mNeonGreen at pH 8.0 at 1.21 Å resolution. The asymmetric unit of the *lan*YFP crystals contains a 16-mer, which can be described as a nonphysiological tetramer of the physiological tetramer (Fig. 1*a*) after comparison of the various interaction surface areas. The eight *A–B*-type interfaces have an area ranging from 843 to 872 Å<sup>2</sup> and the *A–C*-type interfaces have an area ranging from 1020 to 1133 Å<sup>2</sup>, with one outlier at 779 Å in the most agitated tetramer, while the maximal area value for other interfaces is only 346 Å<sup>2</sup>, strongly suggesting that they correspond to crystal contacts. As an illustration of the monomeric state of mNeonGreen (yet not as a proof), the asymmetric unit of

mNeonGreen crystals obtained at pH 8.0 contains only one monomer (Fig. 1c), and the two largest interaction surface areas with symmetry-related monomers are 562 and 742 Å<sup>2</sup>. A comparison of the structures highlights how *lanYFP* was successfully mutated into the monomeric yet still bright mNeonGreen.

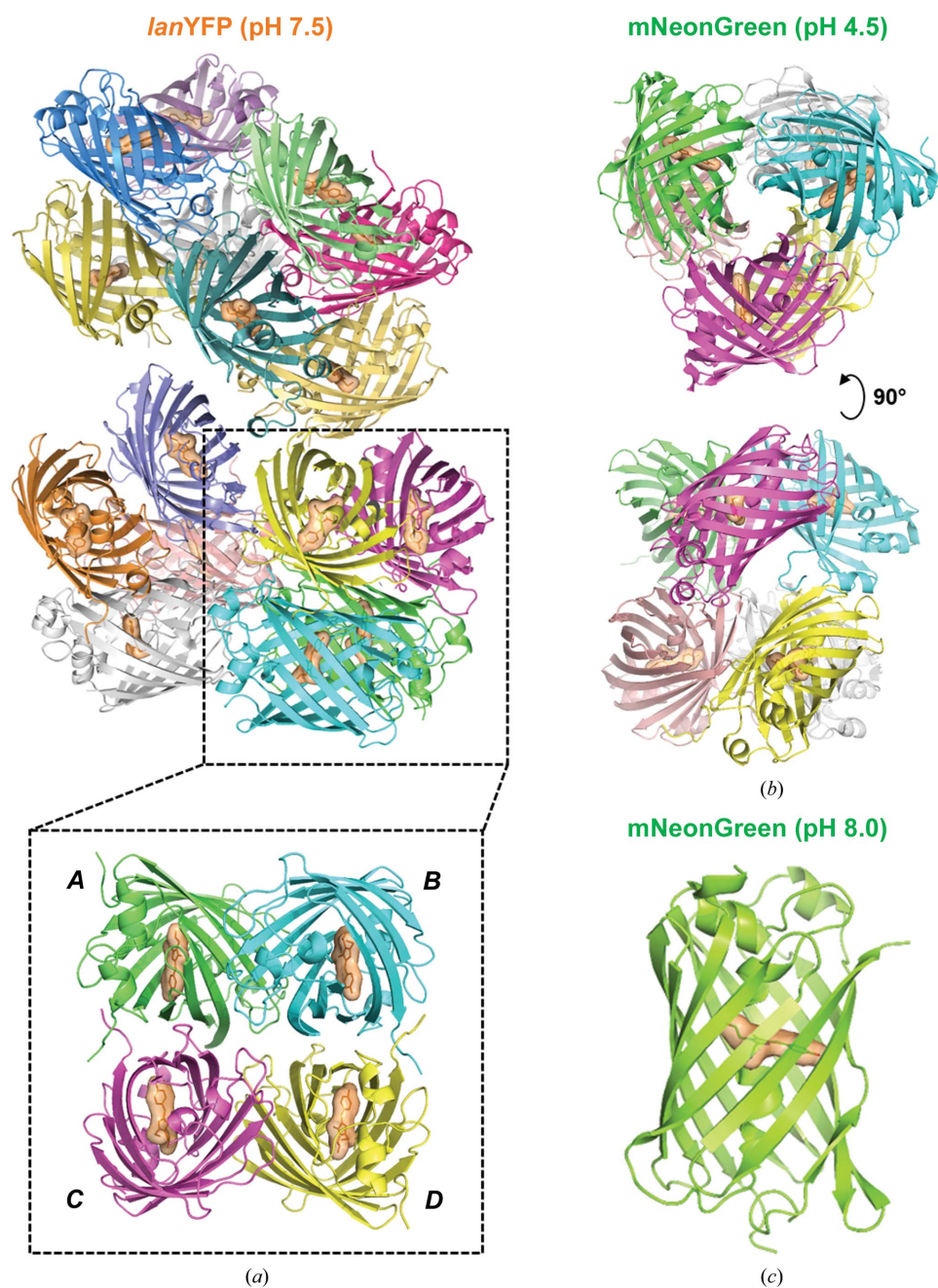
**3.1.1. Mutations at the oligomer-interaction interfaces.** Analysis of surface interactions was performed using the PISA server at <http://www.ebi.ac.uk/pdbe/pisa/> (Krissinel &

Henrick, 2007). The *A–B* interface is composed of residues 14, 16, 17, 19, 85, 87, 89, 91, 95, 97, 99, 101, 103, 116, 118, 120–124, 146, 147, 150, 152, 172, 174 and 176, and the *A–C* interface of residues 137–140, 142, 144, 145, 154, 156, 158, 166, 168, 172, 189, 190, 192, 194–196 and 211–219. Two mutations are located at the *A–B* interface, with N174T essentially being a folding mutation. Ile118 faces His87, Asn101 and Arg103 at the dimer interface, and its replacement by a longer and positively charged lysine is sufficient to disrupt this interface (Fig. 2a).

There are six mutations at the *A–C* interface, but most of the perturbation is ensured only by the V140R mutation, which destabilizes a large hydrophobic patch formed by six phenylalanines and two valines (Fig. 2b). A different mode of interface destabilization was obtained in the D156K and Q168R mutations, which insert positive charges at a location which in *lanYFP* accommodates a negatively charged area in front of a positively charged one.

**3.1.2. Chromophore environment.** Comparison of the chromophore environment in the structures of *lanYFP* and mNeonGreen allows us to propose the structural rationale for the successful propagation of fluorescence properties from one to the other, yet at the price of a 35% loss of overall brightness and of a 7 nm blue shift of both UV–Vis absorption and fluorescence emission maxima, the latter occurring in the evolution step from *dlanYFP* to mNeonGreen. The progressive loss of brightness suggests that the oligomerization state of an FP provides some of the rigidity necessary for efficient chromophore fluorescence within the protein scaffold.

There are two significant changes in the immediate vicinity of the chromophore of mNeonGreen relative to *lanYFP*, one on each side of the conjugated electron cloud. The first one is the localization of a strong spherical electron density at 4.5–4.8 Å from the phenolate ring of the chromophore in *lanYFP* (Fig. 3a) that we modelled as a chloride ion given the interaction distances



**Figure 1**  
Asymmetric unit in the various crystal forms of *lanYFP* and mNeonGreen. (a) *lanYFP* tetramer of tetramers in crystals of space group  $P2_1$  obtained at pH 7.5. (b) mNeonGreen dimer of trimers in crystals of space group  $P2_12_12_1$  obtained at pH 4.5. (c) mNeonGreen monomer in crystals of space group  $P6_32_2$  obtained at pH 8.0.

with the six closest neighbouring groups (average distances of 2.9, 2.9, 3.3, 3.4, 3.6 and 4.0 Å for the most resolved tetramer), the fact that one of these groups is a positively charged arginine and the fact that the *B* factor of the refined chloride compares well with that of the average *B* factor of interacting atoms (32.8 versus 24.0 Å<sup>2</sup>; Carugo, 2014). Given that there is no chloride in the mother-liquor composition, this suggests that *lanYFP* has a very strong affinity for chloride ions, a

property that is also observed in the YFP variant of *A. victoria* GFP (Wachter & Remington, 1999). The presence of chloride at this location is greatly decreased in mNeonGreen thanks to the unexpected carboxylation of Lys143, which introduces a negative charge preventing the binding of any anion nearby (Fig. 3*b*). The carboxylation of lysines is a spontaneous post-translational modification proceeding from dissolved carbon dioxide at basic pH, and is believed to be present in up to 1.3%

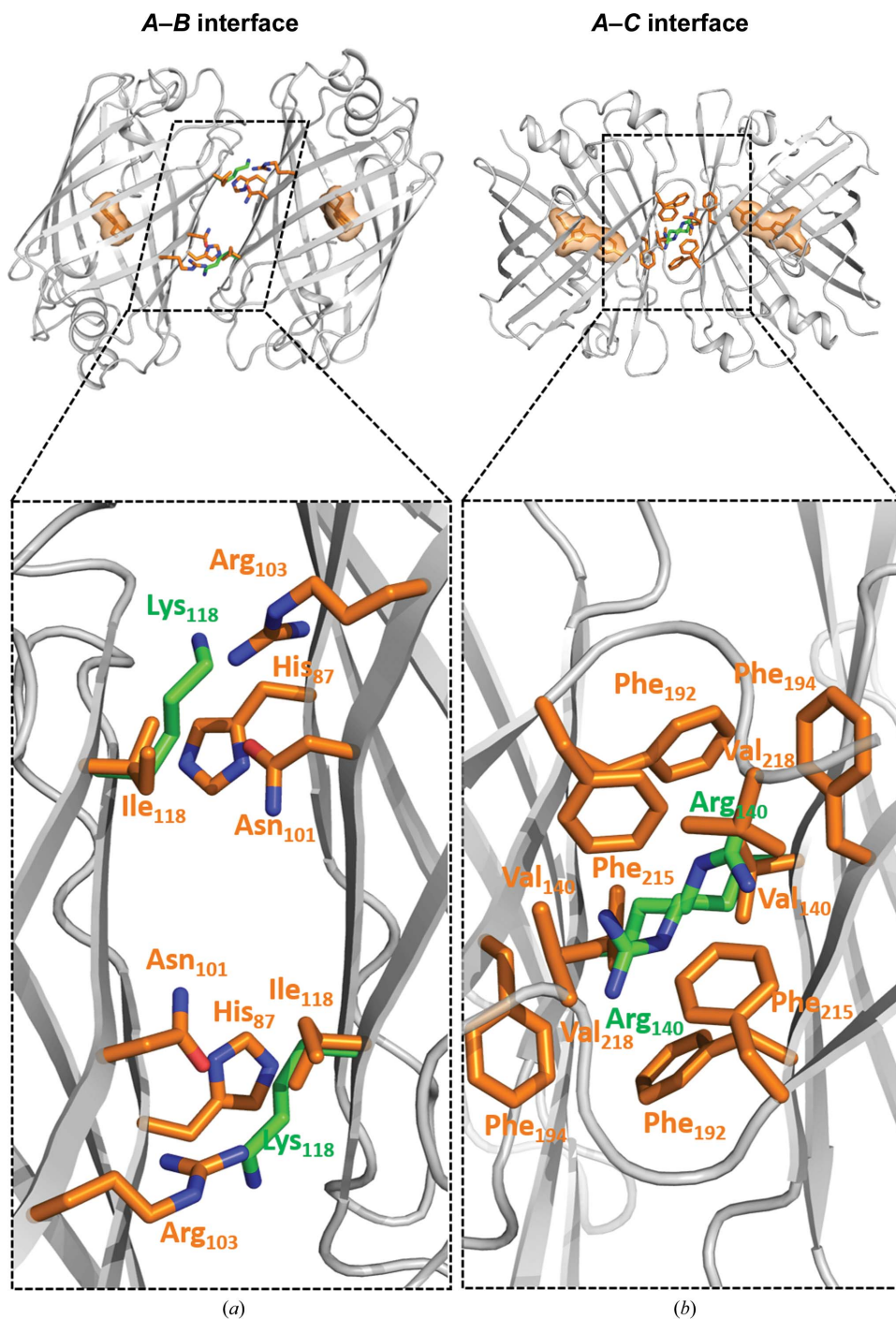


Figure 2  
(a) A–B and (b) A–C interfaces in *lanYFP* (orange side chains) featuring two key mutations ensuring disruption of the interactions in mNeonGreen (green side chains).

of large proteins (Jimenez-Morales *et al.*, 2014). While it would be tempting to attribute a role to this totally unpredictable modification near the chromophore, it is unlikely to significantly control the main fluorescence properties of mNeonGreen, since the M143K mutation was inserted at a late stage in the evolution process, in particular after the blue shift of the fluorescence excitation and emission properties of the protein had occurred. However, this serendipitous modification seems to have greatly reduced the affinity of mNeonGreen for chloride when compared with *lanYFP*.

The most significant structural change is found on the other side of the chromophore (Fig. 3*c*). In *lanYFP*, Pro55, the residue just one residue before the chromophore in the protein sequence, is engaged in a symmetric (red dashed lines in Fig. 3*c*) lone pair– $\pi$  interaction with all five atoms of the imidazolinone ring of the chromophore *via* its carbonyl O atom. In mNeonGreen, the equivalent carbonyl group is translated to the side of the imidazolinone ring by 1.2 Å, resulting in an asymmetric interaction (cyan dashed lines in Fig. 3*c*). The sideways change in polar interaction with the delocalized electron cloud must affect the absorption and emission transition dipole moments of the chromophore, effectively resulting in a shift of both the absorption and emission maxima. The origin of this displacement is to be found in a set of three neighbouring mutations that were introduced in order to restore the fluorescence of monomerized

*dlanYFP*: Q156H, S100V and F115A. In particular, these last two mutations allow the rotation of the side chain of Tyr102, which removes steric hindrance and leads to the formation of a very short hydrogen bond (2.6 Å) between its phenolate O atom and the carbonyl group of His56, the main and side chains of which are displaced (Fig. 3c). The displacement of Pro55 is made all the more visible by the fact that Trp157, with which it is interacting in a methyl- $\pi$  interaction, is pushed away along the same direction (Fig. 3c). Most other mutations introduced to restore fluorescence after the first or the second step of interface disruption are located on loops that are distant from the chromophore, and must be considered as folding mutations at this stage.

### 3.2. Structure of mNeonGreen at acidic pH

We managed to crystallize mNeonGreen at acidic pH (pH 4.5), resulting in crystals with a distinct morphology from those grown at pH 8.0 (Fig. 4a). We were able to solve the structure of the acidic form of mNeonGreen at 1.70 Å resolution, and indeed the space group of this crystal form was

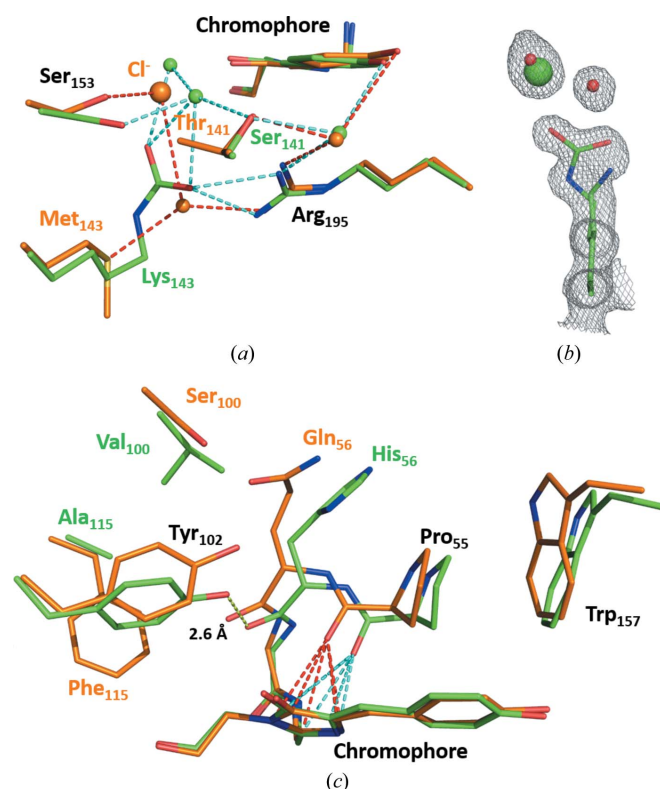


Figure 3

Comparison of the environment of the chromophore in *lanYFP* (orange) and in mNeonGreen (green) at pH 8.0. (a) Close-up of the chloride-binding site in *lanYFP* and the carboxylated lysine in mNeonGreen at physiological pH. (b)  $\sigma_A$ -weighted  $2F_o - F_c$  electron-density map contoured at a  $1.0\sigma$  level around Lys143 in mNeonGreen, the major conformation of which is carboxylated and the minor conformation of which allows the binding of a chloride ion. (c) Close-up of the differences located on the other side of the chromophore. The strong hydrogen bond between Tyr102 and the carbonyl group of His56 in mNeonGreen is represented as a yellow dashed line. The lone pair- $\pi$  interaction between the carbonyl group of Pro55 and the imidazolinone ring of the chromophore is represented as red dashed lines in *lanYFP* and cyan dashed lines in mNeonGreen.

distinct from that we obtained at pH 8.0 ( $P2_12_12_1$  versus  $P6_522$ ). The asymmetric unit contains a hexamer, which can be described as two trimers (Fig. 1b), which are most likely to be nonphysiological given the two maximal interaction areas of 682 and 727 Å<sup>2</sup> within the trimer, which are well below the 840–1140 Å<sup>2</sup> seen in *lanYFP* tetramers. Using *in crystallo*

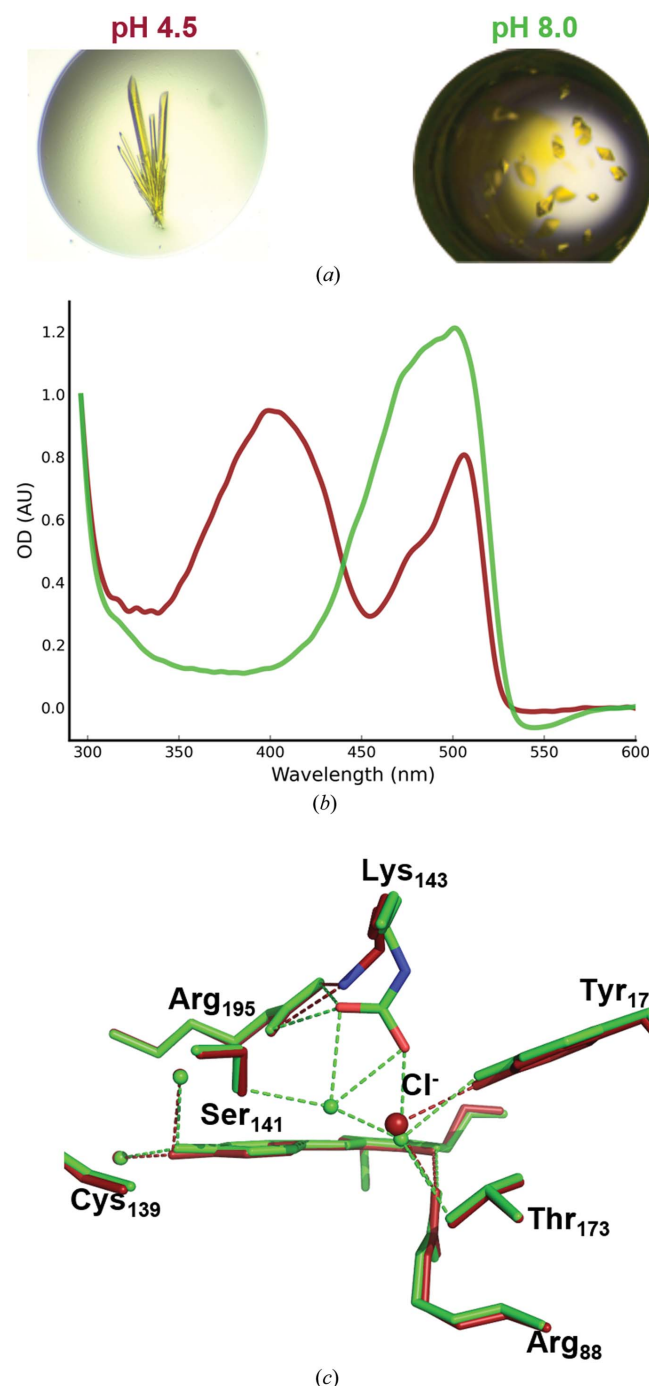


Figure 4

Spectroscopic and structural comparison of mNeonGreen at acidic and near-physiological pH values. (a) mNeonGreen crystal morphologies obtained at pH 4.5 (top) and pH 8.0 (bottom). (b) UV-visible absorption spectra of mNeonGreen crystals at pH 4.5 (dark red) and pH 8.0 (green). (c) Superposition of the chromophore environment in mNeonGreen at pH 4.5 (dark red) and pH 8.0 (green).



UV–visible absorption spectroscopy, we have verified that acidification of the mother liquor leads to an increase in the protonated form of the chromophore, which has an absorption maximum at 403 nm, while the deprotonated form corresponds to the 506 nm peak (Fig. 4*b*).

Comparison of the chromophore environment in both acidic and near-physiological pH conditions only reveals a change in the negatively charged carboxylated Lys143 (Fig. 4*c*). The lysine appears unmodified at acidic pH, leaving room for the binding of a chloride ion at 70% occupancy at the same location as seen for *lanYFP* (100% occupation) and for mNeonGreen (minor occupation of 30%) at pH 8.0. This can be explained by the instability of a carboxylated lysine under acidic conditions, which decarboxylates *via* a probable protonation mechanism (Golemi *et al.*, 2001). However, since the  $pK_a$  of mNeonGreen can be measured at 5.6 or 5.7, whether special care is taken to use chloride-deprived buffers

or not (data not shown), this ion is unlikely to play a role in the protonation mechanism of the chromophore, which thus must relate to the change in hydrogen bonding within the chromophore cavity.

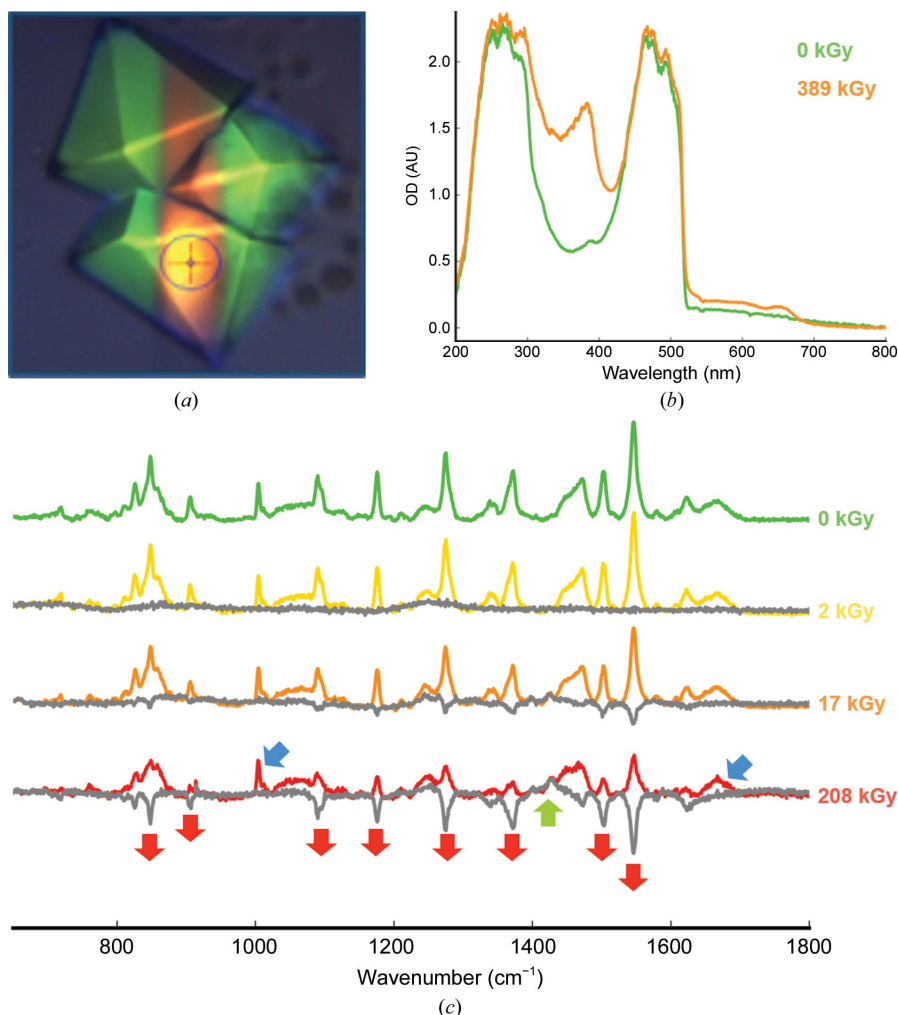
### 3.3. Specific radiation damage in mNeonGreen

While recording our first data set of an mNeonGreen crystal (near-physiological pH form) on beamline ID29 (de Sanctis *et al.*, 2012), we observed on the MD2 video camera that the crystal volume irradiated by the X-ray beam had turned from a bright yellow-green to a dull orange (Fig. 5*a*). In order to understand the structural basis of this colour change, we performed a radiation-damage study correlating optical spectroscopy and X-ray crystallography.

**3.3.1. Optical spectroscopy analysis.** We performed an initial spectroscopic characterization after transporting one

irradiated crystal from the goniometer of beamline ID29 to that of the optical spectroscopy facility ID29S-Cryobench (von Stetten *et al.*, 2015). UV–visible absorption spectra could be measured from a non-irradiated area and an area from which an X-ray data set had been recorded, corresponding to a deposited dose of 389 kGy (Fig. 5*b*). The 0 kGy spectrum shows the expected absorption band between 420 and 520 nm, corresponding to a deprotonated form of the chromophore, yet distorted in shape compared with that in solution, owing to peak saturation resulting from the high optical density of the crystal (~2.0). The 389 kGy spectrum shows the rise of an absorption band centred around 384 nm, which resembles the absorption band of the protonated form of the chromophore observed at acidic pH and thus is likely to be representative of a chromophore that has been protonated upon X-ray irradiation at cryogenic temperature.

In an attempt to further characterize the effects of X-rays on mNeonGreen, we performed an on-line Raman spectroscopy experiment (von Stetten *et al.*, unpublished work) on beamline ID29, in which four Raman spectra were recorded sequentially with increasing doses (Fig. 5*c*). Each spectrum is represented with a difference spectrum corrected for the initial zero-dose spectrum (dark grey). At 2 kGy, the spectrum has not changed. However, at 17 kGy a few strong peaks have started to decrease, and they are significantly smaller at 208 kGy. Examination of the



**Figure 5** X-ray-induced spectroscopic changes of mNeonGreen. (a) Picture of mNeonGreen crystals after a 730 kGy X-ray data collection. (b) UV–visible absorption spectra of the irradiated (orange trace) and non-irradiated (green trace) areas of an mNeonGreen crystal. (c) Series of online Raman spectra measured on an mNeonGreen crystal with increasing X-ray dose. Grey traces correspond to spectra subtracted for the zero-dose spectrum. Blue arrows indicate invariant peaks characteristics of proteins, red arrows indicate decreasing peaks and the green arrow indicates an increasing peak.

**Table 2**  
Principal Raman modes of mNeonGreen.

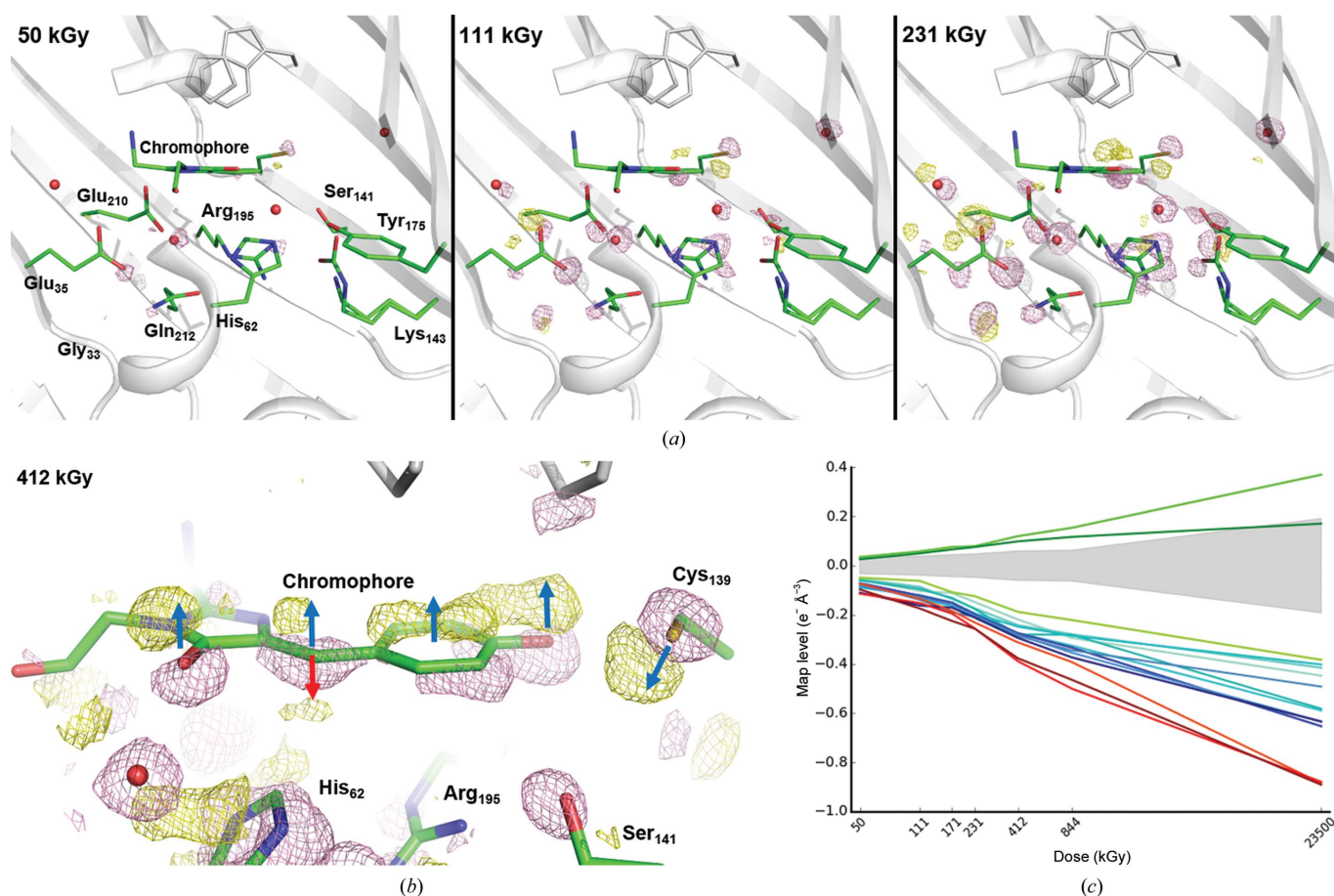
Decreasing (–), increasing (+) or invariant (=) modes with increasing X-ray dose are listed.

Mode description	$\nu$ ( $\text{cm}^{-1}$ )	Change	Reference
Chromophore fingerprint	848	–	Schellenberg <i>et al.</i> (2001)
Chromophore fingerprint	906	–	Schellenberg <i>et al.</i> (2001)
Phenylalanine ring	1005	=	von Stetten <i>et al.</i> (2015)
Chromophore fingerprint	1090	–	Bell <i>et al.</i> (2000)
C–H phenyl bending	1176	–	Bell <i>et al.</i> (2000)
Chromophore fingerprint	1275	–	Bell <i>et al.</i> (2000)
Phenol C–H deformation	1371	–	He <i>et al.</i> (2002)
CH <sub>2</sub> scissor motion	1428	+	Bell <i>et al.</i> (2000)
Phenol 3, C–C stretch of tyrosine ring	1503	–	Schellenberg <i>et al.</i> (2001)
Chromophore exocyclic C=C stretch	1545	–	Schellenberg <i>et al.</i> (2001)
Amide I	1667	=	von Stetten <i>et al.</i> (2015)

usually strong protein stretching modes at  $1005\text{ cm}^{-1}$  (phenylalanine ring band) and  $1667\text{ cm}^{-1}$  (amide I band) (blue arrows in Fig. 5c; von Stetten *et al.*, 2015) shows that the

protein is not affected, and also that most of the strong bands correspond to the pre-resonance Raman signal of the chromophore [even though the excitation wavelength (785 nm) is distant from the absorption maximum of the chromophore at 506 nm]. The eight strongest peaks that decrease upon irradiation (red arrows in Fig. 5c) can be assigned to chromophore modes (Table 2). We have identified one band that increases with dose at  $1428\text{ cm}^{-1}$  (green arrow in Fig. 5c), which we have assigned to a –CH<sub>2</sub> scissoring motion. Together, these results suggest that only the chromophore is affected, that the delocalized  $\pi$ -electron cloud loses resonance with the excitation laser, leading to a decrease in all its modes, and that the only band that increases in intensity could be related to a modification of the methylene bridge.

**3.3.2. Structural analysis.** A series of 23 X-ray data sets were collected one after another from the same crystal, resulting in a ‘dose-lapse’ movie of the specific radiation damage occurring within mNeonGreen. Fourier difference maps ( $n - 1$ ) were calculated between data set  $n$  and data set 1, using the structure factors of each data set and the



**Figure 6**

X-ray-induced structural changes in the chromophore cavity of mNeonGreen. (a) Fourier difference maps  $F_o(n) - F_o(1)$  at increasing X-ray doses contoured at a  $6.0\sigma$  level (pink, negative; yellow, positive). (b) Close-up on the chromophore at a  $4.0\sigma$  level after 412 kGy. Dark blue arrows indicate the upward movement of the whole chromophore and the opposite movement of Cys139. The red arrow indicates the concomitant formation of a kinked configuration of the chromophore upon loss of conjugation at the exocyclic linkage. (c) X-ray-induced electron loss or gain displayed as a function of X-ray dose on a logarithmic scale. Three ensembles of affected groups can be distinguished: damage to Glu35, Glu210 and the water in between is represented in red, damage to the methylene bridge in green and strong damage to other groups listed in Table 3 in blue. The noise level (corresponding to peak heights with a  $\sigma$  level between  $-3.0$  and  $+3.0$ ) at each dose is represented by a light grey shade.

**Table 3**  
X-ray-induced electron loss (or gain), expressed in  $e^- \text{ \AA}^{-3}$ , in mNeonGreen at 100 K.

Dose	50 kGy	111 kGy	171 kGy	231 kGy	412 kGy	844 kGy	23 500 kGy
Fourier difference map	$(F_{o2} - F_{o1})$	$(F_{o3} - F_{o1})$	$(F_{o4} - F_{o1})$	$(F_{o5} - F_{o1})$	$(F_{o8} - F_{o1})$	$(F_{o13} - F_{o1})$	$(F_{o23} - F_{o1})$
1.0 $\sigma$ level	0.011	0.013	0.015	0.016	0.020	0.021	0.064
12 strongest negative peaks							
Glu35	-0.097	-0.174	-0.225	-0.256	-0.377	-0.464	-0.891
Wat(Glu35-Glu210-His62)	-0.114	-0.152	-0.195	-0.255	-0.389	-0.500	-0.879
Glu210	-0.074	-0.134	-0.184	-0.231	-0.312	-0.392	-0.885
Cys139	-0.110	-0.162	-0.171	-0.211	-0.294	-0.373	-0.634
Wat(Thr173-Ser153)	-0.071	-0.139	-0.190	-0.208	-0.286	-0.344	-0.492
His62	-0.084	-0.128	-0.161	-0.203	-0.281	-0.354	-0.634
Gly33	-0.071	-0.122	-0.151	-0.198	-0.275	-0.333	-0.653
Gln212	-0.089	-0.120	-0.150	-0.194	-0.248	-0.285	-0.402
Chromophore (phenolate)	-0.077	-0.134	-0.165	-0.194	-0.275	-0.325	-0.589
Wat(chromophore phenolate-Tyr175)	-0.060	-0.095	-0.131	-0.176	-0.280	-0.281	-0.582
Lys143	-0.055	-0.086	-0.136	-0.173	-0.242	-0.279	-0.415
Chromophore (imidazolinone)	-0.057	-0.103	-0.111	-0.148	-0.219	-0.294	-0.447
Peaks at the methylene-bridge location							
Negative	-0.049	-0.062	-0.108	-0.124	-0.188	-0.221	-0.383
Positive ( $sp^2$ configuration)	0.035	0.058	0.076	0.079	0.120	0.154	0.370
Positive ( $sp^3$ kinked configuration)	0.026	0.050	0.066	0.076	0.099	0.117	0.171

calculated phases of a mNeonGreen structure refined against data set 1. The three maps (2 - 1), (3 - 1) and (5 - 1) are represented in Fig. 6(a), corresponding to accumulated doses of 50, 111 and 231 kGy, respectively. The height of the peaks located on the 12 most-affected groups of the protein (including water molecules) are listed in Table 3 and plotted in Fig. 6(c). The timing of the appearance and the various strengths of the peaks allow us to propose a scenario for the structural description of specific radiation damage occurring to mNeonGreen at early X-ray doses. All peaks are found on one side of the chromophore-binding cavity (Fig. 6a). The initial event appears to be the decarboxylation (negative peak in the maps) of a glutamate residue close to the chromophore, Glu35, which is different from that classically observed to be affected in other UV-damaged or X-ray-damaged fluorescent proteins (van Thor *et al.*, 2002; Royant & Noirclerc-Savoie, 2011; Adam *et al.*, 2009). The next residue affected is actually Glu210, the very residue that was damaged in the previous studies, which appears to be decarboxylated as well, with a negative peak of similar magnitude. However, the number of positive peaks around these two Glu residues suggests that each of the two side chains may reorientate. The most likely scenario is that one of the two is decarboxylated (with a slight preference for Glu35 given that its negative peak is consistently higher in all maps; Table 3), which triggers the reorientation of the other one. The ensuing events are the translation of the whole chromophore in one direction (dark blue arrows) and the movement of Cys139 in the opposite direction, maintaining an interaction with the phenolate O atom (Fig. 6b). These changes are accompanied by the displacement of water molecules and the reorientation of His62, while Arg195 and Lys143 move as well. This results in a drastic change in hydrogen bonding, which leads to the protonation of the chromophore, as observed by UV-Vis absorption spectroscopy (Fig. 5b). However, the loss of resonance is not solely explained by the translation of the whole chromophore and its ensuing protonation. Indeed, further examination of the difference map around the chromophore

shows that there is a second positive peak of similar magnitude next to the exocyclic link (Fig. 6b). This is compatible with the  $sp^2 \rightarrow sp^3$  dehybridization of the central C atom that had been transiently observed during the low-dose X-ray bleaching of the fluorescent protein IrisFP (Adam *et al.*, 2009). It is noteworthy that the peak heights of both positive peaks are similar at low X-ray doses and then progressively differ, with the peak height associated with the dehybridized chromophore falling below the  $\pm 3.0\sigma$  noise level at high doses, in line with the observation that the dehybridization is only transient, thanks to an X-ray-induced reduction/repair mechanism (Adam *et al.*, 2009; Carpentier *et al.*, 2010).

Cumulatively, these observations lead us to propose that X-rays induce fluorescence loss of the chromophore resulting, on one hand, from the translation of the chromophore and its subsequent protonation leading to a blue shift in the absorption maximum which negatively affects the pre-resonance Raman effect and, on the other, from a loss of conjugation at the methylene bridge, separating the two resonant electron clouds of the phenolate and imidazolinone rings and leading to a loss of absorbance at the absorption maximum.

#### 4. Conclusion

Our structural comparison of lanYFP and mNeonGreen confirms the rationale used for the successful evolution process using a combination of rational design and directed evolution. It highlights the key mutations resulting in oligomer interface disruption. It also explains how some of the mutations introduced to help restore fluorescence lead to a concerted rearrangement at the back of the chromophore, resulting in a more constrained environment favouring fluorescence. This rearrangement also causes a displacement of the key interaction between a proline residue and the imidazolinone ring of the chromophore, which must explain the small but significant blue shift of the fluorescence excitation and emission maxima of mNeonGreen when compared with lanYFP.

This correlative X-ray crystallography, UV–Vis absorption and Raman spectroscopy radiation-damage study leads us to interpret the colour change of mNeonGreen crystals along the path of the X-ray beam as an X-ray-induced loss of fluorescence. Compared with other fluorescent proteins, mNeonGreen presents the particularity of having two decarboxylated glutamate residues next to the chromophore instead of one. The colour change originates from the subsequent protonation of the chromophore phenolate group, which leads to a large blue shift of the absorption maximum into the UV region. It also results from the concomitant loss of conjugation of the chromophore at the methylene bridge, effectively separating the two resonant electron clouds of the phenolate and imidazolinone rings, which also leads to a loss of absorbance in the visible, but in this case without the appearance of a distinct absorption band.

### Acknowledgements

DC is the recipient of a PhD fellowship from the University Paris-Sud. We thank Céline Lafaye and Marjolaine Noirclerc-Savoie for contributions at an early stage of the project. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award No. R01GM109984 and by the French National Research Agency under award No. ANR-11-JSV5-0009-01. The ESRF is acknowledged for access to beamlines and facilities for molecular biology *via* its in-house research programme.

### References

- Adam, V., Carpentier, P., Violot, S., Lelimousin, M., Darnault, C., Nienhaus, G. U. & Bourgeois, D. (2009). *J. Am. Chem. Soc.* **131**, 18063–18065.
- Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., Miller, D. J., Wiedenmann, J., Salih, A. & Matz, M. V. (2008). *PLoS One*, **3**, e2680.
- Bell, A. F., He, X., Wachter, R. M. & Tonge, P. J. (2000). *Biochemistry*, **39**, 4423–4431.
- Bui, S., von Stetten, D., Jambrina, P. G., Prangé, T., Colloc'h, N., de Sanctis, D., Royant, A., Rosta, E. & Steiner, R. A. (2014). *Angew. Chem. Int. Ed. Engl.* **53**, 13710–13714.
- Burmeister, W. P. (2000). *Acta Cryst.* **D56**, 328–341.
- Carpentier, P., Royant, A., Weik, M. & Bourgeois, D. (2010). *Structure*, **18**, 1410–1419.
- Carugo, O. (2014). *BMC Struct. Biol.* **14**, 19.
- Deheyne, D. D., Kubokawa, K., McCarthy, J. K., Murakami, A., Porrachia, M., Rouse, G. W. & Holland, N. D. (2007). *Biol. Bull.* **213**, 95–100.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkoporov, A. N., Sander, I., Lukyanov, K. A. & Chudakov, D. M. (2006). *EMBO Rep.* **7**, 1006–1012.
- Golemi, D., Maveyraud, L., Vakulenko, S., Samama, J.-P. & Mobashery, S. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 14280–14285.
- He, X., Bell, A. F. & Tonge, P. J. (2002). *J. Phys. Chem. B*, **106**, 6056–6066.
- Jimenez-Morales, D., Adamian, L., Shi, D. & Liang, J. (2014). *Acta Cryst.* **D70**, 48–57.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kort, R., Hellingwerf, K. J. & Ravelli, R. B. G. (2004). *J. Biol. Chem.* **279**, 26417–26424.
- Kremers, G., Gilbert, S. G., Cranfill, P. J., Davidson, M. W. & Piston, D. W. (2011). *J. Cell Sci.* **124**, 157–160.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Matsui, Y., Sakai, K., Murakami, M., Shiro, Y., Adachi, S.-I., Okumura, H. & Kouyama, T. (2002). *J. Mol. Biol.* **324**, 469–481.
- Matz, M. V., Fradkov, A. F., Labas, Y., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L. & Lukyanov, S. A. (1999). *Nature Biotechnol.* **17**, 969–973.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Pletnev, V. Z., Pletneva, N. V., Lukyanov, K. A., Souslova, E. A., Fradkov, A. F., Chudakov, D. M., Chepurnykh, T., Yampolsky, I. V., Wlodawer, A., Dauter, Z. & Pletnev, S. (2013). *Acta Cryst.* **D69**, 1850–1860.
- Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
- Royant, A. & Noirclerc-Savoie, M. (2011). *J. Struct. Biol.* **174**, 385–390.
- Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
- Schellenberg, P., Johnson, E., Esposito, A. P., Reid, P. J. & Parson, W. W. (2001). *J. Phys. Chem. B*, **105**, 5316–5322.
- Shaner, N. C., Lambert, G. G., Chamma, A., Ni, Y., Cranfill, P. J., Baird, M. A., Sell, B. R., Allen, J. R., Day, R. N., Israelsson, M., Davidson, M. W. & Wang, J. (2013). *Nature Methods*, **10**, 407–409.
- Stetten, D. von, Giraud, T., Carpentier, P., Sever, F., Terrien, M., Dobias, F., Juers, D. H., Flot, D., Mueller-Dieckmann, C., Leonard, G. A., de Sanctis, D. & Royant, A. (2015). *Acta Cryst.* **D71**, 15–26.
- Thor, J. J. van, Gensch, T., Hellingwerf, K. J. & Johnson, L. N. (2002). *Nature Struct. Biol.* **9**, 37–41.
- Tsien, R. Y. (1998). *Annu. Rev. Biochem.* **67**, 509–544.
- Wachter, R. M. & Remington, S. J. (1999). *Curr. Biol.* **9**, R628–R629.
- Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.







**Titre :** Etudes structurales de la dynamique de protéines fluorescentes vertes et jaunes utilisées en imagerie cellulaire

**Mots clés :** Biologie structurale, protéine fluorescente jaune, *Branchiostoma lanceolatum*, chemin de translocation de proton, évolution dirigée.

**Résumé :** Les protéines fluorescentes (PF) homologues d'*AvGFP* (*Green Fluorescent Protein* de la méduse *Aequoria victoria*) sont des outils incontournables de l'imagerie des processus de la cellule vivante. Leurs performances conditionnent la précision de l'analyse quantitative des signaux de fluorescence. Le développement de nouvelles PF demande donc à la fois de parvenir à une forte brillance tout en contrôlant la réponse de la protéine aux variations des paramètres physico-chimiques de la cellule en fonction de la question biologique étudiée. A ce jour, les PF jaunes disponibles montrent une forte sensibilité au pH. Afin d'élaborer des mutants moins sensibles, deux approches ont été considérées : une première consiste à mieux appréhender l'incidence de la dynamique du réseau de liaisons hydrogène entourant le chromophore sur son équilibre acido-basique. La seconde vise à identifier les facteurs structuraux à l'origine de la brillance particulièrement élevée de nouvelles PF jaunes et jaune-vert provenant d'un ver marin, *Branchiostoma lanceolatum*. J'ai d'abord mis au point un algorithme recherchant l'ensemble des liaisons hydrogène présentes au sein d'une protéine et qui étudie leur dynamique au cours de simulations par dynamique moléculaire. Il permet leur agrégation en réseaux, l'identification des réseaux connectés à un atome d'intérêt ainsi que le suivi de leur dynamique. Pour

validation, cet algorithme a été appliqué à la recherche des réseaux de liaisons hydrogène présents au sein de différents mutants d'*AvGFP* pour lesquels un transfert de proton à l'état excité a été étudié expérimentalement. Cet algorithme pourra également servir à comprendre de façon dynamique le mécanisme d'autres systèmes biologiques dont la fonction repose sur le transfert de protons. D'autre part, j'ai résolu la structure de la protéine fluorescente jaune naturelle *lanYFP* de *Branchiostoma lanceolatum*, particulièrement brillante mais à la structure quaternaire tétramérique. Cette protéine a été rendue monomérique par évolution dirigée, ce qui a donné la protéine mNeonGreen à la fluorescence jaune-vert, protéine désormais étalon dans cette gamme spectrale, et dont j'ai également résolu la structure. Mon étude a permis de rationaliser *a posteriori* l'ensemble des mutations introduites au cours de l'évolution. Enfin, j'ai réalisé une étude du dégât d'irradiation spécifique des rayons X permettant de comprendre le changement remarquable de couleur observé sur les cristaux de mNeonGreen après collecte de données de diffraction. L'ensemble des résultats obtenus au cours de ma thèse permet de proposer un cadre de compréhension à la fois théorique et expérimental des déterminants contrôlant les propriétés de fluorescence des PF jaunes.

**Title :** Structural studies of the dynamics of green and yellow fluorescent proteins used in cellular imaging

**Keywords :** Structural biology, yellow fluorescent protein, *Branchiostoma lanceolatum*, proton wire, directed evolution

**Abstract :** Fluorescent Proteins (FPs) homologous to *AvGFP* (Green Fluorescent Protein from the jellyfish *Aequoria victoria*) are versatile tools used in live cell imaging. The amount of information that can be derived from the fluorescence signals depends on the spectroscopic performances of the FP. The development of new FPs should focus on both brightness increase and control of the protein response to physicochemical parameter variations within the cell. Current yellow FPs exhibit a strong sensitivity to pH. In order to engineer less sensitive variants, two complementary approaches have been used: the first one consists in studying the influence of the hydrogen bond network dynamics around the chromophore on its protonation state. In the second one, I have sought to identify the structural determinants of the particularly high brightness of newly discovered yellow FPs from a sea worm, *Branchiostoma lanceolatum*. First, I wrote an algorithm that can identify all hydrogen bonds within a protein and analyse their dynamics along molecular dynamics simulations. It allows for their clustering in networks, the identification of networks connected to a given atom and the

monitoring of their dynamics. The method was validated by using the algorithm on various *AvGFP* mutants for which excited state proton transfer has been experimentally studied. This algorithm should also be useful for the study of other biological systems whose function is based on proton transfer. Besides, I solved the structure of the natural yellow FP *lanYFP* from *Branchiostoma lanceolatum*, which is particularly bright, but presents a tetrameric arrangement. This protein was monomerized by directed evolution, which led to the yellow-green FP mNeonGreen, now a benchmark in this spectral range. I also solved the structure of mNeonGreen, which allowed me to rationalize *a posteriori* the mutations that have been introduced during the evolution process. Finally, I performed a specific radiation damage study in order to explain the remarkable change in colour of mNeonGreen crystals upon X-ray data collection. Altogether, the results of my PhD work provides a theoretical and experimental framework of the determinants that drive the fluorescence properties of yellow FPs.

