



HAL
open science

Traitement des interactions électrostatiques dans les systèmes moléculaires - Etude par simulation numérique de protéines fluorescentes

Mickaël Lelimosin

► **To cite this version:**

Mickaël Lelimosin. Traitement des interactions électrostatiques dans les systèmes moléculaires - Etude par simulation numérique de protéines fluorescentes. Biophysique [physics.bio-ph]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT: . tel-00446463

HAL Id: tel-00446463

<https://theses.hal.science/tel-00446463v1>

Submitted on 12 Jan 2010

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.

THÈSE

présentée par

LELIMOUSIN Mickaël

POUR OBTENIR LE TITRE DE DOCTEUR
DE L'UNIVERSITÉ JOSEPH FOURIER – GRENOBLE I
(Arrêtés ministériels du 5 Juillet 1984 et du 30 Mars 1992)

Spécialité : Chimie physique moléculaire et structurale

**Traitement des interactions électrostatiques
dans les systèmes moléculaires -
Etude par simulation numérique de protéines fluorescentes.**

soutenue le 25 mai 2009 devant le jury composé de :

S. PÉREZ,	Président
I. DEMACHY,	Rapporteur
X. ASSFELD,	Rapporteur
D. PERAHIA,	Examineur
D. BOURGEOIS,	Examineur
M. J. FIELD,	Directeur de thèse

Thèse préparée au Laboratoire de Dynamique Moléculaire,
Institut de Biologie Structurale J.P. Ebel, 41 rue Jules Horowitz 38027 Grenoble Cedex 1.

Je souhaite tout d'abord adresser ma gratitude à toutes les personnes qui ont un jour ou l'autre participé à m'orienter vers ce doctorat. Je pense notamment à certains professeurs d'Université ou certains amis comme Julian, qui m'ont transmis leur passion de la science. Par ailleurs, cette thèse n'aurait jamais pu voir le jour sans les financements du ministère de l'enseignement supérieur et de la recherche et du centre national de la recherche scientifique.

Je remercie Martin Field de m'avoir accepté dans son laboratoire dès le Master 2, et de m'avoir permis de bénéficier de son excellence scientifique tout en m'accordant une grande confiance dans la réalisation de mes travaux.

Je remercie également Dominique Bourgeois d'avoir mis en œuvre avec talent le projet d'étude des protéines fluorescentes. Ce fût un grand plaisir de travailler avec lui et une chance rare d'apprendre tellement auprès de lui-même et des membres de son équipe.

Je remercie Serge Pérez d'avoir toujours été présent pour m'aider depuis le Master 2 et d'avoir accepté d'être président de mon jury de thèse.

Je remercie Isabelle Demachy et Xavier Assfeld pour leur investissement en tant que rapporteurs de mes travaux de thèse. Les discussions que nous avons pu avoir lors de congrès ou lors de la soutenance m'ont beaucoup apporté.

Je remercie David Perahia d'avoir eu la gentillesse d'accepter d'être examinateur de mes travaux de thèse. J'ai beaucoup apprécié l'intérêt qu'il a porté à mon travail et la discussion qui en a résulté le jour de la soutenance.

Je remercie tous les membres passés et présents du Laboratoire de Dynamique Moléculaire. Cette thèse a été avant tout une aventure humaine qui m'a permis de rencontrer des personnes fantastiques. Je remercie les personnes venues des quatre coins de la planète qui m'ont enrichi de leur culture, de leur personnalité, de leurs connaissances et de leur chaleur humaine : merci Smriti, Ignacio, Margherita, Ilaria, Feng, Brian, Laurence, Pasha.

Un merci spécial aux deux personnes qui m'ont courageusement supporté dans ce bureau au cours de ces années. Merci Martin et Grazie Loredana pour l'amitié que nous avons partagée et que nous partagerons encore longtemps, j'en suis sûre.

Je remercie Patricia Amara pour sa profonde générosité, et pour tout ce qu'elle pu m'apporter humainement et professionnellement. Merci de m'avoir permis de réussir mon Master 2 et décrocher cette bourse de thèse, de m'avoir toujours offert de bonnes conditions de travail au laboratoire, de m'avoir formé à tant de choses, de m'avoir soutenu dans les moments difficiles, d'avoir toujours été si irremplaçable ... j'espère sincèrement pouvoir un jour retravailler dans le même couloir que le sien.

Je remercie Aline Thomas pour sa grande gentillesse et le soleil qu'elle a toujours apporté au laboratoire. Je la remercie pour tous ses conseils avisés et le recul scientifique dont elle m'a souvent fait bénéficier. Et je m'excuse auprès de Vincent et Cyril de ne pas avoir eu plus de temps pour venir jouer au foot à Meylan !

Merci à toutes les personnes que j'ai côtoyées avec plaisir à l'institut ; je pense à de nombreuses personnes, dont l'incontournable, le légendaire, le magnifique Georgio Bello Come Stai !

Je remercie chaleureusement l'ex quasi-docteur mais désormais officiel Dr Virgile Adam avec lequel j'ai vécu une aventure peu commune lors de cette thèse en tandem. Ce fut un plaisir et une grande chance de travailler de cette manière lors de ces trois ans et demi de thèse. J'y ai gagné un ami cher et j'ai énormément appris à ses côtés. Je lui souhaite le meilleur pour son post-doctorat et pour tout le reste.

Je remercie Antoine Royant pour cette collaboration très dynamique ! Je le remercie pour tout ce qu'il a fait pour moi et tout ce qu'il m'a apporté lors de la fin de cette thèse. Un grand merci coach !

Pour terminer, je remercie mes proches d'avoir toujours accepté mes choix et de m'avoir soutenu pendant le doctorat. Je remercie tout particulièrement mes parents qui ont toujours été là pour m'aider de toutes les manières que ce soit, et cela de façon exemplaire. Je m'excuse auprès de mes proches d'avoir eu peu de temps à leur consacrer en cette fin de thèse, mais c'était pour une bonne cause ...

Table des matières

Introduction générale	3
Partie A - Traitement des interactions électrostatiques dans les systèmes moléculaires	4
Introduction	7
1 La mécanique quantique	9
1.1 Les fondements	9
1.1.1 La dualité onde-particule	9
1.1.2 L'incertitude d'Heisenberg	10
1.1.3 Les postulats	11
1.2 La chimie quantique	13
1.2.1 Les états stationnaires	13
1.2.2 Le principe variationnel	14
1.2.3 L'hamiltonien moléculaire	14
1.2.4 L'approximation de Born-Oppenheimer	15
1.2.5 La théorie des orbitales moléculaires	16
1.2.6 Le principe d'exclusion de Pauli	18
1.3 La théorie Hartree-Fock	19
1.3.1 L'approximation orbitale	19
1.3.2 L'équation de Hartree-Fock	20
1.3.3 Les équations de Roothaan-Hall	21
1.3.4 Les matrices caractéristiques	22
1.3.5 La transformation symétrique	23
1.3.6 La procédure SCF	24
1.3.7 Le calcul des observables	25

1.3.8	L'énergie de l'état fondamental	26
1.4	La théorie de la fonctionnelle de la densité	27
1.4.1	Le théorème de Kohn-Hohenberg	27
1.4.2	L'équation de Kohn-Sham	27
1.4.3	Les fonctionnelles d'échange-corrélation	28
1.5	L'approche semi-empirique	30
1.5.1	Les principes	30
1.5.2	La matrice densité	31
1.5.3	Les méthodes	31
2	Les charges partielles	33
2.1	Les méthodes de calcul existantes	33
2.1.1	Les analyses de population	33
2.1.2	Les fragmentations de l'espace	35
2.1.3	La reproduction des observables	37
2.1.4	Les méthodes paramétrées	39
2.2	Les méthodes de partition	40
2.2.1	L'utilisation des primitives	41
2.2.2	La quadrature numérique	43
2.2.3	Les schémas de partition	43
2.2.4	L'application aux méthodes <i>ab initio</i> et DFT	47
2.2.5	L'application aux méthodes semi-empiriques	49
2.3	Résultats et discussion	50
2.3.1	Calcul de charges <i>ab initio</i>	50
2.3.2	Calcul de dipôles <i>ab initio</i>	53
2.3.3	Calcul de dipôles DFT	54
2.3.4	Calcul de dipôles semi-empiriques	54
2.3.5	Conclusion des résultats	57
3	Les interactions électrostatiques	59
3.1	La modélisation moléculaire	59
3.1.1	Le système moléculaire	59
3.1.2	La mécanique moléculaire	61
3.1.3	Les potentiels hybrides MQ/MM	64
3.2	L'électrostatique en MQ/MM	66

3.2.1	Le traitement usuel	66
3.2.2	La frontière MQ/MM	66
3.2.3	Le traitement proposé	68
3.2.4	La matrice de Fock	69
3.2.5	Le gradient de l'énergie	71
3.3	Résultats et discussion	73
3.3.1	Calcul d'énergies de formation de dimères	73
Conclusion et perspectives		76
 Partie B - Etude par simulation numérique de protéines fluorescentes		 78
Introduction		81
4	La dynamique structurale des protéines ECFP et Cerulean	83
5	Le mécanisme de photoconversion d'EosFP	127
6	Stabilité thermodynamique des différents états d'IrisFP	175
Conclusion et perspectives		194
 Conclusion générale		 199
 Annexes		 199
A	Informations supplémentaires sur les tests des méthodes de partition	201
B	Base structurale du photoblanchiment d'une protéine fluorescente verte photactivable	217
 Bibliographie		 244

Introduction générale

Depuis plus d'un demi siècle, la biologie subit une révolution de grande ampleur. Les découvertes capitales faites en génétique ont fortement contribué à cette mutation. Elles nous permettent désormais d'explorer la structure tridimensionnelle des protéines, à l'aide principalement des techniques de cristallographie aux rayons X ou de résonance magnétique nucléaire. Cette connaissance permet alors d'appréhender la fonction biologique de la protéine à un niveau atomique. Toutefois, la structure de ces macromolécules est en perpétuelle évolution dans leur conditions physiologiques. Cette dynamique structurale est généralement déterminante pour l'activité de la protéine. Or les techniques expérimentales ne permettent pas toujours de renseigner les processus dynamiques. En cela, la modélisation moléculaire est un complément nécessaire aux approches expérimentales. Ainsi, la simulation des systèmes biologiques permet d'acquérir des informations sur des processus se déroulant à différentes échelles de temps.

La modélisation moléculaire repose directement sur les principes de la physique et de la chimie. Cependant, les équations de la théorie ne peuvent être résolues exactement et des approximations doivent être utilisées dans la description d'un système moléculaire. Il est donc préférable d'avoir en référence des données expérimentales qui permettent de valider les résultats de modélisation. De façon générale, les approximations du modèle doivent être le plus satisfaisantes possible au regard des caractéristiques du système moléculaire (taille, type des atomes, ...) et des propriétés que l'on souhaite calculer. En contrepartie, l'utilisation d'un modèle trop précis peut engendrer des temps de calcul inconcevables de façon courante. Par conséquent, le choix du modèle requiert un compromis entre précision et efficacité. Des efforts importants sont consacrés au développement des techniques de simulation afin d'améliorer conjointement ces deux facteurs. Un élément déterminant est le traitement des interactions électrostatiques prépondérantes dans les systèmes moléculaires. Le travail présenté en première partie de cette thèse propose une nouvelle approche pour le calcul de ces interactions.

De grands progrès ont également été accomplis ces dernières années en biologie cellulaire. Les techniques d'imagerie dans ce domaine ont été révolutionnées par l'utilisation des protéines fluorescentes. Elles ont permis d'améliorer considérablement notre compréhension des processus cellulaires. Le dernier prix Nobel de chimie

a d'ailleurs récompensé les Professeurs Osamu Shimomura, Martin Chalfie et Roger Tsien pour la découverte de la protéine fluorescente GFP (*Green Fluorescent Protein*) et la mise au point de son utilisation comme marqueur. Ces techniques d'imagerie sont en plein essor depuis quinze ans et leur potentiel n'a jamais été aussi prometteur qu'aujourd'hui. Leur développement va de paire avec la modification des protéines fluorescentes pour améliorer leurs propriétés. L'ingénierie de ces marqueurs biologiques nécessite une connaissance fine de leur propriétés physiques et chimiques. En collaboration étroite avec l'équipe de cristallographie cinétique de notre institut, nous avons étudié plusieurs protéines fluorescentes d'intérêt majeur. Ce travail effectué à l'aide de techniques de simulation numérique est présenté dans la seconde partie de cette thèse.

Partie A - Traitement des
interactions électrostatiques dans les
systèmes moléculaires

Introduction

Un système moléculaire est généralement constitué d'un ensemble de molécules et d'ions qui interagissent mutuellement. Parmi ces interactions, celles de type électrostatique sont les plus fortes. Par conséquent, il est fondamental de les considérer avec beaucoup de précautions. Un cas de figure important en biologie est l'interaction des résidus d'une protéine entre eux ou avec des molécules de solvant. Les interactions électrostatiques influencent ainsi la structure globale de la protéine et l'évolution de celle-ci au cours du temps. Ces interactions sont d'ailleurs souvent critiques pour la fonction de la protéine (cf. applications des chap.5 et 6). En particulier, il est connu que les interactions électrostatiques jouent un rôle déterminant dans le pouvoir catalytique des enzymes [1, 2]. De plus, elles subsistent à longue distance contrairement à d'autres types d'interactions plus faibles. Toutefois, ces dernières peuvent également influencer drastiquement les propriétés d'une protéine (cf. chap.4).

Au Laboratoire de Dynamique Moléculaire, nous nous intéressons particulièrement à la modélisation des réactions chimiques dans les protéines. Pour ce faire, il est indispensable de traiter convenablement les interactions électrostatiques entre le site actif d'une protéine et son environnement. Le premier objectif de cette thèse est de proposer un modèle cohérent de ce type d'interactions. L'approche proposée se veut généralisable à différentes méthodes de calcul. Un tel traitement n'existe pas actuellement et il permettrait de répondre à certains problèmes spécifiques de ce domaine. Le projet ambitionne à terme d'augmenter la précision et l'efficacité des méthodes de simulation numérique pour l'étude de la réactivité des systèmes biologiques.

Nous présentons ce travail en trois chapitres progressifs partant des concepts de base pour aboutir aux résultats principaux. Dans le premier chapitre, nous abordons les principes théoriques essentiels à la compréhension de notre travail. Le second chapitre présente l'ensemble des méthodes que nous avons développées pour le calcul de

charges partielles atomiques. Finalement, notre nouveau traitement des interactions électrostatiques dans les systèmes moléculaires est exposé dans le troisième chapitre.

Chapitre 1

La mécanique quantique

L'essentiel des lois physiques nécessaires à la compréhension des systèmes moléculaires est issue de la mécanique quantique. Un bref retour théorique est proposé dans ce chapitre, allant des fondements de cette physique vers les méthodes actuelles de calcul en chimie quantique. Plus spécifiquement, nous abordons les concepts de base des méthodes exposées dans les chapitres suivants.

1.1 Les fondements

La physique quantique a vu le jour au début du XXe siècle pour expliquer des expériences dont les résultats étaient inexplicables par la physique classique.

1.1.1 La dualité onde-particule

La physique classique (dites "newtonienne") permet de décrire l'évolution spatiale et temporelle de deux catégories d'objets qui transportent de l'énergie :

Les ondes. Elles ne sont pas localisées dans l'espace mais sont caractérisées par leur longueur d'onde. Elles transportent de l'énergie de façon délocalisée.

Les particules. Elles sont localisées dans l'espace. La connaissance instantanée de leur position et de leur vitesse permet de calculer leur énergie et leur impulsion.

En 1900, Max Planck réussit à expliquer le spectre du corps noir en exprimant les énergies absorbées et émises sous forme de quanta proportionnels à la fréquence des ondes électromagnétiques [3]. Cinq ans plus tard, Albert Einstein comprend l'effet photoélectrique en affirmant que l'énergie du rayonnement électromagnétique est elle-même quantifiée [4]. Les ondes électromagnétiques révèlent alors un comportement de faisceau de particules appelées photons.

En 1924, Louis de Broglie généralise la dualité observée pour les photons, en associant à toute particule un caractère ondulatoire [5]. Son hypothèse est confirmée trois ans plus tard par des expériences de diffraction d'électrons [6]. Cette généralisation mène à une relation où l'impulsion de toute particule est inversement proportionnelle à sa longueur d'onde associée. Ainsi, la masse importante d'un objet macroscopique induit une longueur d'onde trop courte pour permettre l'observation du caractère ondulatoire.

Louis de Broglie associait une onde réelle à toute particule, or les objets quantiques ne sont ni l'un ni l'autre. Ils se comportent simplement soit comme des ondes (interférences, diffraction), soit comme des particules (effet photoélectrique, diffusion Compton)¹. Des travaux théoriques complémentaires vont alors permettre de définir plus précisément la nature des objets quantiques.

1.1.2 L'incertitude d'Heisenberg

Pour localiser un objet quantique avec précision, un photon doit interagir avec lui. Durant l'interaction, une partie de l'impulsion du photon est transférée à la particule. Il est donc impossible de mesurer avec précision simultanément la position et l'impulsion d'un objet quantique². Werner Heisenberg a déterminé cette relation d'incertitude [7] comme étant :

$$\Delta r \Delta p \geq \hbar/2 \quad (1.1)$$

où Δr et Δp sont les incertitudes respectives sur la position et l'impulsion, et \hbar la constante de Planck divisée par 2π .

De même, il existe une relation d'incertitude entre d'autres couples d'observables, comme celui du temps et de l'énergie :

$$\Delta \mathcal{E} \Delta t \geq \hbar/2 \quad (1.2)$$

¹Une analogie simple à la dualité onde-corpuscule est la projection d'un cylindre sur deux murs perpendiculaires que l'on observe soit comme un rectangle soit comme un cercle, bien que l'objet ne soit ni l'un ni l'autre.

²A l'échelle macroscopique, cette incertitude de mesure est négligeable et les lois de Newton sont par conséquent applicables.

où $\Delta\mathcal{E}$ et Δt sont les incertitudes respectives sur l'énergie et le temps. Ainsi, la durée de vie d'un état excité détermine la largeur naturelle de la bande spectroscopique associée.

1.1.3 Les postulats

Entre 1920 et 1930, plusieurs scientifiques éminents ont défini précisément les lois de cette physique. Tandis que Werner Heisenberg développait la mécanique quantique sous une forme matricielle, Erwin Schrödinger l'établissait comme une mécanique ondulatoire [8]. Les membres de l'Ecole de Copenhague soutenaient que la mécanique quantique pouvait être définie par un ensemble de postulats.

Postulat 1 *L'état d'un système physique est complètement défini à tout instant t par la connaissance de son vecteur d'état $|\psi(t)\rangle$.*

La fonction d'onde du système $\psi(\{\vec{r}_a\}, t)$ est la représentation du vecteur d'état dans l'espace des coordonnées, tel que :

$$\psi(\{\vec{r}_a\}, t) = \langle \vec{r}_1, \vec{r}_2, \dots, \vec{r}_n | \psi(t) \rangle \quad (1.3)$$

où $\{\vec{r}_a\}$ est l'ensemble des coordonnées des particules $1, 2, \dots, n$ du système.

Le carré du module de la fonction d'onde représente la densité de probabilité de trouver le système dans la configuration spatiale $\{\vec{r}_a\}$. La fonction d'onde doit être normée, car la probabilité de trouver le système dans tout l'espace est égale à un, tel que :

$$\begin{aligned} 1 &= \langle \psi(t) | \psi(t) \rangle \\ &= \iiint \psi^*(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, t) \psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, t) d\vec{r}_1, d\vec{r}_2, \dots, d\vec{r}_n \\ &= \iiint |\psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, t)|^2 d\vec{r}_1, d\vec{r}_2, \dots, d\vec{r}_n \end{aligned} \quad (1.4)$$

Postulat 2 *A chaque observable de la mécanique classique, il correspond un opérateur linéaire et hermitique en mécanique quantique.*

Postulat 3 *Dans toute mesure d'une observable associée à un opérateur \hat{A} , les seules valeurs qui seront toujours obtenues sont les valeurs propres a_k qui satisfont :*

$$\hat{\mathcal{A}}|\psi_k\rangle = a_k|\psi_k\rangle \quad (1.5)$$

Les vecteurs propres $|\psi_k\rangle$ de $\hat{\mathcal{A}}$ sont orthonormaux entre eux, tel que :

$$\langle\psi_l|\psi_k\rangle = \delta_{lk} = \begin{cases} 1 & \text{si } l = k \\ 0 & \text{si } l \neq k \end{cases} \quad (1.6)$$

où δ_{lk} est le delta de Kronecker.

Le vecteur d'état du système contient toutes les informations, notamment celles concernant l'observable \mathcal{A} en étant une combinaison linéaire de ses états (ou vecteurs) propres :

$$|\psi(t)\rangle = \sum_{k=1}^K |\psi_k\rangle c_k(t) \quad (1.7)$$

où K est le nombre d'états propres de \mathcal{A} accessibles au système.

Lors d'une mesure de la propriété \mathcal{A} , la probabilité de trouver la valeur a_k est :

$$\begin{aligned} p_k(t) &= \langle\psi(t)|\psi_k\rangle \langle\psi_k|\psi(t)\rangle \\ &= c_k^*(t) c_k(t) \\ &= |c_k(t)|^2 \end{aligned} \quad (1.8)$$

Si deux opérateurs commutent, il est possible de leur trouver une base de vecteurs propres communs et par conséquent de mesurer simultanément avec précision les 2 observables associées. Dans le cas contraire, il existe une relation d'incertitude entre les deux propriétés physiques (cf § 1.1.2).

Postulat 4 *La valeur moyenne de toute grandeur observable correspondant à l'opérateur $\hat{\mathcal{A}}$ est donnée par :*

$$\langle\mathcal{A}\rangle = \langle\psi(t)|\hat{\mathcal{A}}|\psi(t)\rangle \quad (1.9)$$

Postulat 5 *Le vecteur d'état d'un système évolue dans le temps suivant l'équation de Schrödinger dépendante du temps :*

$$i\hbar \frac{\partial}{\partial t}|\psi(t)\rangle = \hat{\mathcal{H}}(t)|\psi(t)\rangle \quad (1.10)$$

où $\hat{\mathcal{H}}(t)$ est l'opérateur hamiltonien associé au système.

1.2 La chimie quantique

La chimie quantique est le domaine de la mécanique quantique consacré à l'étude des systèmes moléculaires.

1.2.1 Les états stationnaires

En chimie quantique, les électrons et les noyaux sont considérés comme particules élémentaires car ils se conservent lors des réactions³. Un électron gravitant autour d'un noyau est assimilable à un courant électrique. Par conséquent, l'électron devrait rayonner de l'énergie et alors se rapprocher du noyau pour finalement tomber sur lui. En 1913, Niels Bohr postula que les électrons ne se déplacent autour du noyau que sur des orbites déterminées et énergétiquement stables, correspondant à des états stationnaires [9]. Il n'y a émission de lumière que lorsque l'électron passe d'une orbite à une autre.

Dans le référentiel propre à la molécule, on peut alors considérer que les charges sont au repos et que les interactions entre celles-ci sont de type électrostatique et non pas électromagnétique. Ainsi, l'hamiltonien d'un système moléculaire au repos est indépendant du temps, ce qui est en accord avec le principe de conservation d'énergie pour un système isolé. En conséquence, les variables de temps et d'espace peuvent être séparées dans la fonction d'onde du système :

$$\psi(\{\vec{r}_a\}, \{\vec{R}_A\}, t) = \Phi(\{\vec{r}_a\}, \{\vec{R}_A\}) e^{-i\mathcal{E}t/\hbar} \quad (1.11)$$

Même dans les cas où l'hamiltonien évolue dans le temps, comme lors d'une expérience de spectroscopie, seuls les états stationnaires $\Phi_n(\{\vec{r}_a\}, \{\vec{R}_A\})$ seront généralement recherchés. Ils sont les états propres de l'hamiltonien du système au repos et ainsi solutions de l'équation de Schrödinger indépendante du temps :

$$\hat{\mathcal{H}} \Phi(\{\vec{r}_a\}, \{\vec{R}_A\}) = \mathcal{E} \Phi(\{\vec{r}_a\}, \{\vec{R}_A\}) \quad (1.12)$$

³Leurs coordonnées seront notées respectivement en minuscule et en majuscule.

1.2.2 Le principe variationnel

Tout état approché $\tilde{\Phi}$ d'un système possède une énergie supérieure ou égale à celle de l'état fondamental exact, tel que :

$$\underbrace{\langle \tilde{\Phi} | \hat{\mathcal{H}} | \tilde{\Phi} \rangle}_E \geq \underbrace{\langle \Phi_0 | \hat{\mathcal{H}} | \Phi_0 \rangle}_{\mathcal{E}_0} \quad (1.13)$$

Une extension de ce principe est la méthode des variations linéaires, dans laquelle la fonction d'essai $\tilde{\Phi}(\vec{r})$ représentant l'état du système est une combinaison linéaire de plusieurs fonctions :

$$\tilde{\Phi}(\vec{r}) = \sum_{i=1}^L f_i(\vec{r}) c_i \quad (1.14)$$

La fonction d'essai optimale est obtenue en minimisant l'énergie par rapport aux coefficients $\{c_i\}$. Ces dérivations mènent à un système d'équations linéaires qui s'écrit sous la forme d'un déterminant séculaire :

$$|\underline{H} - \underline{E} \underline{S}| = 0 \quad (1.15)$$

où \underline{H} et \underline{E} sont les matrices respectives de l'hamiltonien et de l'énergie du système. \underline{S} est la matrice de recouvrement dont un élément S_{ij} représente le volume partagé par 2 fonctions $f_i(\vec{r})$ et $f_j(\vec{r})$ dans tout l'espace, tel que :

$$S_{ij} = \iiint f_i^*(\vec{r}) f_j(\vec{r}) d\vec{r} \quad (1.16)$$

La résolution du déterminant séculaire (1.15) fournit L solutions $\tilde{\Phi}_n(\vec{r})$ qui constituent une représentation approchée des états de plus basse énergie. Si un nombre infini de fonctions était utilisé, alors les solutions représenteraient exactement tous les états stationnaires du système.

1.2.3 L'hamiltonien moléculaire

En unités atomiques, l'hamiltonien non-relativiste d'un système de N électrons et M noyaux est :

$$\begin{aligned} \hat{\mathcal{H}} = & - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{a=1}^N \frac{1}{2} \nabla_a^2 - \sum_{A=1}^M \sum_{a=1}^N \frac{Z_A}{|\vec{R}_A - \vec{r}_a|} \\ & + \sum_{a=1}^N \sum_{b>a}^N \frac{1}{|\vec{r}_a - \vec{r}_b|} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \end{aligned} \quad (1.17)$$

où \vec{R}_A , M_A et Z_A sont la position, la masse et le numéro atomique respectifs du noyau A ; \vec{r}_a est la position de l'électron a .

Les deux premiers termes sont les opérateurs d'énergie cinétique des noyaux et des électrons, définis par le laplacien ∇^2 . Les trois derniers termes définissent l'ensemble des interactions électrostatiques entre particules par la loi de Coulomb. La présence des distances inter-particules, dans ces derniers, rend impossible la séparation des variables et ainsi la résolution exacte de l'équation de Schrödinger indépendante du temps (1.12) pour un système moléculaire. Des approximations doivent être utilisées pour trouver les solutions approchées de l'équation.

1.2.4 L'approximation de Born-Oppenheimer

Par la différence de leur masse respective, les échelles de temps des noyaux et des électrons peuvent être séparées dans la fonction d'onde tel que :

$$\Phi(\{\vec{r}_a\}, \{\vec{R}_A\}) = \Phi^{el}(\{\vec{r}_a\}; \{\vec{R}_A\}) \Phi^{nuc}(\{\vec{R}_A\}) \quad (1.18)$$

où les coordonnées nucléaire $\{\vec{R}_A\}$ deviennent des paramètres pour la fonction d'onde électronique $\Phi^{el}(\{\vec{r}_a\}; \{\vec{R}_A\})$.

A l'échelle de temps la plus rapide, les électrons se déplacent dans le champs des noyaux fixes et l'hamiltonien moléculaire (1.17) devient :

$$\hat{\mathcal{H}} = \underbrace{-\sum_{a=1}^N \frac{1}{2} \nabla_a^2 - \sum_{A=1}^M \sum_{a=1}^N \frac{Z_A}{|\vec{R}_A - \vec{r}_a|}}_{\hat{\mathcal{H}}^{el}} + \sum_{a=1}^N \sum_{b>a}^N \frac{1}{|\vec{r}_a - \vec{r}_b|} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \quad (1.19)$$

Le terme d'interaction entre noyaux étant constant, le problème (1.12) se réduit à la résolution de la structure électronique :

$$\hat{\mathcal{H}}^{el} \Phi^{el}(\{\vec{r}_a\}; \{\vec{R}_A\}) = \mathcal{E}^{el}(\{\vec{R}_A\}) \Phi^{el}(\{\vec{r}_a\}; \{\vec{R}_A\}) \quad (1.20)$$

où l'énergie électronique $\mathcal{E}^{el}(\{\vec{R}_A\})$ est calculée pour une géométrie $\{\vec{R}_A\}$ donnée.

A l'échelle de temps la plus lente, les noyaux se déplacent dans le champs moyen

des électrons et l'hamiltonien moléculaire (1.17) devient :

$$\begin{aligned}
 \hat{\mathcal{H}} &= -\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \underbrace{\left\langle -\sum_{a=1}^N \frac{1}{2} \nabla_a^2 - \sum_{A=1}^M \sum_{a=1}^N \frac{Z_A}{|\vec{R}_A - \vec{r}_a|} + \sum_{a=1}^N \sum_{b>a}^N \frac{1}{|\vec{r}_a - \vec{r}_b|} \right\rangle}_{\mathcal{H}^{el}} \\
 &+ \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \\
 &= -\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \underbrace{\mathcal{E}^{el}(\{\vec{R}_A\})}_{\mathcal{E}^{pot}(\{\vec{R}_A\})} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \quad (1.21)
 \end{aligned}$$

où $\mathcal{E}^{pot}(\{\vec{R}_A\})$ constitue l'énergie potentielle du système.

Dans l'approximation de Born-Oppenheimer [10], les noyaux se déplacent sur des surfaces d'énergie potentielle correspondant aux états électroniques Φ_n^{el} de la molécule. Cette approximation "adiabatique" est en général parfaitement valide. Toutefois, elle peut devenir discutable dans certains cas, notamment en photochimie. Avant toute exploration des surfaces d'énergie potentielle, il est nécessaire de résoudre le problème (1.20) de la structure électronique ⁴.

1.2.5 La théorie des orbitales moléculaires

L'équation de Schrödinger (1.20) ne peut être résolue analytiquement que pour l'atome d'hydrogène et les ions hydrogénoïdes. Les solutions de l'équation correspondent alors à des orbitales atomiques que l'électron peut occuper. Pour les atomes à plusieurs électrons, seule une forme approchée des orbitales peut être déterminée. Le terme de répulsion entre les électrons dans l'hamiltonien (1.19) rend impossible la résolution exacte de l'équation de Schrödinger. En chimie générale, une forme similaire aux orbitales des ions hydrogénoïdes est communément utilisée pour décrire les orbitales des atomes polyélectroniques. Leur numéro atomique doit cependant être considéré afin de caractériser convenablement la taille et l'énergie des orbitales.

Pour une molécule, la théorie des orbitales moléculaires énonce que les orbitales spatiales sont une combinaison linéaire d'orbitales atomiques [11]. Cette théorie permet de comprendre la nature des liaisons chimiques ainsi que de nombreux phéno-

⁴Afin d'alléger les notations, l'indice "el" et les paramètres $\{\vec{R}_A\}$ seront désormais omis.

mènes expérimentaux (diamagnétisme du dioxygène, conductivité du graphite, ...).

Pour déterminer les orbitales d'une molécule, il est peu commode d'utiliser les orbitales issues des ions hydrogenoides qui sont le produit d'une harmonique sphérique et d'une fonction radiale. John Slater proposa de remplacer le polynôme de la composante radiale par une fonction exponentielle plus simple à manipuler [12]. Bien qu'elles décrivent convenablement les orbitales atomiques, les fonctions de Slater conduisent à des intégrales difficiles à calculer ; ainsi, des gaussiennes (*Gaussian Type Orbital*) appelées "primitives" leur sont généralement préférées :

$$\eta^{GTO}(\alpha, \vec{r} - \vec{R}_A) = N (x - x_A)^k (y - y_A)^l (z - z_A)^m e^{-\alpha(\vec{r} - \vec{R}_A)^2} \quad (1.22)$$

où \vec{R}_A est la position du noyau atomique A de coordonnées cartésiennes (x_A, y_A, z_A) et N le facteur de normalisation de la primitive. L'exposant α définit la largeur de la fonction gaussienne afin de caractériser une couche électronique (nombre quantique principal). La symétrie des différents types d'orbitales atomiques (s, p, ...) est respectée par les facteurs k , l et m dont la somme est égale au nombre quantique azimutal (0, 1, ...).

Une combinaison linéaire de primitives (*Contracted Gaussian Functions*) est cependant nécessaire pour retrouver une forme juste des orbitales de chaque atome :

$$\eta_{\mu}^{CGF}(\vec{r} - \vec{R}_A) = \sum_{g=1}^K \eta_g^{GTO}(\alpha, \vec{r} - \vec{R}_A) D_{g\mu} \quad (1.23)$$

où $D_{g\mu}$ sont les coefficients de contraction, dont les valeurs optimales sont tabulées tout comme celles des exposants α ; K est le nombre de primitives utilisées pour constituer les fonctions gaussiennes contractées, lesquelles correspondent aux fonctions de base disponibles pour représenter une orbitale atomique.

Différents types de base de fonctions sont utilisées en chimie quantique et une nomenclature spécifique leur est associée. Les bases minimales décrivent une orbitale atomique à l'aide d'une seule fonction de base. Par exemple, la fonction unique d'une orbitale atomique utilisée par les bases STO-3G est une combinaison de 3 primitives qui permet de retrouver la forme approchée d'une orbitale de Slater (*Slater Type Orbital*). Les bases étendues utilisent plusieurs fonctions de base pour décrire une orbitale atomique. Pour une base 6-31G, l'extension est limitée aux orbitales atomiques

susceptibles d'évoluer au cours du temps : les orbitales de cœur sont représentées par une fonction de base composée de 6 primitives, tandis que les orbitales de valence sont somme de deux fonctions de base composées l'une de 3 primitives et l'autre d'une seule. Pour les orbitales de valence, le poids de chacune des 2 fonctions de base est optimisé lors du calcul variationnel. Cette flexibilité permet de mieux ajuster l'extension spatiale d'une orbitale donnée dans son environnement moléculaire. Dans une base 6-31+G*, des fonctions de base supplémentaires dites de diffusion (+) permettent de représenter la densité électronique loin des noyaux. La déformation d'une orbitale atomique par l'approche d'un autre atome est prise en considération par des orbitales de polarisation (*). D'autres bases étendues communément utilisées sont celles de type cc-pVXZ (*correlation-consistent polarized valence X zeta*) qui utilisent X fonctions de base pour décrire une orbitale atomique de valence.

Au final, une orbitale moléculaire obtenue par le calcul est une combinaison linéaire de fonctions de base tel que :

$$\phi_i(\vec{r}) = \sum_{\mu=1}^L \eta_{\mu}(\vec{r}) C_{\mu i} \quad (1.24)$$

Plus le jeu de fonctions de base est étendu, plus la fonction d'essai possède de paramètres variationnels $C_{\mu i}$ et mieux l'orbitale moléculaire est déterminée.

1.2.6 Le principe d'exclusion de Pauli

Comme toute particule quantique, l'électron possède une propriété intrinsèque de spin associée à son moment angulaire. Cette propriété a été découverte expérimentalement afin d'expliquer l'effet Zeeman [13] et le dédoublement de la raie jaune du spectre d'émission du sodium [14].

Un électron possède deux états propres de spin $|\alpha\rangle$ et $|\beta\rangle$ pour lesquels la composante selon l'axe z du moment angulaire de spin est égale respectivement à $+1/2$ et $-1/2$. L'état électronique d'un système de spin total S est de multiplicité $2S+1$ relative au nombre d'états propres de la composante selon z du moment angulaire de spin. Par exemple, un état triplet (S=1) est constitué de trois micro-états dégénérés de composantes de spin $+1$, 0 et -1 selon l'axe z.

Toutefois, l'hamiltonien électronique non-relativiste de notre équation (1.20) dépend uniquement des coordonnées spatiales des électrons et ne peut pas expliquer certains phénomènes tels que ceux causés par le couplage spin-orbite. Seuls des hamiltoniens prenant en considération le spin peuvent être utilisés pour interpréter la structure hyperfine de spectres, comme le permet notamment celui de la théorie quantique relativiste formulée par Paul Dirac [15]. La fonction d'onde électronique est pour sa part fonction des coordonnées spatiales \vec{r} de l'électron et de sa coordonnée de spin w , et devient ainsi $\Phi(\{\vec{x}_a\})$ avec pour coordonnée électronique :

$$\vec{x} = (\vec{r}, w) \quad (1.25)$$

Du fait de l'indiscernabilité des électrons, la fonction d'onde électronique doit être antisymétrique par rapport à l'échange des coordonnées \vec{x} de 2 électrons, tel que l'énonce le principe d'exclusion de Pauli :

$$\Phi(\vec{x}_1, \dots, \vec{x}_i, \dots, \vec{x}_j, \dots, \vec{x}_N) = - \Phi(\vec{x}_1, \dots, \vec{x}_j, \dots, \vec{x}_i, \dots, \vec{x}_N) \quad (1.26)$$

Dans un système moléculaire, chaque électron occupe une spin-orbitale qui est le produit d'une orbitale spatiale et d'une fonction de spin :

$$\chi(\vec{x}) = \begin{cases} \phi(\vec{r})\alpha(w) \\ \text{ou} \\ \phi(\vec{r})\beta(w) \end{cases} \quad (1.27)$$

Le principe d'exclusion de Pauli impose que seuls 2 électrons de spins opposés peuvent occuper la même orbitale spatiale [16].

1.3 La théorie Hartree-Fock

La méthode Hartree-Fock est un moyen d'obtenir des solutions approchées de l'équation de Schrödinger pour des systèmes à plusieurs électrons.

1.3.1 L'approximation orbitale

Les électrons sont corrélés entre eux par le fait que la probabilité de trouver un électron dans une position donnée dépend directement de la position instantanée des autres électrons du système. L'approximation orbitale consiste à négliger cette corrélation électronique en formulant la fonction d'onde ⁵ d'un système à N électrons

⁵Nous conservons la notation Φ et \mathcal{E} pour la fonction d'onde électronique exacte et nous adoptons la notation Ψ et E pour les solutions approchées de type Hartree-Fock.

comme le produit des spin-orbitales occupées [17] :

$$\Psi^{HP}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \chi_a(\vec{x}_1)\chi_b(\vec{x}_2) \cdots \chi_q(\vec{x}_N) \quad (1.28)$$

Toutefois, le produit de Hartree Ψ^{HP} ne respecte pas le principe d'antisymétrie [18] et il est nécessaire d'écrire la fonction d'onde sous forme de déterminant de Slater :

$$\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = (N!)^{-1/2} \begin{vmatrix} \chi_a(\vec{x}_1) & \chi_b(\vec{x}_1) & \cdots & \chi_q(\vec{x}_1) \\ \chi_a(\vec{x}_2) & \chi_b(\vec{x}_2) & \cdots & \chi_q(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_a(\vec{x}_N) & \chi_b(\vec{x}_N) & \cdots & \chi_q(\vec{x}_N) \end{vmatrix} \quad (1.29)$$

où $(N!)^{-1/2}$ est le facteur de normalisation du déterminant.

1.3.2 L'équation de Hartree-Fock

Dans le cadre de l'approximation orbitale, la fonction d'essai antisymétrique la plus simple pour décrire l'état fondamental électronique d'un système à N électrons est le déterminant de Slater (1.29) que l'on peut réécrire ⁶ comme :

$$|\Psi_0\rangle = |\chi_a\chi_b \cdots \chi_q\rangle \quad (1.30)$$

La flexibilité du calcul variationnel réside dans le choix des spin-orbitales. En minimisant l'expression de l'énergie associée au déterminant par rapport au choix de ses spin-orbitales, nous obtenons une équation aux valeurs propres appelée équation de Hartree-Fock qui détermine les spin-orbitales optimales :

$$\underbrace{\left(-\frac{1}{2}\nabla_1^2 - \sum_{A=1}^M \frac{Z_A}{|\vec{R}_A - \vec{r}_1|} + \left\langle \sum_{e \neq 1}^N \frac{1}{|\vec{r}_1 - \vec{r}_e|} \right\rangle \right)}_{\hat{f}_1} \chi(\vec{x}_1) = \mathcal{E} \chi(\vec{x}_1) \quad (1.31)$$

où \hat{f}_1 est un opérateur effectif mono-électronique appelé opérateur de Fock. Les deux premiers termes de \hat{f}_1 rassemblent les opérateurs mono-électroniques relatifs à l'énergie cinétique de l'électron et son attraction avec les noyaux⁷. Le troisième terme caractérise le potentiel moyen ressenti par l'électron 1 causé par la présence de chacun des autres. De ce fait, l'opérateur \hat{f}_1 dépend des spin-orbitales des électrons différents de 1 et par conséquent de ces fonctions propres. Ainsi, l'équation de Hartree-Fock est

⁶Cette notation conserve uniquement la diagonale du déterminant complet en respectant l'ordre des coordonnées d'électrons de 1 à N.

⁷Ces termes sont désignés généralement par l'opérateur monoélectronique \hat{h}_1 .

non-linéaire et doit être résolue itérativement.

L'astuce de l'approximation Hartree-Fock (HF) est de remplacer le problème à plusieurs électrons (1.20) par un problème à un seul électron (1.31) dans lequel la répulsion électronique est traitée de façon moyennée. Le déterminant de Slater (1.30) est constitué des spin-orbitales de plus basse énergie. La meilleure estimation de l'énergie de l'état fondamental que l'on puisse obtenir à partir de cette fonction d'onde est appelée la limite Hartree-Fock. La différence entre cette valeur limite et la valeur exacte correspond à l'énergie de corrélation.

1.3.3 Les équations de Roothaan-Hall

Pour un système moléculaire dont l'état fondamental est un singulet, il est d'usage d'appliquer la méthode RHF (*Restricted Hartree-Fock*). Dans ce traitement, les orbitales spatiales de plus basse énergie sont contraintes d'être doublement occupées par des électrons de spins opposés. Ces orbitales appelées couches fermées constituent le déterminant de Slater de l'état fondamental. La méthode RHF permet de transformer l'équation générale de Hartree-Fock (1.31) en une nouvelle équation aux valeurs propres :

$$\hat{f}_1 \phi_i(\vec{r}_1) = \varepsilon_i \phi_i(\vec{r}_1) \quad (1.32)$$

où \hat{f}_1 est l'opérateur de Fock et ε_i l'énergie de l'orbitale spatiale $\phi_i(\vec{r}_1)$.

Pour un système moléculaire de spin total non nul, certaines orbitales spatiales appelées couches ouvertes ne contiennent qu'un électron. Dans ce cas, il est nécessaire d'utiliser la méthode UHF (*Unrestricted Hartree-Fock*) pour laquelle deux jeux d'orbitales spatiales différents sont déterminés pour les spin-orbitales de type α et pour celles de type β . Les équations définissant les deux systèmes de spin demeurent similaires à celles du traitement RHF.

L'application de la théorie des orbitales moléculaires (1.24) à l'équation RHF (1.32) mène aux équations de Roothaan-Hall [19, 20] :

$$\sum_{\nu=1}^L F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_{\nu=1}^L S_{\mu\nu} C_{\nu i} \quad i = 1, 2, \dots, L \quad (1.33)$$

L'élément $F_{\mu\nu}$ de la matrice de Fock $\underline{\underline{F}}$ est :

$$F_{\mu\nu} = \iiint \phi_{\mu}^*(\vec{r}_1) \hat{f}_1 \phi_{\nu}(\vec{r}_1) d\vec{r}_1 \quad (1.34)$$

L'élément $S_{\mu\nu}$ de la matrice de recouvrement $\underline{\underline{S}}$ est :

$$S_{\mu\nu} = \iiint \phi_{\mu}^*(\vec{r}_1) \phi_{\nu}(\vec{r}_1) d\vec{r}_1 \quad (1.35)$$

Le système d'équations (1.33) peut être réécrit sous forme matricielle :

$$\underline{\underline{F}} \underline{\underline{C}} = \underline{\underline{S}} \underline{\underline{C}} \underline{\underline{\varepsilon}} \quad (1.36)$$

où $\underline{\underline{C}}$ est la matrice carrée des coefficients linéaires $C_{\mu i}$:

$$\underline{\underline{C}} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1L} \\ C_{21} & C_{22} & \cdots & C_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ C_{L1} & C_{L2} & \cdots & C_{LL} \end{pmatrix} \quad (1.37)$$

et $\underline{\underline{\varepsilon}}$ est la matrice diagonale des énergies ε_i :

$$\underline{\underline{\varepsilon}} = \begin{pmatrix} \varepsilon_1 & & & \mathbf{0} \\ & \varepsilon_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \varepsilon_L \end{pmatrix} \quad (1.38)$$

1.3.4 Les matrices caractéristiques

Les développements précédents mènent à deux matrices caractéristiques :

La matrice densité

La densité électronique est définie comme le carré de la fonction d'onde électronique (cf § 1.1.3 : postulat 1). Si L orbitales HF sont déterminées à partir des fonctions de bases, alors nous pouvons écrire la densité électronique comme :

$$\rho(\vec{r}) = \sum_{i=1}^L n_i \phi_i^*(\vec{r}) \phi_i(\vec{r}) \quad (1.39)$$

$$\begin{aligned} &= \sum_{i=1}^L n_i \left(\sum_{\nu=1}^L \eta_{\nu}^*(\vec{r}) C_{\nu i}^* \right) \left(\sum_{\mu=1}^L \eta_{\mu} C_{\mu i}(\vec{r}) \right) \\ &= \sum_{\mu=1}^L \sum_{\nu=1}^L \underbrace{\sum_{i=1}^L n_i C_{\mu i} C_{i\nu}^{\dagger}}_{P_{\mu\nu}} \eta_{\nu}^*(\vec{r}) \eta_{\mu}(\vec{r}) \end{aligned} \quad (1.40)$$

où n_i est le nombre d'électrons (0,1 ou 2) occupant l'orbitale ϕ_i et $P_{\mu\nu}$ un élément de la matrice densité $\underline{\underline{P}}$.

La matrice de Fock

La densité électronique pouvant être définie soit par la matrice $\underline{\underline{C}}$ ou la matrice $\underline{\underline{P}}$, il est facile d'exprimer les éléments de la matrice de Fock RHF en fonction de cette densité :

$$F_{\mu\nu} = \underbrace{(\mu|\hat{h}_1|\nu)}_{H_{\mu\nu}^{core}} + \underbrace{\sum_{\lambda=1}^L \sum_{\sigma=1}^L P_{\lambda\sigma} [(\mu\nu|\sigma\lambda) - \frac{1}{2}(\mu\lambda|\sigma\nu)]}_{G_{\mu\nu}} \quad (1.41)$$

où $H_{\mu\nu}^{core}$ est un élément de la matrice de l'hamiltonien de coeur $\underline{\underline{H}}^{core}$ et $G_{\mu\nu}$ un élément de la partie biélectronique $\underline{\underline{G}}$ de la matrice de Fock. La présence des éléments $P_{\lambda\sigma}$ de la matrice densité rappelle que les équations de Roothaan-Hall doivent être résolues itérativement.

Le coût principal d'un calcul RHF réside dans la détermination des intégrales biélectroniques à quatre centres $(\mu\nu|\sigma\lambda)$ et de la matrice $\underline{\underline{G}}$. Le temps d'un calcul RHF est ainsi proportionnel au nombre de fonctions de base à la puissance quatre, soit L^4 .

1.3.5 La transformation symétrique

Les équations de Roothaan-Hall (1.36) ne correspondent pas à un problème aux valeurs propres, puisque le jeu des fonctions de base n'est pas orthonormal. Une orthogonalisation des fonctions de base est d'abord nécessaire afin de pouvoir diagonaliser la matrice de Fock. Une des procédures possible est la transformation symétrique.

L'étape initiale consiste à diagonaliser la matrice de recouvrement par transformation unitaire :

$$\underline{\underline{T}}^\dagger \underline{\underline{S}} \underline{\underline{T}} = \underline{\underline{s}} \quad (1.42)$$

où $\underline{\underline{s}}$ est une matrice diagonale contenant les valeurs propres de $\underline{\underline{S}}$.

$\underline{\underline{T}}$ est la matrice unitaire qui contient les coefficients associés aux vecteurs propres de $\underline{\underline{S}}$.

L'étape suivante consiste à obtenir la matrice de transformation $\underline{\underline{X}}$:

$$\underline{\underline{X}} \equiv \underline{\underline{S}}^{-\frac{1}{2}} = \underline{\underline{T}} \underline{\underline{s}}^{-\frac{1}{2}} \underline{\underline{T}}^\dagger \quad (1.43)$$

Si on applique la matrice de transformation à la matrice $\underline{\underline{C}}$, alors :

$$\underline{\underline{C'}} = \underline{\underline{X}}^{-1} \underline{\underline{C}} \quad \underline{\underline{C}} = \underline{\underline{X}} \underline{\underline{C'}} \quad (1.44)$$

la substitution dans les équations de Roothaan-Hall donne :

$$\underline{\underline{F}} \underline{\underline{X}} \underline{\underline{C'}} = \underline{\underline{S}} \underline{\underline{X}} \underline{\underline{C'}} \underline{\underline{\varepsilon}} \quad (1.45)$$

et la multiplication par $\underline{\underline{X}}^\dagger$ donne⁸ :

$$\underbrace{\underline{\underline{X}}^\dagger \underline{\underline{F}} \underline{\underline{X}}}_{\underline{\underline{F'}}} \underline{\underline{C'}} = \underbrace{\underline{\underline{X}}^\dagger \underline{\underline{S}} \underline{\underline{X}}}_{\underline{\underline{I}}} \underline{\underline{C'}} \underline{\underline{\varepsilon}} \quad (1.46)$$

où $\underline{\underline{F'}}$ est la matrice de Fock transformée et $\underline{\underline{I}}$ la matrice identité.

On obtient finalement un problème aux valeurs propres :

$$\underline{\underline{F'}} \underline{\underline{C'}} = \underline{\underline{C'}} \underline{\underline{\varepsilon}} \quad (1.47)$$

Une diagonalisation de la matrice $\underline{\underline{F'}}$ permet d'obtenir les valeurs et vecteurs propres voulus.

1.3.6 La procédure SCF

Comme nous l'avons mentionné au paragraphe 1.3.2, la détermination de la fonction d'onde Hartree Fock nécessite un calcul itératif. Les étapes de la procédure SCF (*Self Consistent Field*) d'un calcul RHF sont brièvement énumérées ci-dessous :

1. Définition du système $(\{\vec{R}_A\}, \{Z_A\}, N)$ et des fonctions de base $(\{\eta_\mu(\vec{r})\})$.
2. Calcul des matrices $\underline{\underline{S}}$, $\underline{\underline{H}}^{core}$ et des intégrales $(\mu\nu|\sigma\lambda)$.
3. Diagonalisation de la matrice $\underline{\underline{S}}$ et calcul de la matrice $\underline{\underline{X}}$.
4. Suggestion initiale de la matrice densité $\underline{\underline{P}}$.
5. Calcul de la matrice $\underline{\underline{G}}$ à partir de la matrice $\underline{\underline{P}}$ et des intégrales $(\mu\nu|\sigma\lambda)$.
6. Calcul de la matrice de Fock $\underline{\underline{F}} = \underline{\underline{H}}^{core} + \underline{\underline{G}}$.
7. Calcul de la matrice de Fock transformée $\underline{\underline{F'}} = \underline{\underline{X}}^\dagger \underline{\underline{F}} \underline{\underline{X}}$.
8. Diagonalisation de la matrice $\underline{\underline{F'}}$ pour obtenir $\underline{\underline{C'}}$ et $\underline{\underline{\varepsilon}}$.

⁸Tout comme la matrice de recouvrement, $\underline{\underline{X}}$ est symétrique et donc égale à son adjoint.

9. Calcul de la matrice $\underline{\underline{C}} = \underline{\underline{X}} \underline{\underline{C}}'$.
10. Calcul de la nouvelle matrice densité $\underline{\underline{P}}$ à partir de $\underline{\underline{C}}$.
11. Comparaison de la nouvelle matrice densité avec la précédente : si le critère de convergence n'est pas satisfait, retour à l'étape 5 avec la nouvelle matrice $\underline{\underline{P}}$.
12. Obtention de $\underline{\underline{P}}$ pouvant servir au calcul de propriétés.

1.3.7 Le calcul des observables

Au même titre que la fonction d'onde, la matrice densité contient toutes les informations concernant la structure électronique du système à l'état fondamental. Par conséquent, il est possible de retrouver toute propriété du système à partir de la matrice densité.

La plupart des observables sont une somme de contributions monoélectroniques et il est possible d'écrire leur valeur moyenne comme :

$$\langle A \rangle = \sum_{\mu=1}^L \sum_{\nu=1}^L P_{\mu\nu} A_{\nu\mu} \quad (1.48)$$

où $A_{\nu\mu}$ est un élément matriciel de l'opérateur monoélectronique associé à l'observable A .

En revanche, si une observable est caractérisée par une somme de contributions biélectroniques alors la valeur moyenne sera :

$$\langle B \rangle = \frac{1}{2} \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{\eta=1}^L \sum_{\lambda=1}^L P_{\mu\nu} P_{\eta\lambda} B_{\nu\mu\lambda\eta} \quad (1.49)$$

où $B_{\nu\mu\lambda\eta}$ est un élément matriciel de l'opérateur biélectronique associé à l'observable B .

L'énergie de l'état fondamental E_0 étant définie par un somme d'opérateurs monoélectroniques et biélectroniques (1.41), nous pouvons l'écrire comme :

$$E_0 = \sum_{\mu=1}^L \sum_{\nu=1}^L P_{\mu\nu} \left(H_{\nu\mu}^{core} + \frac{1}{2} G_{\nu\mu} \right) \quad (1.50)$$

1.3.8 L'énergie de l'état fondamental

L'expression (1.50) est le résultat d'un calcul RHF réalisé à l'aide d'un jeu de fonctions de base. Cette énergie RHF peut s'exprimer autrement en fonction des orbitales spatiales, tel que :

$$E_0 = 2 \sum_{i=1}^{N/2} \langle i | \hat{h}_1 | i \rangle + \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} 2 \underbrace{\langle ii | jj \rangle}_{J_{ij}} - \underbrace{\langle ij | ji \rangle}_{K_{ij}} \quad (1.51)$$

J_{ij} est appelée intégrale de Coulomb :

$$\langle ii | jj \rangle = \iiint \iiint |\phi_i(\vec{r}_1)|^2 \frac{1}{|\vec{r}_1 - \vec{r}_2|} |\phi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \quad (1.52)$$

L'intégrale de Coulomb représente la répulsion coulombienne classique entre les densités électroniques $|\phi_i(\vec{r}_1)|^2$ et $|\phi_j(\vec{r}_2)|^2$.

K_{ij} est appelée intégrale d'échange :

$$\langle ij | ji \rangle = \iiint \iiint \phi_i^*(\vec{r}_1) \phi_j(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \phi_j^*(\vec{r}_2) \phi_i(\vec{r}_2) d\vec{r}_1 d\vec{r}_2 \quad (1.53)$$

L'intégrale d'échange représente le phénomène non classique que deux électrons de même spin s'évitent et donc réduisent la répulsion électronique.

D'une manière générale, l'énergie Hartree-Fock s'exprime dans le jeu des spin-orbitales occupées, tel que :

$$E_0 = \sum_{a=1}^N \langle a | \hat{h}_1 | a \rangle + \sum_{a=1}^N \sum_{b>a}^N \underbrace{\langle ab | ab \rangle - \langle ab | ba \rangle}_{\langle ab || ab \rangle} \quad (1.54)$$

$$\text{avec} \quad \langle ab || ab \rangle = \iiint \iiint \chi_a^*(\vec{x}_1) \chi_b^*(\vec{x}_2) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_a(\vec{x}_1) \chi_b(\vec{x}_2) d\vec{x}_1 d\vec{x}_2$$

L'énergie HF est en résumé une somme des contributions $\langle a | \hat{h}_1 | a \rangle$ de chaque électron et des interactions $\langle ab || ab \rangle$ entre paire unique d'électrons tel que :

$$\langle ab || ab \rangle = \begin{cases} J_{ab} & \text{si spins opposés} \\ J_{ab} - K_{ab} & \text{si spins parallèles} \end{cases} \quad (1.55)$$

Les intégrales de Coulomb et d'échange étant positives, la prise en compte des interactions d'échange entre électrons de même spin diminue E_0 afin de l'approcher de la valeur exacte de l'énergie à l'état fondamental. Des méthodes complémentaires permettent de calculer la corrélation électronique qui subsiste entre électrons de spins opposés. Cependant, l'utilisation des méthodes post-Hartree-Fock augmente rapidement le coût de calcul.

1.4 La théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité ou DFT (*Density Functional Theory*) est une approche alternative aux calculs Hartree Fock. Elle permet de décrire l'état fondamental d'un système en incluant directement la corrélation électronique.

1.4.1 Le théorème de Kohn-Hohenberg

En utilisant le principe variationnel, Walter Kohn et Pierre Hohenberg démontrent en 1964 que l'énergie d'un système à l'état fondamental est une fonctionnelle unique de la densité électronique [21]. Du fait que la fonction d'essai n'est plus la fonction d'onde mais la densité électronique, le principe variationnel caractéristique de la DFT devient :

$$\mathcal{E}_0[\tilde{\rho}] \geq \mathcal{E}_0[\rho] \quad (1.56)$$

Dans le cadre de l'approximation de Born-Oppenheimer, la fonctionnelle de la densité définissant l'énergie électronique fondamentale \mathcal{E}_0 s'écrit :

$$\mathcal{E}_0[\rho] = \mathcal{T}[\rho] + \sum_{A=1}^M \iiint \frac{Z_A \rho(\vec{r})}{|\vec{R}_A - \vec{r}|} d\vec{r} + \frac{1}{2} \iiint \iiint \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + \mathcal{E}_{xc}[\rho] \quad (1.57)$$

Les deux premiers termes correspondent à l'énergie cinétique des électrons et à l'énergie d'interaction électrostatique noyaux-électrons. Les deux derniers termes correspondent aux énergies de répulsion coulombienne et d'échange-correlation électroniques.

1.4.2 L'équation de Kohn-Sham

Le terme cinétique de l'énergie posant un problème majeur, Walter Kohn et Lu Sham proposent en 1965 l'utilisation de l'approximation orbitale (cf § 1.3.1) pour

déterminer la fonctionnelle de la densité [22]. Ainsi, ce terme devient la somme des énergies cinétiques de chaque électron :

$$T = \sum_{a=1}^N \langle \chi_a | -\frac{1}{2} \nabla_1^2 | \chi_a \rangle \quad (1.58)$$

La densité électronique s'écrit par conséquent de la même manière (1.39) que dans le formalisme Hartree-Fock. La flexibilité du calcul variationnel réside alors dans le choix des spin-orbitales constituant la densité. En minimisant l'expression de la fonctionnelle (1.57) par rapport à la densité, nous obtenons une équation aux valeurs propres appelée équation de Kohn-Sham qui détermine les spin-orbitales optimales :

$$\underbrace{\left(-\frac{1}{2} \nabla_1^2 - \sum_{A=1}^M \frac{Z_A}{|\vec{R}_A - \vec{r}_1|} + \iiint \frac{\rho(\vec{r})}{|\vec{r}_1 - \vec{r}|} d\vec{r} + \frac{\partial \mathcal{E}_{xc}[\rho]}{\partial \rho} \right)}_{\hat{f}_1^{KS}} \chi(\vec{x}_1) = \mathcal{E} \chi(\vec{x}_1) \quad (1.59)$$

où \hat{f}_1^{KS} est un opérateur effectif mono-électronique appelé opérateur de Kohn-Sham. Les deux premiers termes de \hat{f}_1^{KS} rassemblent les opérateurs mono-électroniques relatifs à l'énergie cinétique de l'électron et son attraction avec les noyaux. Le troisième terme définit la répulsion coulombienne ressentie par l'électron 1 causée par la densité électronique du système. De ce fait, l'opérateur \hat{f}_1^{KS} dépend de la densité électronique et par conséquent de ses spin-orbitales propres. Ainsi, l'équation de Kohn-Sham est non-linéaire et doit être résolue itérativement par une procédure similaire à celle du champ auto-cohérent (cf § 1.3.6). Comme pour les méthodes Hartree-Fock, un jeu de fonctions de base est utilisé pour former les orbitales moléculaires. La forme du terme d'échange-correlation n'étant pas connue, nous l'exprimons pour le moment par la dérivée de sa fonctionnelle par rapport à la densité.

La densité exacte du système pourrait alors être déterminée si la fonctionnelle d'échange-correlation corrigeait l'approximation faite sur le calcul de l'énergie cinétique des électrons indépendants. En appliquant la densité exacte à la fonctionnelle (1.57), l'énergie exacte du système à l'état fondamental serait obtenue. Or, la fonctionnelle d'échange-correlation n'est pas connue et des formes approchées doivent être utilisées.

1.4.3 Les fonctionnelles d'échange-corrélation

Un enjeu permanent de cette théorie est de paramétrer une fonctionnelle d'échange-correlation qui soit la plus proche possible de sa forme analytique réelle.

Approximation de la densité locale

L'approximation de la densité locale ou LDA (*Local Density Approximation*) permet de formuler une fonctionnelle d'échange-corrélation pour une densité uniforme d'électrons :

$$E_{xc}^{LDA}[\rho] = \iiint \rho(\vec{r}) \varepsilon_x[\rho(\vec{r})] d\vec{r} + \iiint \rho(\vec{r}) \varepsilon_c[\rho(\vec{r})] d\vec{r} \quad (1.60)$$

où $\varepsilon_x[\rho(\vec{r})]$ et $\varepsilon_c[\rho(\vec{r})]$ sont les énergies respectives d'échange et de corrélation par électron.

La forme analytique de la fonctionnelle d'échange est connue pour un gaz uniforme d'électron, tel que :

$$\varepsilon_x[\rho(\vec{r})] = -\frac{3}{4} \left(\frac{3\rho(\vec{r})}{\pi} \right)^{1/3} \quad (1.61)$$

Pour l'énergie de corrélation, plusieurs fonctionnelles ont été paramétrées à partir de calculs Monte-Carlo quantique sur le gaz uniforme d'électrons [23].

Approximation du gradient généralisé

Dans le cas des systèmes réels où la densité n'est pas uniforme, l'approximation LDA n'est valable que localement. La prise en compte du gradient de la densité dans la fonctionnelle permet de traiter convenablement les zones de grande variation de densité. Différentes fonctionnelles ont été développées dans l'approximation du gradient généralisé ou GGA (*Generalized Gradient Approximation*), ce qui a permis une nette amélioration des résultats pour ces systèmes.

La fonctionnelle BLYP (*Becke Lee Yang Parr*) est une fonctionnelle de type GGA qui regroupe deux améliorations par rapport à celles de type LDA : Axel Becke a tout d'abord proposé une correction ΔE_x^{B88} du terme d'échange [24] ; Chengteh Lee, Weitao Yang et Robert Parr ont ensuite développé une fonctionnelle de corrélation E_c^{LYP} tenant également compte du gradient de la densité [25].

Les fonctionnelles hybrides

Une amélioration supplémentaire a été apportée aux fonctionnelles de type GGA, en traitant une partie des interactions d'échange par l'opérateur d'échange Hartree-Fock appliqué au déterminant de Slater des spin-orbitales de Kohn-Sham. La fonctionnelle d'échange-corrélation la plus utilisée est sans conteste la fonctionnelle hybride

B3LYP (*Becke 3 parameters Lee Yang Parr*) [26]. Elle propose une amélioration de la fonctionnelle BLYP en incluant une fonctionnelle d'échange E_x^{HF} de type Hartree-Fock, tel que :

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a(E_x^{HF} - E_x^{LDA}) + b\Delta E_x^{B88} + c(E_c^{LYP} - E_c^{LDA}) \quad (1.62)$$

où $a \approx 0,2$, $b \approx 0,7$ et $c \approx 0,8$ sont des paramètres qui ont été déterminés empiriquement.

1.5 L'approche semi-empirique

Les méthodes Hartree-Fock sont dites *ab initio* car elles n'ont aucun recours à des données expérimentales. Pour les méthodes *ab initio* et DFT, la taille des systèmes étudiés ne peut excéder quelques dizaines d'atomes. En revanche, des systèmes de plusieurs centaines d'atomes peuvent être traités couramment par les méthodes semi-empiriques qui pour leur part utilisent des données expérimentales.

1.5.1 Les principes

Les méthodes semi-empiriques les plus courantes ont pour vocation de réduire le nombre d'intégrales à calculer. Elles sont fondées sur les approximations suivantes :

- Seules les électrons de valence sont traités explicitement. Leurs orbitales sont représentées par une base minimale de fonctions.
- Les intégrales de recouvrement sont négligées lors de la procédure SCF.
- Les intégrales bi-électroniques à 3 ou 4 centres sont supposées nulles ; la prise en considération des intégrales bi-électroniques à 1 ou 2 centres dépend de la méthode utilisée.

Ces approximations sont compensées par l'utilisation de données expérimentales :

- Les éléments non diagonaux de la matrice \underline{H}^{core} sont estimés au moyen de relations empiriques qui considèrent que ces termes sont proportionnels à l'intégrale de recouvrement des orbitales atomiques concernées.
- Les intégrales mono et bi-électroniques à 1 centre sont estimées à partir de données spectroscopiques d'atomes et d'ions, en particulier à partir de potentiels d'ionisation et d'affinités électroniques.
- Les intégrales bi-électroniques à 2 centres sont paramétrisées afin que les méthodes reproduisent au mieux les données expérimentales ou tirées de calculs *ab initio* de haute précision obtenues pour un grand nombre de composés.

La réduction du nombre d'intégrales bi-électroniques à calculer transforme l'ordre du temps de calcul de la matrice de Fock en L^2 (carré du nombre de fonctions de la base minimale). Cependant, le temps d'un calcul semi-empirique est proportionnel à L^3 en conséquence de l'étape de diagonalisation de la matrice de Fock.

1.5.2 La matrice densité

Lors de la procédure SCF, l'approximation $\underline{S} = \underline{I}$ simplifie considérablement la résolution des équations de Roothaan-Hall car elle permet d'accéder directement à l'équation aux valeurs propres (1.47). Ainsi, la matrice densité déterminée lors d'un calcul semi-empirique est définie par :

$$P'_{\mu\nu} = \sum_{i=1}^L n_i C'_{\mu i} C'_{i\nu}{}^\dagger \quad (1.63)$$

Afin de comprendre la nature de cette matrice densité, nous utilisons la transformation symétrique à partir d'un élément de la matrice densité exacte :

$$\begin{aligned} P_{\mu\nu} &= \sum_{i=1}^L n_i C_{\mu i} C_{i\nu}{}^\dagger \\ &= \sum_{i=1}^L n_i \left(\sum_{a=1}^L X_{\mu a} C'_{a i} \right) \left(\sum_{b=1}^L C'_{i b}{}^\dagger X_{b\nu}^\dagger \right) \\ &= \sum_{a=1}^L \sum_{b=1}^L X_{\mu a} \left(\sum_{i=1}^L n_i C'_{a i} C'_{i b}{}^\dagger \right) X_{b\nu}^\dagger \\ &= \sum_{a=1}^L \sum_{b=1}^L X_{\mu a} P'_{ab} X_{b\nu}^\dagger \end{aligned} \quad (1.64)$$

Au final, la matrice densité exacte peut être retrouvée par transformation symétrique de la matrice densité approchée⁹ :

$$\underline{P} = \underline{X} \underline{P}' \underline{X}^\dagger \quad (1.65)$$

1.5.3 Les méthodes

Afin d'améliorer les approximations considérées, une série de méthodes semi-empiriques a progressivement été développée depuis un demi-siècle. Pendant cette thèse, cinq d'entre elles parmi les plus évoluées ont été utilisées :

⁹Il doit toutefois être gardé à l'esprit que la matrice densité approchée d'un calcul semi-empirique décrit uniquement les électrons de valence du système.

MNDO (*Modified Neglect of Diatomic Overlap*) [27]

AM1 (*Austin Model 1*) [28]

RM1 (*Reparameterized Model 1*) [29]

PM3 (*Parametric Method 3*) [30]

PDDG/PM3 (*Pairwise Distance Directed Gaussian*) [31]

Ces méthodes reposent sur le même type d'approximations, mais diffèrent principalement par leur paramétrisation. Elles fournissent généralement des résultats fiables pour des composés organiques, notamment pour des systèmes conjugués. Cependant, leur capacité à traiter des atomes diffus comme le soufre ou le phosphore est généralement critique. Par ailleurs, la paramétrisation et l'application des métaux demeurent très difficiles.

Lors d'applications, il est préférable d'utiliser plusieurs méthodes semi-empiriques afin de vérifier la reproductibilité des résultats. Pour ce faire, AM1 et PDDG constituent des méthodes de référence utilisant chacune des modes de paramétrisation différents. Enfin, il est important de disposer de données expérimentales ou de calculs quantiques plus précis afin de juger de la validité des résultats.

Chapitre 2

Les charges partielles

Les charges partielles permettent de représenter de façon simple la répartition de la densité électronique dans un système moléculaire. Elles sont fréquemment appliquées à la loi de Coulomb pour calculer les interactions électrostatiques. Par ailleurs, les charges partielles sont des indices importants pour la compréhension de différences structurales ou de mécanismes réactionnels. De façon générale, l'obtention de charges partielles reportant fidèlement l'état réel des atomes dans une molécule est ainsi un enjeu important. Dans le cadre de cette thèse, une détermination précise de la répartition des charges est une étape cruciale dans notre traitement des interactions électrostatiques (cf chap.3).

2.1 Les méthodes de calcul existantes

Le concept de charges partielles demeure artificiel car elles ne sont pas des observables physiques. De ce fait, différentes définitions existent sans qu'aucune ne soit réellement plus légitime qu'une autre [32, 33]. Du moins, la méthode de calcul doit être choisie en fonction des propriétés que l'on souhaite analyser. Il est même utile d'essayer plusieurs méthodes afin d'avoir un recul critique sur les charges obtenues. Les principales méthodes de calcul sont présentées ci-dessous.

2.1.1 Les analyses de population

Les méthodes de chimie quantique sont généralement utilisées dans le formalisme des jeux de fonctions de base centrées sur les atomes (cf § 1.2.5). Une approche immédiate pour le calcul de charges partielles est par conséquent l'analyse des populations électroniques constituées par ces fonctions de base.

Les charges de Mulliken

En intégrant dans l'espace l'expression (1.39) de la densité électronique, nous retrouvons le nombre total d'électrons :

$$N = \sum_{\mu=1}^L \sum_{\nu=1}^L P_{\mu\nu} S_{\nu\mu} \equiv \sum_{\mu=1}^L (\underline{P}\underline{S})_{\mu\mu} \quad (2.1)$$

Il est possible de décomposer la population électronique en 2 termes, tel que :

$$N = \sum_{\mu=1}^L P_{\mu\mu} S_{\mu\mu} + \sum_{\mu=1}^L \sum_{\nu \neq \mu}^L P_{\mu\nu} S_{\nu\mu} \quad (2.2)$$

où le premier terme correspond à la somme des populations nettes de chaque fonction et le second à celle des populations de recouvrement entre fonctions de base.

A partir de cette décomposition de la population électronique, Robert Mulliken proposa une définition pour les charges partielles atomiques [34] :

$$q_A = Z_A - \sum_{\mu \in A}^L P_{\mu\mu} S_{\mu\mu} - \sum_{\mu \in A}^L \sum_{\nu \neq \mu}^L P_{\mu\nu} S_{\nu\mu} \quad (2.3)$$

Cette définition revient à partager entre deux atomes une partie égale (la moitié) de la population de recouvrement de deux fonctions de base leur appartenant.

Nous pouvons écrire finalement les charges de Mulliken comme :

$$q_A = Z_A - \sum_{\mu \in A}^L \sum_{\nu=1}^L P_{\mu\nu} S_{\nu\mu} \equiv Z_A - \sum_{\mu \in A}^L (\underline{P}\underline{S})_{\mu\mu} \quad (2.4)$$

L'analyse de Mulliken est simple, rapide et donne une représentation correcte des charges partielles avec des jeux de fonctions de base de taille petite ou moyenne. Mais l'utilisation d'orbitales diffuses donne lieu à des variations imprévisibles des charges partielles. En effet, une fonction diffuse décrit principalement¹ la densité électronique sur les atomes voisins à celui sur lequel elle est centrée et donc attribuée par l'analyse de Mulliken. Ainsi, l'extension des jeux de fonctions de base ne mène pas à une convergence des charges de Mulliken.

¹La densité radiale de l'orbitale atomique est maximale sur les atomes voisins.

Cette analyse de population permet également de calculer un autre concept théorique d'importance, l'ordre de liaison [35] :

$$B_{AB} = \sum_{\mu \in A}^L \sum_{\nu \in B}^L (\underline{P} \underline{S})_{\mu\nu} (\underline{P} \underline{S})_{\nu\mu} \quad (2.5)$$

Les charges de Löwdin

Une autre définition de charges partielles a été proposée par Per-Olov Löwdin [36]. Il utilisa judicieusement la propriété suivante sur les traces de matrices afin de reformuler la relation (2.1), tel que :

$$N = tr(\underline{P} \underline{S}) = tr(\underline{S}^{\frac{1}{2}} \underline{P} \underline{S}^{\frac{1}{2}}) \quad (2.6)$$

Par cette propriété, les charges de Löwdin sont définies comme :

$$q_A = Z_A - \sum_{\mu \in A}^L (\underline{S}^{\frac{1}{2}} \underline{P} \underline{S}^{\frac{1}{2}})_{\mu\mu} \equiv Z_A - \sum_{\mu \in A}^L P'_{\mu\mu} \quad (2.7)$$

En comparaison avec l'analyse de Mulliken, celle de Löwdin n'effectue aucun partage de population de recouvrement mais conserve le problème rencontré avec des orbitales diffuses.

L'ordre de liaison devient dans l'analyse de Löwdin :

$$B_{AB} = \sum_{\mu \in A}^L \sum_{\nu \in B}^L P'_{\mu\nu} P'_{\nu\mu} \quad (2.8)$$

Les charges NAO

L'analyse des NAO (*Natural Atomic Orbitals*) et des NBO (*Natural Bond Orbitals*) est une procédure de diagonalisation de la matrice densité [37]. Cette méthode permet de redistribuer la densité électronique dans des orbitales naturelles atomiques et moléculaires. Des charges atomiques peuvent alors être dérivées de ces orbitales naturelles. Bien que moins dépendante aux jeux de fonctions de base que les analyses de Mulliken et Löwdin, cette méthode nécessite un temps de calcul plus important.

2.1.2 Les fragmentations de l'espace

Le concept de charges partielles revient à partager entre les atomes la densité électronique de la molécule. Ainsi, une possibilité de calcul des charges est de partager l'espace de la molécule en régions atomiques.

Les charges AIM

La méthode de Bader ou AIM (*Atoms In Molecules*) considère la densité électronique comme une somme des densités appartenant aux atomes [38]. Le volume propre de chaque atome dans la molécule est délimité par une surface sur laquelle le flux de la densité est nulle, tel que :

$$\nabla\rho(\vec{r}) \cdot \vec{n} = 0 \quad (2.9)$$

où \vec{n} est le vecteur normal à la surface.

Cette définition fournit une analyse topologique intéressante. L'évaluation des charges AIM s'avère peu dépendante au jeu de fonctions de base utilisé. Cependant, la méthode de Bader est très coûteuse et accentue généralement la polarisation des liaisons.

Les charges de Hirshfeld

La méthode de Hirshfeld ou *stockholder* (actionnaires) partage la densité électronique de la molécule en fonction de la densité atomique initiale de chaque atome [39]. Cette démarche permet de calculer des charges partielles atomiques :

$$q_A = Z_A - \iiint \frac{\rho_A^{at}(\vec{r})}{\sum_{I=1}^M \rho_I^{at}(\vec{r})} \rho(\vec{r}) d\vec{r} \quad (2.10)$$

où $\rho_I^{at}(\vec{r})$ est la densité électronique de l'atome I isolé à la même position que dans la molécule.

La forme analytique complexe du facteur de pondération nécessite une résolution numérique (cf § 2.2.2). Cette méthode présente une faible dépendance aux jeux de fonctions de base, mais attribue généralement aux liaisons un caractère trop covalent [40].

Il existe des approches voisines à celle d'Hirshfeld, telle que la méthode VDD (*Voronoi Deformation Density*) [41]. La méthode des polyèdres de Voronoi est un partage géométrique de l'espace de la molécule. Les droites reliant un atome aux autres sont coupées en leur centre par un plan perpendiculaire, formant un polyèdre de Voronoi qui définit le volume propre de l'atome [42, 43]. La méthode des atomes de Stewart est un ajustement de densités atomiques sphériques par rapport à la densité

électronique de la molécule [44]. Les charges partielles de ces approches sont difficiles à obtenir et généralement de valeur trop grandes.

2.1.3 La reproduction des observables

Le calcul des charges partielles a souvent pour but de caractériser les interactions électrostatiques du système avec son environnement. Il est donc judicieux de définir les charges partielles par rapport à des observables telles que le moment dipolaire ou le potentiel électrostatique.

Les charges GAPT

La méthode GAPT (*Generalized atomic polar tensor*) définit les charges partielles à partir d'un tenseur atomique caractéristique [45] :

$$\underline{\underline{V}}^{APT} = \begin{pmatrix} \frac{\partial \mu_x}{\partial x_A} & \frac{\partial \mu_x}{\partial y_A} & \frac{\partial \mu_x}{\partial z_A} \\ \frac{\partial \mu_y}{\partial x_A} & \frac{\partial \mu_y}{\partial y_A} & \frac{\partial \mu_y}{\partial z_A} \\ \frac{\partial \mu_z}{\partial x_A} & \frac{\partial \mu_z}{\partial y_A} & \frac{\partial \mu_z}{\partial z_A} \end{pmatrix} \quad (2.11)$$

où μ_x, μ_y, μ_z sont les composantes du moment dipolaire de la molécule, tandis que x_A, y_A, z_A sont les coordonnées cartésiennes de l'atome A .

Les charges GAPT sont définies comme le tiers de la trace du tenseur polaire atomique, tel que :

$$q_A^{GAPT} = \frac{1}{3} \left(\frac{\partial \mu_x}{\partial x_A} + \frac{\partial \mu_y}{\partial y_A} + \frac{\partial \mu_z}{\partial z_A} \right) \quad (2.12)$$

Cette méthode est particulièrement coûteuse du fait qu'elle nécessite le calcul de la dérivée seconde de l'énergie. Par conséquent, elle n'est utilisée généralement qu'après un calcul de fréquences de vibration. Tout comme le moment dipolaire, les charges GAPT sont dépendantes du niveau de calcul de la fonction d'onde. Elles fournissent des résultats précis lorsque la corrélation électronique est suffisamment bien traitée dans le calcul de la densité électronique.

Les charges ESP

La méthode ESP (*ElectroStatic Potential*) [46, 47, 48] permet d'obtenir un jeu de charges partielles qui reproduit au mieux le potentiel électrostatique calculé sur une surface de Connolly² de la molécule. Une fonction d'ajustement est appliquée aux

²Une surface de Connolly est tissée sur les sphères de Van der Waals des atomes de la molécule.

points \vec{r}_g d'une grille placée sur la surface moléculaire :

$$\mathcal{F} = \sum_{g=1}^{points} (V^{obs}(\vec{r}_g) - V^{fit}(\vec{r}_g))^2 \quad (2.13)$$

L'observable potentiel électrostatique V^{obs} est de forme analytique :

$$V^{obs}(\vec{r}_g) = \sum_{A=1}^M \frac{Z_A}{|\vec{r}_g - \vec{R}_A|} - \iiint \frac{\rho(\vec{r})}{|\vec{r}_g - \vec{r}|} d\vec{r} \quad (2.14)$$

Le potentiel électrostatique ajusté V^{fit} est de forme analytique :

$$V^{fit}(\vec{r}_g) = \sum_{A=1}^M \frac{q_A}{|\vec{r}_g - \vec{R}_A|} \quad (2.15)$$

Les charges partielles sont optimales lorsqu'elles conduisent la fonction d'ajustement à un minimum, tel que :

$$\frac{\partial \mathcal{F}}{\partial q_A} = -2 \sum_{g=1}^{points} \frac{V^{obs}(\vec{r}_g) - V^{fit}(\vec{r}_g)}{|\vec{r}_g - \vec{R}_A|} = 0 \quad (2.16)$$

En outre, il est nécessaire d'imposer à cette minimisation une contrainte sur la somme des charges partielles par la méthode des multiplicateurs de Lagrange.

La méthode ESP donnent des charges partielles de référence pour de petites molécules [49]. Dans le cas de grands systèmes moléculaires, les charges ESP sont convenables pour les atomes proches de la surface de Connolly mais sont mal évaluées pour les atomes "enfouis" dans le système. Par ailleurs, les charges ESP sont très sensibles à de faibles changements de conformation.

Les charges RESP

La méthode RESP (*Restrained ElectroStatic Potential*) [50] utilise une fonction de pénalité qui permet d'éviter les charges souvent trop élevées pour les atomes enfouis. D'autres contraintes sont imposées afin de permettre la transférabilité des résultats d'un système à un autre. La molécule est divisée en fragments dont la charge totale de chacun est maintenue neutre. Par ailleurs, la symétrie est prise en compte en égalisant les charges des hydrogènes d'un méthyle. Les contraintes RESP réduisent également la sensibilité aux changements conformationnels. Une solution alternative mais plus coûteuse est l'ajustement des charges sur le potentiel électrostatique calculé pour plusieurs conformations [51].

2.1.4 Les méthodes paramétrées

Bien que les charges ESP soient appropriées pour traiter de petites molécules, leur temps de calcul est non négligeable. Pour évaluer les charges d'une série de ligands ou de molécules organiques, il est parfois préférable d'avoir recours à des méthodes paramétrées plus rapides.

Les charges basées sur l'électronégativité

Beaucoup de méthodes ont été proposées pour calculer les charges atomiques à partir de paramètres. Les méthodes les plus abouties sont basées sur le principe d'égalisation des électronégativités [52]. Cette approche a principalement été développée par Johann Gasteiger [53], William Goddard III [54] et Patrick Bultinck [55, 56]. Elle constitue un des moyens les plus efficaces pour définir les charges atomiques des molécules étudiées en *drug design*.

Les charges CM

La méthode CM (*Charge Model*) développée par Christopher Cramer et Donald Truhlar permet une correction des charges partielles de Löwdin par le biais d'une fonction paramétrée. La fonction de correction de la charge atomique dépend des ordres de liaison impliquant l'atome et son environnement covalent. Les paramètres de la fonction ont été initialement optimisés par rapport aux moments dipolaires expérimentaux d'un grand nombre de molécules organiques.

La méthode CM permet d'obtenir des charges partielles de haute qualité pour un temps de calcul peu élevé. Néanmoins, elle nécessite une paramétrisation spécifique pour les différents types de liaison covalente du modèle et pour chaque méthode de chimie quantique associée à un jeu de fonctions de base donné. Ce modèle de charge a été paramétré pour certaines méthodes semi-empiriques [57], puis certaines méthodes *ab initio* et DFT [58].

Les charges RLPA

La méthode RLPA (*Redistributed Löwdin Population Analysis*) [59] utilise une approche similaire au modèle CM afin de corriger les charges de Löwdin obtenues avec une base 6-31+G*. Les populations venant des fonctions diffuses sont redistribuées

par une fonction paramétrée, ce qui permet de corriger l'incompatibilité de l'analyse de Löwdin avec l'utilisation d'orbitales diffuses (cf § 2.1.1).

Les charges AM1-BCC

La méthode AM1-BCC (*Bond Charge Corrections*) est une correction des charges de Löwdin-AM1 par le biais d'une fonction paramétrée caractérisant la topologie de la molécule [60, 61]. Les paramètres de corrections AM1-BCC des divers types de liaisons covalentes ont été initialement optimisés sur une série d'environ 2700 molécules afin de reproduire leur potentiel électrostatique calculé au niveau HF/6-31G*. La simplicité et la paramétrisation *ab initio* du modèle AM1-BCC permet de calculer très rapidement des charges de haute qualité pour des ligands organiques.

2.2 Les méthodes de partition

Notre objectif est d'utiliser le concept de charges partielles pour calculer les interactions électrostatiques mises en jeu dans les systèmes biologiques. L'environnement d'un atome pouvant changer brutalement au cours d'une réaction, nous souhaitons une méthode qui n'impose pas de topologie donnée au système. De plus, nous envisageons une approche standard pouvant être appliquée à toute méthode de chimie quantique quelque soit le niveau de calcul. Pour ces différentes raisons, les méthodes paramétrées sont inadaptées à notre objectif.

Par ailleurs, la forme analytique des fonctions utilisées doit être facilement dérivable et l'évolution des charges doit être numériquement stable par rapport à de faibles changements de conformation. La méthode doit de même être capable de définir un système de taille importante. Ainsi, les méthodes de type ESP sont elles aussi inadaptées à notre projet. Les méthodes de type numériques pourrait être envisagées, bien que leur forme analytique soit souvent trop compliquée pour assurer une dérivation facile et un calcul rapide.

De façon générale, nous utilisons des fonctions de base pour le calcul de la densité électronique. Les méthodes de Mulliken et Löwdin répondent ainsi à nos critères de sélection. La simplicité de ces méthodes étant toutefois critiquable, nous avons tenté d'améliorer ces analyses de populations reposant sur les fonctions de base.

2.2.1 L'utilisation des primitives

Afin d'améliorer le partage de la densité électronique, il est judicieux de faire une analyse de population basée sur les primitives. Le changement de base (1.23) mène à reformuler la population électronique (2.1), tel que :

$$N = \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{g=1}^K \sum_{h=1}^K P_{\mu\nu} D_{\nu g}^\dagger S'_{gh} D_{h\mu} \equiv \sum_{\mu=1}^L (\underline{P} \underline{D}^\dagger \underline{S}' \underline{D})_{\mu\mu} \quad (2.17)$$

où \underline{S}' est la matrice de recouvrement des fonctions primitives et \underline{D} la matrice des coefficients de contraction de gaussiennes.

La forme gaussienne des primitives permet de calculer facilement les éléments de la matrice de recouvrement \underline{S}' . Prenons par exemple une primitive qui représente une orbitale atomique de type 1s :

$$\eta_{1s}^{GTO}(\alpha, \vec{r} - \vec{R}_A) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha|\vec{r}-\vec{R}_A|^2} \quad (2.18)$$

où l'exposant α détermine la largeur de la fonction gaussienne centrée en \vec{R}_A .

Le produit de deux primitives de type 1s forme une troisième gaussienne de même type (voir figure 2.1), tel que :

$$\eta_{1s}^{GTO}(\alpha, \vec{r} - \vec{R}_A) \eta_{1s}^{GTO}(\beta, \vec{r} - \vec{R}_B) = K_{AB} \eta_{1s}^{GTO}(p, \vec{r} - \vec{R}_P) \quad (2.19)$$

$$\text{avec} \quad \begin{cases} p & = \alpha + \beta \\ \vec{R}_P & = \frac{\alpha\vec{R}_A + \beta\vec{R}_B}{\alpha + \beta} \\ K_{AB} & = \left(\frac{2\alpha\beta}{(\alpha + \beta)\pi}\right)^{3/4} e^{-\frac{\alpha\beta}{\alpha + \beta}|\vec{R}_A - \vec{R}_B|^2} \end{cases}$$

L'analyse de Mulliken attribue une part égale de la population de recouvrement aux deux atomes qui l'ont créée. Ce partage ne prend pas en compte la position du recouvrement (centrée sur P) par rapport aux deux atomes (A et B). Or, un calcul standard de chimie quantique crée des recouvrements distribués tout le long de la distance interatomique. D'autre part, seuls les atomes impliqués dans la création de la population de recouvrement se voient attribuer une charge. Là encore, la position

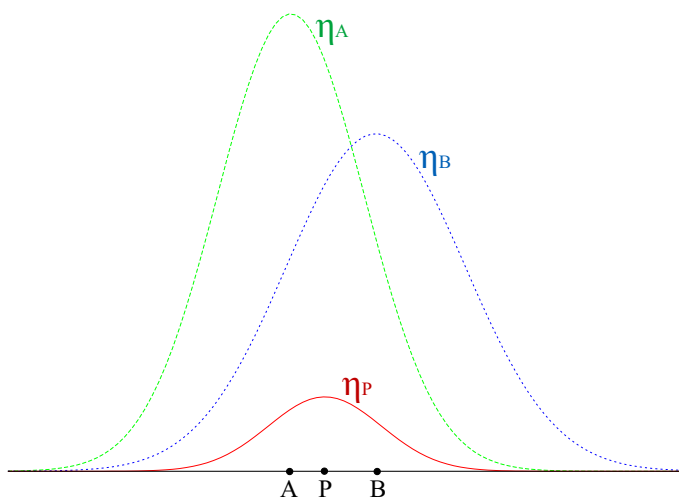


FIG. 2.1: Recouvrement η_P de 2 primitives η_A et η_B de type 1s.

du recouvrement n'est pas considérée par rapport à celle des autres atomes de la molécule, bien qu'elle puisse être centrée sur l'un d'entre eux. Prenons l'exemple du dioxyde de carbone : certaines populations de recouvrement de fonctions de base des 2 oxygènes sont centrées sur le carbone central, bien qu'elles soient finalement attribuées aux 2 oxygènes.

Nous proposons ici une analyse de population qui rectifie les lacunes de la méthode de Mulliken. Pour toutes les populations électroniques formées dans la base des primitives, un partage réaliste est défini entre tous les atomes du système tel que :

$$S'_{gh} = \sum_{A=1}^M w_A^{gh} S'_{gh} \quad , \quad \text{avec} \quad \sum_{A=1}^M w_A^{gh} = 1 \quad (2.20)$$

où w_A^{gh} est le poids attribué à l'atome A dans le partage de la population du recouvrement.

Une charge partielle atomique est alors définie comme :

$$q_A = Z_A - \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{g=1}^K \sum_{h=1}^K P_{\mu\nu} D_{\nu g}^\dagger \left(w_A^{gh} S'_{gh} \right) D_{h\mu} \quad (2.21)$$

Les coefficients de contraction de gaussiennes et les éléments de la matrice densité n'interviennent que dans la multiplication de l'intégrale de recouvrement afin de déterminer la population électronique associée. Chaque intégrale de recouvrement S'_{gh}

est l'objet d'une partition atomique. Les indices g et h représentent les deux primitives à l'origine du recouvrement, à partir desquelles il est possible de retrouver la position du recouvrement et celles des atomes qui l'ont créé. Pour le calcul des facteurs de partition atomique w_A^{gh} , nous nous sommes inspirés des techniques de quadrature numérique utilisées en DFT.

2.2.2 La quadrature numérique

Si la forme analytique d'une propriété est connue, nous pouvons déterminer sa valeur totale par l'intégration suivante :

$$I = \iiint F(\vec{r}) d\vec{r} \quad (2.22)$$

où $F(\vec{r})$ est la valeur de la propriété en un point \vec{r} donné.

Le calcul de l'intégrale I peut être approximé par discrétisation, tel que :

$$I \approx \sum_{g=1}^G v(\vec{r}_g) F(\vec{r}_g) \quad (2.23)$$

où G est le nombre de points de la grille et $v(\vec{r}_g)$ le poids d'intégration numérique d'un point \vec{r}_g de la grille.

En DFT, la forme complexe des fonctionnelles d'échange-corrélation (cf § 1.4.3) ne permet pas une résolution analytique et des techniques numériques doivent être utilisées. Un schéma général de quadrature numérique a été proposé par Axel Becke pour le calcul des propriétés moléculaires [62]. L'intégration numérique s'effectue avec des grilles centrées sur chaque atome de la molécule, tel que :

$$I \approx \sum_{A=1}^M \sum_{g_A=1}^{G_A} v(\vec{r}_{g_A}) w_A(\vec{r}_{g_A}) F(\vec{r}_{g_A}) \quad (2.24)$$

où $w_A(\vec{r}_{g_A})$ est un facteur de partition³ attribué à l'atome A au point \vec{r}_{g_A} d'une grille centrée sur son noyau.

2.2.3 Les schémas de partition

Les facteurs de partition de l'intégration numérique (2.24) attribuent le degré d'appartenance d'un point de la molécule à un atome donné. Un point à proximité

³Cette partition atomique facilite le calcul numérique en transformant une intégration à plusieurs centres en une somme d'intégrations à un seul centre de symétrie sphérique.

d'un noyau sera totalement attribué à l'atome correspondant. Au delà d'une certaine distance du noyau, le facteur de partition atomique sera nul. Le problème principal est de déterminer une partition correcte pour des positions internucléaires. Il convient pour cela de concevoir la densité électronique de la molécule comme un ensemble de densités électroniques atomiques qui se recouvrent. Cette démarche corrobore à la notre et c'est pourquoi nous avons choisi d'appliquer les facteurs de partition de quadrature numérique à notre analyse de population.

Le schéma original

Dans son schéma de partition, Axel Becke proposa d'utiliser les coordonnées elliptiques (λ, μ, θ) définies ci-dessous par les relations (2.25) et la figure 2.2.

$$\lambda_{AB} = \frac{|\vec{R}_A - \vec{R}_P| + |\vec{R}_B - \vec{R}_P|}{|\vec{R}_A - \vec{R}_B|} \quad \mu_{AB} = \frac{|\vec{R}_A - \vec{R}_P| - |\vec{R}_B - \vec{R}_P|}{|\vec{R}_A - \vec{R}_B|} \quad (2.25)$$

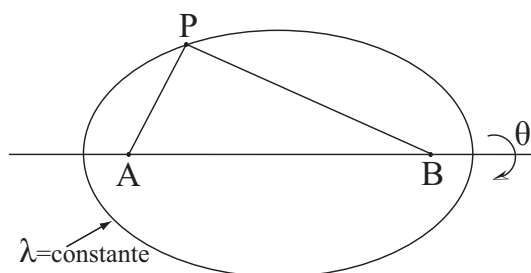


FIG. 2.2: Définition des coordonnées elliptiques.

Une fonction de partition est définie entre 2 atomes par :

$$s(\mu_{AB}) = \frac{1}{2} [1 - f_k(\mu_{AB})] \quad (2.26)$$

La fonction $f_k(\mu_{AB})$ est une fonction composée où k est l'ordre d'itération⁴, tel que :

$$f_3(\mu_{AB}) = p \{p [p(\mu_{AB})]\} \quad (2.27)$$

Le polynôme de Becke est de forme :

$$p(\mu_{AB}) = \frac{3}{2} \mu_{AB} - \frac{1}{2} \mu_{AB}^3 \quad (2.28)$$

⁴L'ordre d'itération 3 est reconnu optimal pour la quadrature numérique DFT.

Afin de satisfaire la condition (2.20), le facteur de partition atomique normalisé issu du schéma de Becke est défini comme :

$$w_A^{gh} = \frac{\prod_{B \neq A}^M s(\mu_{AB})}{\sum_{C=1}^M \prod_{D \neq C}^M s(\mu_{CD})} \quad (2.29)$$

L'extension du schéma original

Nous avons également développé un schéma où les fonctions de Becke sont appliquées directement à l'analyse de population de Mulliken. La partition multiatomique du schéma original est remplacée par une partition diatomique entre les atomes à l'origine de la population de recouvrement. Les facteurs de partition atomique deviennent alors :

$$w_A^{gh} = s(\mu_{AB}) \quad \text{et} \quad w_B^{gh} = 1 - s(\mu_{AB}) \quad (2.30)$$

où A et B sont les atomes à l'origine de la population du recouvrement S'_{gh} .

Dans l'analyse de Mulliken, les facteurs w_A^{gh} et w_B^{gh} sont fixés à $1/2$. Avec la fonction de partition, nous prenons en considération la position du recouvrement par rapport à tous les atomes dans le schéma multiatomique, et par rapport aux atomes A et B dans le schéma diatomique. Pour les deux schémas, nous avons itéré de 1 à 6 fois le polynôme de Becke afin de tester toutes les fonctions de partition envisageables.

La partition linéaire

Nous avons par ailleurs développé une méthode que l'on nommera "linéaire". Dans la fonction de partition de cette méthode, les polynômes de Becke sont simplement remplacés par la valeur de la coordonnée elliptique, tel que :

$$s(\mu_{AB}) = \frac{1}{2} [1 - \mu_{AB}] \quad (2.31)$$

Nous avons uniquement appliqué la fonction au schéma diatomique afin d'accéder à une méthode de calcul efficace. En effet, la fonction de partition (2.31) peut facilement être remplacée dans ce cas par les coordonnées cartésiennes. Les facteurs de partition atomique deviennent alors :

$$w_A^{gh} = \frac{|\vec{R}_B - \vec{R}_P|}{|\vec{R}_A - \vec{R}_B|} \quad \text{et} \quad w_B^{gh} = 1 - \frac{|\vec{R}_B - \vec{R}_P|}{|\vec{R}_A - \vec{R}_B|} \quad (2.32)$$

La figure 2.3 permet de comparer le profil de la fonction linéaire (2.31) avec les fonctions de partition de Becke (2.26).

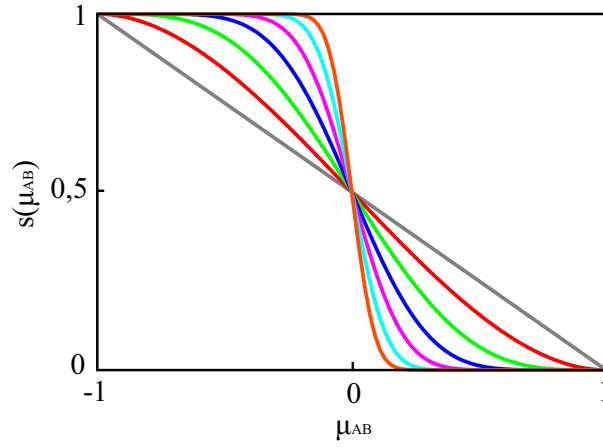


FIG. 2.3: Profils des fonctions de partition linéaire (gris) et de Becke d'ordres d'itération de 1 (rouge) à 6 (orange).

Le traitement hétéronucléaire

Dans son schéma de quadrature numérique, Axel Becke propose également un traitement spécifique pour les systèmes hétéronucléaires. Il utilise les rayons de Bragg-Slater qui constituent une définition empirique des rayons covalents atomiques [63, 64]. Sa démarche est de remplacer dans la fonction de partition la coordonnée initiale μ_{AB} par la coordonnée ν_{AB} , tel que :

$$\nu_{AB} = \mu_{AB} + a_{AB} (1 - \mu_{AB}^2) \quad (2.33)$$

Le facteur d'échelle a_{AB} caractérise la différence de taille entre les atomes par :

$$a_{AB} = \frac{\Omega_B^2 - \Omega_A^2}{4\Omega_A\Omega_B} \quad (2.34)$$

où Ω_A et Ω_B sont les rayons de Bragg-Slater respectifs des atomes A et B .

Ce changement de coordonnée permet de modéliser la taille respective des atomes en déplaçant la fonction de partition $s(\mu_{AB})$ le long de l'axe de la coordonnée elliptique. Afin que la fonction de partition demeure monotone, il est toutefois nécessaire de poser une limite à ce déplacement :

$$|a_{AB}| \leq \frac{1}{2} \quad (2.35)$$

Cette limite concerne les atomes dont le rapport des rayons de Bragg-Slater est supérieur à 2,4.

Les fonctions de Stratmann

Afin de mettre au point des méthodes DFT à croissance linéaire, Eric Stratmann et ses collègues ont proposé une alternative⁵ aux fonctions de Becke pour le calcul numérique de l'énergie d'échange-corrélation [65]. Ils suggeraient de remplacer $f_k(\mu_{AB})$ par $g(\mu_{AB}; a)$ dans la fonction de partition (2.26), tel que :

$$g(\mu_{AB}; a) = \begin{cases} -1 & \mu_{AB} \leq -a \\ z(\mu_{AB}; a) & -a < \mu_{AB} < +a \\ +1 & \mu_{AB} \geq +a \end{cases} \quad (2.36)$$

$$\text{avec} \quad z(\mu_{AB}; a) = \frac{1}{16} \left[35 \left(\frac{\mu_{AB}}{a} \right) - 35 \left(\frac{\mu_{AB}}{a} \right)^3 + 21 \left(\frac{\mu_{AB}}{a} \right)^5 - 5 \left(\frac{\mu_{AB}}{a} \right)^7 \right]$$

où a est une valeur seuil qui détermine la grandeur d'une sphère atomique dans laquelle une position est totalement attribuée à l'atome.

Un seuil a de 0,64 était suggéré dans le papier original [65], bien que la valeur de 0,86 soit généralement retenue en pratique. Nous avons tracé le profil des fonctions de Stratmann avec ces valeurs seuils et différents ordres d'itération du polynôme $z(\mu_{AB}; a)$. Le tracé de ces fonctions se superposent aux fonctions de Becke et donneraient ainsi les mêmes résultats. Par conséquent, seules les fonctions de Becke ont été implémentées, bien que les fonctions de Stratmann offrent une perspective d'efficacité⁶ intéressante.

2.2.4 L'application aux méthodes *ab initio* et DFT

Les schémas de partition ont tout d'abord étaient appliqués aux méthodes *ab initio* et DFT, par le biais d'une analyse de multipôles.

La définition des multipôles

L'énergie d'interaction électrostatique d'une distribution de charge avec un potentiel extérieur est définie comme :

$$E_{elec} = \iiint \rho(\vec{r}) V(\vec{r}) d\vec{r} \quad (2.37)$$

⁵Nous précisons que les auteurs n'applique pas de traitement hétéroatomique.

⁶Le seuil a permet d'éviter un grand nombre de calculs à proximité des noyaux atomiques.

Le potentiel électrostatique peut se développer en série de Taylor au voisinage d'une position de référence \vec{r}_0 , tel que :

$$V(\vec{r}) = V(\vec{r}_0) + (\vec{r} - \vec{r}_0) \left(\frac{\partial V(\vec{r}_0)}{\partial \vec{r}} \right) + \frac{1}{2} (\vec{r} - \vec{r}_0)^2 \left(\frac{\partial^2 V(\vec{r}_0)}{\partial \vec{r}^2} \right) + \dots \quad (2.38)$$

L'inclusion du développement (2.38) dans la définition (2.37) fait apparaître la notion de multipôles :

$$E_{elec} = \underbrace{\iiint \rho(\vec{r}) d\vec{r}}_q V(\vec{r}_0) + \underbrace{\iiint \rho(\vec{r}) (\vec{r} - \vec{r}_0) d\vec{r}}_{\underline{\mu}} \left(\frac{\partial V(\vec{r}_0)}{\partial \vec{r}} \right) + \frac{1}{2} \underbrace{\iiint \rho(\vec{r}) (\vec{r} - \vec{r}_0)^2 d\vec{r}}_{\underline{\underline{\Theta}}} \left(\frac{\partial^2 V(\vec{r}_0)}{\partial \vec{r}^2} \right) + \dots \quad (2.39)$$

où q , $\underline{\mu}$ et $\underline{\underline{\Theta}}$ sont les multipôles d'ordre 0 (charge), 1 (dipôle) et 2 (quadrupôle) respectivement.

L'analyse des multipôles distribués

Les multipôles permettent de définir la symétrie d'une distribution de charge soumise ou non à un potentiel extérieur. Plus le développement multipolaire est important, plus la distribution de charge est clairement définie et plus son énergie d'interaction avec un potentiel extérieur est précise. Une méthode simple pour obtenir une telle description est l'analyse des multipôles distribués [66]. Cette analyse est spécifiquement destinée à une fonction d'onde calculée à partir de fonctions de base gaussiennes. Les propriétés analytiques des primitives sont en fait utilisées pour définir les multipôles créées aux points P de recouvrement de celles-ci (voir figure 2.1). L'étendu du développement multipolaire dépend de la symétrie des primitives. Par exemple, si deux primitives de type p (soit 2 dipôles) se recouvrent, la population de recouvrement est définie par une charge, un dipôle et un quadrupôle.

Anthony Stone a développé sa méthode dans un programme appelé GDMA (*Gaussian Distributed Multipoles Analysis*). L'analyse nécessite la matrice densité $\underline{\underline{P}}$ et la définition du jeu de fonctions de base d'un calcul effectué par le programme de chimie quantique Gaussian [67]. GDMA permet d'effectuer l'analyse pour un ensemble de

points choisis. L'analyse des multipôles est faite d'abord aux points P de recouvrement des primitives. Les multipôles d'un point P sont ensuite attribués au point le plus proche parmi ceux ayant été choisis. Notre utilisation du programme étant destiné au partage des populations de recouvrement des primitives, nous avons sélectionné naturellement l'ensemble des points P. Par ailleurs, seuls les multipôles d'ordres 0 nous intéressaient pour nos méthodes de partition de la densité électronique. En complément des calculs Gaussian et GDMA, nous avons implémenté en Python l'ensemble de la procédure pour le calcul des charges partielles.

2.2.5 L'application aux méthodes semi-empiriques

L'approche précédente basée sur les multipôles distribués ne peut être appliquée aux méthodes semi-empiriques, car la matrice densité calculée $\underline{\underline{P'}}$ est biaisée par l'approximation $\underline{\underline{S}}=\underline{\underline{I}}$ (cf § 1.5). De façon générale, les méthodes de partition doivent obligatoirement être appliquées à la matrice densité exacte $\underline{\underline{P}}$.

Les charges de Löwdin peuvent directement être obtenues à partir de la matrice densité du calcul, grâce à la définition standard :

$$q_A = Z'_A - \sum_{\mu \in A}^L P'_{\mu\mu} \quad (2.40)$$

où Z'_A est la charge nucléaire de l'atome A auquel on soustrait le nombre d'électrons de coeur non pris en compte dans le calcul semi-empirique de la matrice densité $\underline{\underline{P'}}$.

La détermination des charges de Mulliken nécessite le calcul des matrices de transformation symétrique, tel que :

$$q_A = Z'_A - \sum_{\mu \in A}^L \sum_{x=1}^L \sum_{y=1}^L X_{\mu x} P'_{xy} X_{y\mu}^{-1} \equiv Z'_A - \sum_{\mu \in A}^L (\underline{\underline{X}} \underline{\underline{P'}} \underline{\underline{X}}^{-1})_{\mu\mu} \quad (2.41)$$

Pour les méthodes de partition, une charge partielle atomique est définie comme :

$$q_A = Z'_A - \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{x=1}^L \sum_{y=1}^L \sum_{g=1}^K \sum_{h=1}^K X_{\mu x} P'_{xy} X_{y\nu}^\dagger D_{\nu g}^\dagger \left(w_A^{gh} S'_{gh} \right) D_{h\mu} \quad (2.42)$$

Ces méthodes de calcul de charges partielles pour des fonctions d'onde semi-empiriques ont été implémentées en Fortran90 dans la librairie fDynamo [68]. Ce

programme développé au laboratoire est destiné à l'étude des mécanismes réactionnels dans les protéines.

Plus généralement, les développements matriciels présentés dans ce manuscrit permettent de définir les méthodes de façon théorique. En pratique, ces développements doivent être calculés de façon fragmentée afin d'assurer l'efficacité des algorithmes. Par ailleurs, la quantité d'information stockée pendant le calcul doit être limitée. Bien que ce type de précautions soit toujours respecté, il est essentiel de tester la qualité des résultats avant toute optimisation de la programmation. Ainsi, les tests ci-dessous jugent de la précision des méthodes sans préoccupation d'efficacité.

2.3 Résultats et discussion

Afin de mettre à l'épreuve la validité de nos méthodes de partition, nous avons effectué un ensemble de tests sur des molécules de petite taille.

Pour des raisons de clarté, nous adoptons la nomenclature suivante :

- DEN : observable calculée à partir de la matrice densité.
- ESP : ajustement des charges par rapport au potentiel électrostatique.
- MLK et LWD : analyse des populations de Mulliken et Löwdin.
- LNR : méthode de partition linéaire.
- DHB1 à DHB6 : di(D)-homo(H)atomique partition de Becke ($k=1$ à 6).
- DXB1 à DXB6 : di(D)-hétéro(X)atomique partition de Becke ($k=1$ à 6).
- MHB1 à MHB6 : multi(M)-homo(H)atomique partition de Becke ($k=1$ à 6).
- MXB1 à MXB6 : multi(M)-hétéro(X)atomique partition de Becke ($k=1$ à 6).

2.3.1 Calcul de charges *ab initio*

Nous avons tout d'abord testé la dépendance des méthodes de calcul de charges partielles aux jeux de fonctions de base. La géométrie d'une molécule d'eau a initialement été optimisée au niveau HF/6-31G**. La densité électronique Hartree-Fock calculée avec différents jeux de fonctions de base a servi au calcul des charges partielles des différentes méthodes disponibles. Les valeurs de la charge de l'oxygène sont reportées dans le tableau 2.1 pour les principales méthodes.

Pour ce calcul ainsi que les suivants, le degré d'itération de la fonction de Becke n'influe que dans une moindre mesure sur les résultats. Toutefois, l'ordre d'itération 3 est jugé optimal et seuls ses résultats sont présentés ici. Les résultats des autres ordres d'itération sont donnés en Annexe A, dont le tableau A.1 pour la charge de l'oxygène de l'eau.

Base	ESP	MLK	LNR	DHB3	DXB3	MHB3	MXB3
STO-3G	-0,64	-0,38	-0,44	-0,49	-0,67	-0,46	-0,61
3-21G	-0,89	-0,74	-0,70	-0,64	-0,81	-0,61	-0,76
6-31G*	-0,82	-0,90	-0,83	-0,78	-0,95	-0,78	-0,93
6-311G**	-0,80	-0,50	-0,49	-0,48	-0,75	-0,46	-0,72
6-311++G**	-0,83	-0,51	-0,50	-0,49	-0,76	-0,48	-0,74
cc-pVTZ	-0,75	-0,48	-0,49	-0,50	-0,75	-0,47	-0,71
cc-pVQZ	-0,75	-0,52	-0,52	-0,52	-0,81	-0,49	-0,76
cc-pV5Z	-0,74	-0,57	-0,57	-0,57	-0,85	-0,57	-0,83
cc-pV6Z	-0,74	-0,40	-0,46	-0,50	-0,69	-0,53	-0,73

TAB. 2.1: Charge de l'oxygène de l'eau pour différentes méthodes à partir d'une fonction d'onde HF calculée avec différents jeux de fonctions de base.

Tout d'abord, nous constatons que seules les charges ESP convergent avec l'augmentation du nombre de fonctions de base et l'amélioration conjuguée de la densité électronique. Les charges obtenues avec des jeux de fonctions de base étendus constituent des valeurs de référence pour les petites molécules et ainsi nos différents tests. La méthode ESP a toutefois tendance à accentuer légèrement la polarité des liaisons.

Les charges de Mulliken sont pour leur part dépendantes du jeu de fonctions de base utilisé. Des charges raisonnables sont généralement obtenues pour des bases de taille moyenne telles que 3-21G ici. Pour des bases plus étendues, l'utilisation de fonctions diffuses rend imprévisible la variation des charges partielles (cf § 2.1.1). Dans le cas présent, la charge de l'oxygène passe de -0,9 avec 6-31G* à -0,5 avec 6-311G**. Ce genre de problème n'assure pas de bons résultats de la méthode avec des jeux de fonctions de base étendus.

Les différents schémas de partition homonucléaire présentent des résultats proches de ceux de l'analyse de Mulliken, malgré une sensibilité aux fonctions diffuses quelque peu réduite. Les schémas de partition hétéronucléaire donnent de meilleurs résultats. En particulier, l'augmentation du nombre de fonctions de base semble faire converger

les charges MXB3 vers les valeurs références des charges ESP. Ainsi le schéma de partition le plus complexe donne pour la molécule d'eau les résultats souhaités.

La procédure précédente a été appliquée ensuite à la molécule de formaldéhyde. Les valeurs de la charge du carbone sont reportées dans le tableau 2.2 pour les principales méthodes. Les résultats de toutes les méthodes sont reportés dans le tableau A.2.

Base	ESP	MLK	LNR	DHB3	DXB3	MHB3	MXB3
STO-3G	0,36	0,09	0,21	0,27	-0,11	0,33	-0,01
3-21G	0,49	0,14	0,28	0,33	-0,01	0,51	0,23
6-31G*	0,45	0,14	0,26	0,34	-0,10	0,53	0,15
6-311G**	0,44	0,21	0,37	0,47	0,04	0,61	0,27
6-311++G**	0,48	0,13	0,34	0,46	0,05	0,64	0,34
cc-pVTZ	0,42	0,21	0,38	0,49	0,07	0,72	0,41
cc-pVQZ	0,42	0,33	0,46	0,54	0,09	0,82	0,51
cc-pV5Z	0,42	0,41	0,54	0,62	0,13	0,92	0,58
cc-pV6Z	0,42	0,23	0,39	0,50	-0,01	0,64	0,21

TAB. 2.2: Charge du carbone du formaldéhyde pour différentes méthodes à partir d'une fonction d'onde HF calculée avec différents jeux de fonctions de base.

Ce second test sur l'atome central d'une molécule de plus grande taille met davantage à l'épreuve les considérations énumérées dans la section 2.2.1. Les charges obtenues pour le carbone montrent une nette démarcation entre chacune des méthodes de partition et l'analyse de Mulliken. Les schémas de partition di-homo-atomiques présentent une amélioration des charges de Mulliken. En particulier, les charges LNR s'approchent des valeurs ESP.

Les charges DXB3 sont proches de zéro, tandis que celles de la méthode MHB3 sont trop élevées. Les charges MXB3 présentent de grandes variations avec les changements de jeux de fonctions de base. Au final, les traitements hétéronucléaires et multiatomiques ne permettent pas d'attribuer une charge convenable à l'atome de carbone central du formaldéhyde.

Afin de s'affranchir de la contradiction des résultats obtenus pour ces deux molécules, il est nécessaire de tester les méthodes sur un grand nombre de systèmes différents en taille et composition. Les statistiques obtenues sur un grand ensemble de systèmes permettent d'obtenir une indication de valeur sur la qualité des méthodes.

2.3.2 Calcul de dipôles *ab initio*

Nous avons testé la capacité des charges partielles à reproduire le moment dipolaire d'un grand nombre de molécules. L'observable moment dipolaire peut être défini à partir de la densité électronique d'un calcul quantique, tel que :

$$\vec{\mu}_{DEN} = \sum_{A=1}^M Z_A \left(\vec{R}_A - \vec{R}_C \right) - \sum_{\mu=1}^L \sum_{\nu=1}^L P_{\mu\nu} (\nu|\hat{r} - \hat{R}_C|\mu) \quad (2.43)$$

où \vec{R}_C est le centre des charges nucléaires de la molécule.

Le moment dipolaire de la molécule peut par ailleurs être calculé à partir de charges partielles, tel que :

$$\vec{\mu}_{CHARGES} = \sum_{A=1}^M q_A \left(\vec{R}_A - \vec{R}_C \right) \quad (2.44)$$

La géométrie de 200 molécules a initialement été optimisée au niveau B3LYP/6-31G*. La liste des molécules et leur dipôle expérimental est reportée dans le tableau A.3. La densité électronique Hartree-Fock de ces structures a été calculée avec différents jeux de fonctions de base. Les charges partielles ont été déterminées à partir de cette densité. Les 200 dipôles ont ensuite été calculés pour chaque méthode et comparés aux valeurs expérimentales. Les erreurs absolues moyennes sont reportées dans le tableau 2.3 pour les principales méthodes.

Base	DEN	ESP	MLK	LNR	DHB3	DXB3	MHB3	MXB3
3-21G	0,43	0,44	1,38	1,07	0,72	0,97	0,98	1,33
6-31G*	0,35	0,37	0,98	0,84	0,75	1,18	0,83	1,41
6-311G**	0,44	0,46	1,87	0,73	0,52	0,97	2,84	3,01

TAB. 2.3: Erreur absolue moyenne dans le calcul des 200 dipôles (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir d'une fonction d'onde HF calculée avec différents jeux de fonctions de base.

La méthode ESP atteste ici sa capacité à reproduire précisément la distribution des charges dans de petites molécules, aussi bien que le décrit la fonction d'onde. La méthode de Mulliken est beaucoup moins précise, notamment avec l'utilisation de jeux de fonctions de base étendus comme pour 6-311G**. Ce problème est drastiquement accentué pour les méthodes de partition multiatomique. En revanche, les méthodes

de partition diatomique améliorent de façon satisfaisante les charges de Mulliken. En particulier, la méthode DHB3 obtient de bons résultats qui s'approchent de ceux de la méthode ESP avec l'augmentation du nombre de fonctions de base. La disparition de la dépendance de cette analyse de population aux jeux de fonctions de base est un résultat important afin d'envisager des applications ultérieures.

2.3.3 Calcul de dipôles DFT

La même procédure a été effectuée pour le calcul des moments dipolaires à partir de la fonctionnelle DFT-B3LYP. L'ensemble des 200 dipôles calculés avec chaque méthode a été comparé aux valeurs expérimentales. Les erreurs absolues moyennes sont reportées dans le tableau 2.4 pour les principales méthodes.

Base	DEN	ESP	MLK	LNR	DHB3	DXB3	MHB3	MXB3
3-21G	0,29	0,30	0,99	0,78	0,53	0,70	0,73	0,92
6-31G*	0,22	0,23	0,60	0,52	0,48	0,86	0,53	1,01
6-311G**	0,25	0,27	1,53	0,60	0,45	0,76	2,46	2,58

TAB. 2.4: Erreur absolue moyenne dans le calcul des 200 dipôles (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir d'une fonction d'onde B3LYP calculée avec différents jeux de fonctions de base.

Les résultats DFT sont tout à fait comparables à ceux des méthodes *ab initio*. Nous obtenons avec DHB3 les résultats les plus intéressants, bien que le niveau de précision de la méthode ESP ne soit pas atteint.

2.3.4 Calcul de dipôles semi-empiriques

La même procédure a de nouveau été appliquée, en utilisant désormais des fonctions d'onde semi-empiriques. Cependant, l'énergie des structures a été tout d'abord minimisée avec le potentiel semi-empirique de calcul. Les dipôles de ces structures optimisées ont été calculés avec chaque méthode et comparés aux valeurs expérimentales. Les erreurs absolues moyennes sur les 200 dipôles sont reportées dans le tableau 2.5 pour les principales méthodes.

Le tableau A.4 réunit les statistiques de tous les calculs de dipôles *ab initio*, DFT et semi-empiriques. Il doit être noté que l'ensemble des 200 molécules n'a pas pu être traité avec PDDG et RM1. Les paramètres du brome n'existent pas pour PDDG

(10 molécules concernées), tandis que ceux du silicium n'existent pas pour RM1 (6 molécules concernées).

Méthode	DEN	LDW	MLK	LNR	DHB3	DXB3	MHB3	MXB3
MNDO	0,44	0,76	0,68	0,63	0,64	0,87	0,64	0,90
PM3	0,40	0,74	0,53	0,56	0,71	0,77	0,68	0,71
AM1	0,39	0,62	0,52	0,59	0,69	0,89	0,69	0,94
PDDG	0,36	0,65	0,56	0,61	0,87	1,10	0,83	1,10
RM1	0,35	0,64	0,60	0,62	0,73	1,08	0,70	1,08

TAB. 2.5: Erreur absolue moyenne dans le calcul des 200 dipôles (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir de différentes fonctions d'onde semi-empiriques.

La méthode ESP n'étant pas disponible dans la librairie fDynamo, les résultats de référence deviennent ceux de la densité électronique de la fonction d'onde semi-empirique. En ce qui concerne les méthodes traditionnelles, l'analyse de Mulliken donne de meilleurs résultats que celle de Löwdin. Cette amélioration est probablement due au fait que l'analyse de population est effectuée à partir de la matrice densité exacte. Les problèmes rencontrés avec les méthodes *ab initio* disparaissent du fait que les fonctions de base sont peu diffuses pour les méthodes semi-empiriques. Ainsi, les résultats de la méthode de Mulliken sont raisonnables, bien qu'ils demeurent éloignés de ceux obtenus directement à partir de la fonction d'onde.

Pour nos méthodes de partition, l'utilisation de fonctions de base étroites engendrent des résultats similaires pour les partitions di- ou multi-atomiques de Becke. Le traitement multiatomique n'est donc pas nécessaire pour des fonctions d'onde semi-empiriques. En revanche, le traitement hétéronucléaire donne de moins bons résultats que le traitement homonucléaire. Ce dernier présente toutefois des résultats moins bons que ceux de l'analyse de Mulliken. Seule la méthode de partition linéaire donne des résultats similaires à l'analyse de Mulliken.

Pour analyser plus finement les résultats, nous avons recalculé les statistiques en classant les molécules par familles. Alors que les résultats des calculs *ab initio* et DFT sont parfaitement homogènes, ceux des méthodes semi-empiriques peuvent être légitimement séparés en deux classes principales. Une différence marquée est en effet constatée entre les molécules contenant simplement les quatre éléments hydrogène,

carbone, oxygène, azote et les autres molécules. Les erreurs absolues moyennes sur le calcul des 113 dipôles de la première classe de molécules sont reportées dans le tableau 2.6.

Méthode	DEN	LDW	MLK	LNR	DHB3	DXB3	MHB3	MXB3
MNDO	0,46	0,96	0,69	0,56	0,51	0,64	0,52	0,62
PM3	0,42	0,84	0,57	0,46	0,54	0,71	0,50	0,63
AM1	0,39	0,76	0,53	0,51	0,57	0,84	0,54	0,85
PDDG	0,36	0,68	0,45	0,46	0,63	0,90	0,57	0,81
RM1	0,36	0,69	0,43	0,42	0,56	0,88	0,51	0,82

TAB. 2.6: Erreur absolue moyenne dans le calcul des 113 dipôles des molécules constitués uniquement des atomes H, C, N et O (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir de différentes fonctions d'onde semi-empiriques.

Les tendances notées pour l'ensemble des 200 molécules subsistent, bien que des caractéristiques marquées apparaissent pour cette famille de 113 molécules. Tout d'abord, la médiocrité de l'analyse de Löwdin est accentuée pour cette famille de molécules "HCNO". Cette analyse est effectuée à partir de la matrice densité \underline{P}' . Le contraire est en revanche observé pour les autres méthodes qui utilisent la matrice densité exacte \underline{P} . Les résultats des 87 dipôles de la seconde classe de molécules sont reportées dans le tableau 2.7.

Méthode	DEN	LDW	MLK	LNR	DHB3	DXB3	MHB3	MXB3
MNDO	0,40	0,51	0,67	0,72	0,80	1,18	0,78	1,26
PM3	0,38	0,60	0,48	0,69	0,93	0,84	0,91	0,82
AM1	0,39	0,45	0,51	0,69	0,86	0,96	0,88	1,06
PDDG	0,37	0,61	0,72	0,84	1,23	1,38	1,22	1,53
RM1	0,34	0,57	0,85	0,89	0,98	1,36	0,96	1,45

TAB. 2.7: Erreur absolue moyenne dans le calcul des 87 dipôles des molécules constitués des atomes F, Br, Cl, Si, P et S (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir de différentes fonctions d'onde semi-empiriques.

Les tendances observées pour la première famille de molécules sont logiquement inversées pour la seconde. Aucune différence entre les deux familles n'est pourtant observée dans le calcul des observables dipôles. Le phénomène ne remet donc pas en question les méthodes semi-empiriques pour cette famille de molécules. La diffusion des orbitales atomiques est la différence la plus évidente entre les 2 familles de

molécules. La qualité des analyses de population dépend alors de la matrice densité utilisée (\underline{P}' ou \underline{P}) et probablement de la diffusion des fonctions de base du système étudié.

2.3.5 Conclusion des résultats

En résumé, ce sont nos schémas de partition les plus simples qui donnent les meilleurs résultats. Les méthodes de partition di-homoatomiques LNR et DHB3 apportent la meilleure réponse aux limites de l'analyse de Mulliken pour des fonctions d'onde *ab initio* et DFT. Les deux méthodes donnent de bons résultats sans être affectées par l'utilisation de fonctions de base diffuses.

L'analyse de Mulliken décrit convenablement les fonctions d'onde dans le cas des méthodes semi-empiriques. Les méthodes LNR et DHB3 donnent des résultats de qualité similaire à celle de Mulliken pour ce type de potentiels. Toutefois, un problème affecte l'ensemble de ces méthodes pour le traitement d'atomes diffus avec les fonctions d'onde semi-empiriques. Une vigilance particulière doit être conservée lors de l'application de ces méthodes à ce cas de figure.

Nos schémas de partition les plus complexes donnent les résultats les moins satisfaisants. Les schémas de partition multiatomiques ne sont pas convenables pour des fonctions d'onde *ab initio* et DFT. Ils sont par ailleurs inutiles pour l'analyse des fonctions d'onde semi-empiriques. Le traitement hétéronucléaire donne en général de moins bons résultats que le traitement homonucléaire.

Chapitre 3

Les interactions électrostatiques

Les méthodes de calcul de charges partielles du chapitre précédent ont été développées dans le but d'être utilisées pour le traitement des interactions électrostatiques dans les systèmes moléculaires. Leur application en tant que telle est présentée dans ce nouveau chapitre.

3.1 La modélisation moléculaire

La modélisation moléculaire est une représentation des systèmes chimiques qui a pour but d'expliquer un grand nombre de leurs propriétés. Le modèle doit être assez simple pour permettre une compréhension rapide et assez complexe pour représenter correctement le système étudié. Par ailleurs, l'évolution du modèle doit être facilement prédictible par des techniques de simulation (cf partieB).

3.1.1 Le système moléculaire

La structure d'un système moléculaire peut être définie par une matrice $\underline{\underline{R}}$ de dimension $3 * M$ qui contient les coordonnées cartésiennes des M atomes. Ce système évolue constamment entre des configurations d'équilibre de basse énergie.

L'énergie du système

Au voisinage d'une structure $\underline{\underline{R}}_{eq}$ à l'équilibre, l'énergie potentielle du système peut s'écrire en série de Taylor :

$$E^{pot} = E_{eq}^{pot} + \underline{\underline{G}}_{eq}^\dagger (\underline{\underline{R}} - \underline{\underline{R}}_{eq}) + \frac{1}{2} (\underline{\underline{R}} - \underline{\underline{R}}_{eq})^\dagger \underline{\underline{H}}_{eq} (\underline{\underline{R}} - \underline{\underline{R}}_{eq}) + \dots \quad (3.1)$$

où $\underline{\underline{G}}_{eq}$ et $\underline{\underline{H}}_{eq}$ sont les matrices du gradient et de l'hessien de l'énergie à l'équilibre. Le gradient de l'énergie est nul à l'équilibre. Le terme de l'hessien modélise le caractère

harmonique du puit de potentiel. Les termes d'ordres supérieurs rendent compte du caractère anharmonique qui s'accroît en s'éloignant de l'équilibre.

Le gradient de l'énergie

La matrice gradient \underline{G} est de dimension $3*M$, étant constituée de M vecteurs :

$$\underline{g}_A = \frac{\partial E^{pot}}{\partial \vec{R}_A} = \begin{pmatrix} \frac{\partial E^{pot}}{\partial x_A} \\ \frac{\partial E^{pot}}{\partial y_A} \\ \frac{\partial E^{pot}}{\partial z_A} \end{pmatrix} \quad A = 1, 2, \dots, M \quad (3.2)$$

Une optimisation de géométrie permet d'accéder à un minimum local, proche de la structure de départ. En pratique, le critère de convergence pour la minimisation de l'énergie est défini comme :

$$G_{RMS} = \sqrt{\frac{1}{3M} \sum_{A=1}^M \underline{g}_A^\dagger \underline{g}_A} \approx 0 \quad (3.3)$$

L'hessien de l'énergie

La matrice hessien \underline{H} est de dimension $3M*3M$, étant constituée de M^2 tenseurs :

$$\underline{H}_{AB} = \frac{\partial^2 E^{pot}}{\partial \vec{R}_A \partial \vec{R}_B} = \begin{pmatrix} \frac{\partial^2 E^{pot}}{\partial x_A \partial x_B} & \frac{\partial^2 E^{pot}}{\partial x_A \partial y_B} & \frac{\partial^2 E^{pot}}{\partial x_A \partial z_B} \\ \frac{\partial^2 E^{pot}}{\partial y_A \partial x_B} & \frac{\partial^2 E^{pot}}{\partial y_A \partial y_B} & \frac{\partial^2 E^{pot}}{\partial y_A \partial z_B} \\ \frac{\partial^2 E^{pot}}{\partial z_A \partial x_B} & \frac{\partial^2 E^{pot}}{\partial z_A \partial y_B} & \frac{\partial^2 E^{pot}}{\partial z_A \partial z_B} \end{pmatrix} \quad (3.4)$$

La diagonalisation de la matrice hessien fournit les modes normaux du système et leur fréquence de vibration. Une molécule non-linéaire possède $3M-6$ modes de vibration indépendants ($3M-5$ pour une molécule linéaire) et $3M-6$ coordonnées internes représentant les degrés de liberté du système. Une description en coordonnées cartésiennes possède 6 degrés de liberté supplémentaires qui définissent la position et l'orientation du système dans l'espace. De même, les 6 modes non vibrationnels correspondent aux modes de rotation et de translation dans les 3 directions de l'espace.

Nous pouvons écrire l'énergie totale du système moléculaire comme :

$$E = E^{pot} + E^{vib} + E^{rot} + E^{trans} \quad (3.5)$$

où E^{pot} , E^{vib} , E^{rot} , E^{trans} sont les énergies potentielle, de vibration, de rotation et de translation respectivement. Les trois derniers termes correspondent à l'énergie cinétique des noyaux. Cette décomposition énergétique suppose que les différents types

d'états sont indépendants. Cette approximation est tout à fait juste pour les points stationnaires¹ d'un système isolé (phase gaz). Pour les autres cas comme une protéine en solution, les couplages entre états sont non négligeables et seul un échantillonnage représentatif des états accessibles au système (espace de phases) permet d'obtenir l'ensemble des niveaux d'énergie correspondants. Les lois de la mécanique statistique permettent ensuite de retrouver des grandeurs thermodynamiques.

Chaque mode de vibration obéit aux lois de la mécanique quantique : les niveaux d'énergie sont quantifiés et leur écart est proportionnel à la fréquence de vibration. La spectroscopie infrarouge permet de provoquer les transitions de niveaux pour chaque mode et d'obtenir ainsi le spectre des fréquences de vibration. Pour un système de petite taille, il est possible de reproduire ce spectre par un calcul de mécanique quantique. Pour un système de taille importante, le spectre devient très complexe et les méthodes de chimie quantique trop coûteuses. Il est néanmoins possible de calculer l'hessien de l'énergie d'une protéine par la mécanique classique. Les modes de vibration obtenus donnent un aperçu instructif des mouvements que peut subir le système à différentes échelles de temps.

3.1.2 La mécanique moléculaire

La mécanique moléculaire considère la molécule comme un ensemble de particules classiques. Les atomes d'un système sont définis comme des sphères chargées et reliées entre elles par des ressorts. Une fonction d'énergie spécifique à ce modèle permet de considérer les propriétés intrinsèques² d'une molécule.

Le champ de force

Un champ de force est l'association d'une fonction d'énergie potentielle et d'un ensemble de paramètres déterminés pour une famille de composés. Les paramètres sont généralement obtenus pour des conditions d'équilibre de l'état fondamental. Cette paramétrisation est faite à partir de données expérimentales (spectroscopie, cristallographie) et de calculs quantiques de haute précision.

¹Un point de la surface d'énergie potentielle est dit "stationnaire" lorsque le gradient de son énergie est nul (cas des minima, maxima et point selles).

²même purement quantiques

Afin d'étudier les macromolécules biologiques, différents champs de force ont été développés [69, 70, 71, 72]. Leur fonction d'énergie potentielle est similaire et seul leur mode de paramétrisation les différencie. Pendant cette thèse, le champs de force OPLS (*Optimized Potential for Liquid Simulations*) [71] a été utilisé. La forme simple de sa fonction empirique donne une description rapide et correcte d'une protéine, tel que :

$$E^{pot} = E_{liaisons} + E_{angles} + E_{dièdres} + E_{impropres} + E_{elec} + E_{vdW} \quad (3.6)$$

où les quatre premiers termes de l'énergie potentielle correspondent aux interactions liantes et les deux derniers aux interactions non liantes (électrostatique, Van der Waals). Les paramètres OPLS des liaisons, des angles de valence et des dièdres impropres sont en grande partie issus du champ de force AMBER (*Assisted Model Building with Energy Refinement*) [70]. Le développement d'OPLS se focalise principalement sur l'optimisation des paramètres de torsion et ceux des interactions non liantes .

L'énergie des liaisons

L'énergie d'étirement des liaisons est définie par la loi de Hooke :

$$E_{liaisons} = \sum_{l=1}^{liaisons} \frac{1}{2} k_l (l - l_0)^2 \quad (3.7)$$

où k_l est la constante de force de la liaison, l et l_0 sont les longueurs de la liaison à un instant donné et à l'équilibre respectivement.

L'énergie des angles de valence

L'énergie de variation des angles de valence est également définie par un potentiel harmonique :

$$E_{angles} = \sum_{\theta=1}^{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (3.8)$$

où k_θ est la constante de force de l'angle de valence, θ et θ_0 sont les valeurs de l'angle à un instant donné et à l'équilibre respectivement.

L'énergie des angles dièdres

L'énergie de rotation autour des angles dièdres est définie par une série de Fourier :

$$E_{dièdres} = \sum_{\phi=1}^{dièdres} \frac{V_1^\phi}{2} (1 + \cos \phi) + \frac{V_2^\phi}{2} (1 - \cos 2\phi) + \frac{V_3^\phi}{2} (1 + \cos 3\phi) \quad (3.9)$$

où V_i^ϕ est la constante de force du dièdre ϕ pour un ordre de périodicité i .

L'énergie des dièdres impropres

L'énergie de distorsion hors du plan des dièdres impropres est également définie par une série de Fourier :

$$E_{impropres} = \sum_{\omega=1}^{impropres} \frac{V_1^\omega}{2}(1 + \cos \omega) + \frac{V_2^\omega}{2}(1 - \cos 2\omega) + \frac{V_3^\omega}{2}(1 + \cos 3\omega) \quad (3.10)$$

où V_i^ω est la constante de force de l'impropre ω pour un ordre de périodicité i .

L'énergie électrostatique

L'énergie des interactions électrostatiques est définie par la loi de Coulomb :

$$E_{elec} = \sum_{A=1}^M \sum_{B>A}^M \frac{f_{AB}}{4\pi\epsilon_0\epsilon} \frac{q_A q_B}{|\vec{R}_A - \vec{R}_B|} \quad (3.11)$$

où ϵ_0 et ϵ sont les permittivités respectives du vide et du milieu ; q_A est la charge partielle de l'atome A et \vec{R}_A sa position ; f_{AB} est un facteur d'échelle empirique appliqué aux interactions non liantes : il est nul quand A et B sont distants de une ou deux liaisons, égal à $1/2$ quand A et B sont distants de trois liaisons et égal à 1 dans les autres cas.

Le potentiel coulombien (3.11) est largement utilisé pour l'étude des protéines. Les charges partielles sont généralement déterminées par une méthode de type ESP (cf § 2.1.3). Certains champs de force dédiés aux molécules organiques utilisent des multipôles d'ordre supérieur (cf § 2.2.4) pour une description plus fine des interactions électrostatiques [73]. Un tel traitement est trop complexe pour une modélisation efficace des protéines et il est généralement peu utilisé pour les petites molécules. De plus, la grande dépendance des multipôles aux conformations structurales limite la transférabilité de paramètres [74, 75, 76]. Une approximation plus rude est la fixation des charges atomiques ponctuelles. Certains champs de force dit "polarisables" introduisent un terme d'induction qui modélise la capacité du nuage électronique d'un atome à évoluer par rapport à son environnement. Cette amélioration permet de faire évoluer la charge atomique pendant une simulation, mais implique un temps de calcul plus long.

L'énergie de Van Der Waals

L'énergie des interactions de Van der Waals est définie par un potentiel de Lennard-Jones :

$$E_{VdW} = \sum_{A=1}^M \sum_{B>A}^M 4 f_{AB} \sqrt{\varepsilon_{AA} \varepsilon_{BB}} \left[\left(\frac{\sqrt{\sigma_{AA} \sigma_{BB}}}{|\vec{R}_A - \vec{R}_B|} \right)^{12} - \left(\frac{\sqrt{\sigma_{AA} \sigma_{BB}}}{|\vec{R}_A - \vec{R}_B|} \right)^6 \right] \quad (3.12)$$

où les paramètres ε et σ déterminent respectivement la profondeur et la position du puit de potentiel.

Le terme en $1/R^{12}$ représente l'interaction répulsive d'échange qui traduit les phénomènes classique et quantique (cf § 1.2.6) que deux nuages électroniques ne peuvent occuper le même espace. Le terme en $1/R^6$ représente l'interaction attractive de dispersion qui traduit une fluctuation corrélée des nuages électroniques de deux atomes.

3.1.3 Les potentiels hybrides MQ/MM

La mécanique moléculaire permet d'étudier efficacement les changements conformationnels des molécules, mais sa paramétrisation impose un état électronique et une configuration donnée du système. Seule la mécanique quantique est appropriée pour traiter les changements d'états électroniques, les processus réactionnels et tout phénomène qui implique une grande variation de la structure électronique. Mais le coût élevé des calculs de mécanique quantique limite considérablement la taille des systèmes qu'il est possible d'examiner. Un moyen pour étudier efficacement les réactions en solution ou dans les systèmes biologiques est de former des potentiels hybrides couplant la mécanique quantique (MQ) et la mécanique moléculaire (MM) [77, 78, 79].

Les potentiels hybrides MQ/MM modélise le site réactif d'un système par la mécanique quantique et son environnement par la mécanique moléculaire, comme décrit par la figure 3.1. Ce partage est associé à un hamiltonien mixte de type :

$$\hat{H} = \hat{H}_{MQ} + \hat{H}_{MM} + \hat{H}_{MQ/MM} + \hat{H}_{limites} \quad (3.13)$$

où \hat{H}_{MQ} et \hat{H}_{MM} sont les hamiltoniens propres aux régions MQ et MM décrites respectivement par la mécanique quantique et la mécanique moléculaire ; $\hat{H}_{MQ/MM}$ est l'hamiltonien qui décrit l'interaction entre les atomes MQ et les atomes MM ; $\hat{H}_{limites}$ décrit les interactions du système étudié avec son environnement à longue distance

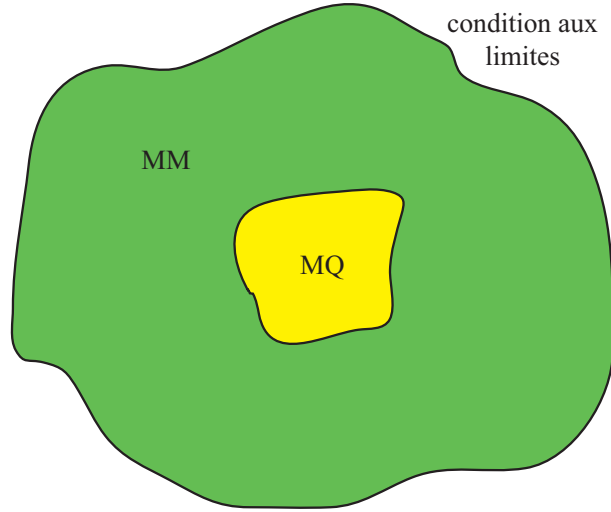


FIG. 3.1: Division du système moléculaire par un potentiel hybride MQ/MM.

appelé communément "conditions aux limites".

L'équation de Schrödinger associé à l'hamiltonien (3.13) prend la forme :

$$\hat{H} \Psi(\{\vec{r}_q\}; \{\vec{R}_Q\}, \{\vec{R}_M\}) = E^{pot}(\{\vec{R}_Q\}, \{\vec{R}_M\}) \Psi(\{\vec{r}_q\}; \{\vec{R}_Q\}, \{\vec{R}_M\}) \quad (3.14)$$

où les positions nucléaires $\{\vec{R}_Q\}$ des atomes MQ et $\{\vec{R}_M\}$ des atomes MM sont des paramètres pour la fonction d'onde électronique Ψ (cf § 1.2.4). Par conséquent, les termes énergétiques dépendant uniquement des positions nucléaires peuvent être traités comme des constantes lors de la résolution de l'équation (3.14), tel que :

$$E^{pot} = \langle \Psi | \hat{H}_{MQ} + \hat{H}_{MQ/MM}^{elec} + \hat{H}_{limites(MQ)} | \Psi \rangle + E_{MM}^{pot} + E_{MQ/MM}^{VdW} + E_{limites(MM)} \quad (3.15)$$

L'hamiltonien \hat{H}_{MQ} est similaire à l'hamiltonien (1.19). L'hamiltonien $\hat{H}_{MQ/MM}^{elec}$ décrivant les interactions électrostatiques entre les régions MQ et MM requiert une attention particulière et celles-ci seront abordées dans la prochaine section. Les conditions aux limites influençant la partie MQ sont résolues par le calcul variationnel, tandis que l'influence sur la partie MM est traitée séparément. L'énergie potentielle E_{MM}^{pot} de la partie MM correspond à l'expression (3.6). Les interactions de Van der Waals MQ/MM sont traitées par le potentiel de Lennard-Jones (3.12), en conservant le facteur d'échelle du champ de force (cf § 3.1.2).

3.2 L'électrostatique en MQ/MM

3.2.1 Le traitement usuel

La manière traditionnelle de traiter les interactions électrostatiques entre les atomes MQ et les atomes MM est décrite par l'hamiltonien ci-dessous :

$$\hat{H}_{MQ/MM}^{elec} = \sum_{M=1}^{M_M} \sum_{Q=1}^{M_Q} \frac{q_M Z_Q}{|\vec{R}_M - \vec{R}_Q|} - \sum_{M=1}^{M_M} \sum_{q=1}^{N_Q} \frac{q_M}{|\vec{R}_M - \vec{r}_q|} \quad (3.16)$$

où les M_M charges partielles atomiques q_M interagissent avec les M_Q charges nucléaires Z_Q dans le premier terme et les N_Q électrons dans le second terme.

L'énergie associée à l'hamiltonien (3.16) s'écrit :

$$E_{MQ/MM}^{elec} = \sum_{M=1}^{M_M} q_M \left\{ \sum_{Q=1}^{M_Q} \frac{Z_Q}{|\vec{R}_M - \vec{R}_Q|} - \iiint \frac{\rho(\vec{r})}{|\vec{R}_M - \vec{r}|} d\vec{r} \right\} \quad (3.17)$$

Le terme impliquant les charges nucléaires est recalculé à chaque changement de géométrie, tandis que la polarisation de la densité électronique par les charges MM est considérée lors du calcul SCF.

Lorsqu'il n'existe aucune liaison covalente entre les parties MQ et MM, les interactions électrostatiques sont naturellement calculées entre toutes les paires d'atomes MQ/MM. Lorsque la division du système impose la coupure de liaisons covalentes, aucun facteur d'échelle n'est généralement appliqué à la frontière contrairement aux interactions de Van der Waals.

3.2.2 La frontière MQ/MM

Bien que les interactions entre les régions MQ et MM soient définies, la coupure de liaisons covalentes doit être traitée convenablement. Tout d'abord, les interactions liantes perturbées par une telle division MQ/MM doivent être modélisées afin d'assurer la continuité du système à l'interface des deux régions. La méthode habituelle est de conserver dans le calcul de mécanique moléculaire toutes les coordonnées internes qui implique au minimum un atome MM. Cette définition permet d'exclure la région MQ du calcul empirique, tout en assurant sa cohésion avec la région MM.

Il est ensuite nécessaire de compléter la densité électronique de la région MQ, sous peine de conserver des radicaux au niveau des liaisons rompues par la division MQ/MM. La méthode la plus simple à mettre en application, et de ce fait la plus couramment utilisée, est la méthode de l'atome de lien [78, 79]. Elle consiste à ajouter un atome d'hydrogène à chaque coupure afin de compléter le système MQ, comme le montre la figure 3.2.

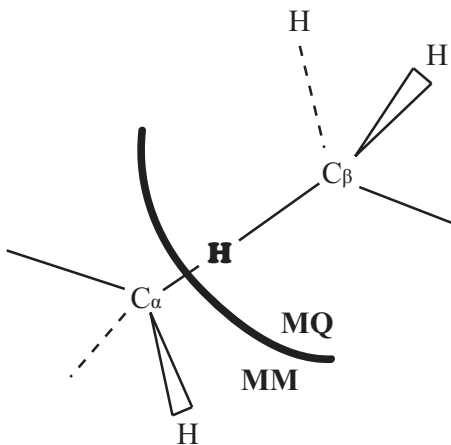


FIG. 3.2: Atome de lien à la frontière MQ/MM.

L'atome de lien permet d'effectuer un calcul convenable de la fonction d'onde électronique. Dans le traitement habituel, la seule interaction considérée de l'atome de lien avec son environnement MM est l'interaction électrostatique de la densité électronique avec les charges MM. L'atome de lien étant placé à proximité de l'atome MM frontière (Fig 3.2), la méthode souffre d'une surpolarisation de la densité électronique par la charge MM dans cette région. Une redistribution de la charge MM en question permet d'atténuer ces interactions à courte distance [80, 81]. L'alternative majeure à l'atome de lien est l'utilisation des orbitales hybrides gelées pour saturer le système MQ [82, 83]. Ce type de méthodes est plus rigoureux mais difficile à mettre en application. Il nécessite notamment une paramétrisation spécifique pour chaque méthode de chimie quantique.

Une étude comparative a montré que l'atome de lien donne des résultats de précision comparable aux autres méthodes à condition que la distribution des charges soit traitée correctement à l'interface MQ/MM [80]. Le problème de surpolarisation

due à la fuite de la densité sur les atomes MM est peu important avec les méthodes semi-empiriques, car les fonctions de base utilisées sont peu diffuses. En revanche, l'application de méthodes *ab initio* avec des jeux de fonctions de base étendus nécessite plus de vigilance [84, 85]. L'expansion des potentiels hybrides *ab initio* et les problèmes mentionnés ci-dessus incitent au développement d'une méthode standard pour le calcul des interactions électrostatiques MQ/MM [86].

3.2.3 Le traitement proposé

Nous proposons un nouveau traitement des interactions électrostatiques MQ/MM, tel que :

$$E_{MQ/MM}^{elec} = \sum_{M=1}^{M_M} \sum_{Q=1}^{M_Q} \frac{q_M q_Q}{|\vec{R}_M - \vec{R}_Q|} \quad (3.18)$$

où $E_{MQ/MM}^{elec}$ est la somme des interactions électrostatiques entre les M_M charges partielles atomiques q_M du champs de force et les M_Q charges partielles atomiques q_Q calculées à partir de la fonction d'onde électronique.

Cette nouvelle approche répond à plusieurs nécessités concernant l'application des potentiels hybrides. Tout d'abord, elle propose un traitement plus cohérent des interactions électrostatiques. L'interaction calculée habituellement entre une répartition précise (densité) de charge MQ et approximative (points) de charge MM est asymétrique et probablement inappropriée. La définition (3.18) propose un traitement logique des interactions électrostatiques entre des charges décrites au même niveau dans la partie MQ et dans la partie MM.

De plus, elle devrait permettre de réduire fortement le coût d'un calcul MQ/MM. La détermination des interactions électrostatiques MQ/MM peut être l'étape limitante du calcul pour un grand système moléculaire. Si des milliers d'atomes MM interagissent avec un site MQ de taille réduite, autant d'intégrales doivent être théoriquement ³ calculées afin de déterminer les interactions électrostatiques de la densité électronique avec chacune des charges atomiques MM. Ce cas de figure récurrent en biochimie motive l'adoption du potentiel (3.18), du fait que le calcul des charges partielles doit être beaucoup moins coûteux que le calcul des intégrales.

³L'élimination des interactions à très longue distance réduit le nombre d'intégrales à calculer.

Le troisième intérêt est d'éliminer les problèmes de surpolarisation de la densité électronique par les charges MM. La polarisation de la densité est toujours prise en compte à chaque pas de la procédure SCF, mais elle est désormais induite par les interactions des charges ponctuelles MQ et MM. L'atome de lien n'est pas pris en compte dans notre calcul (3.18), ce qui supprime les interactions à courte portée rencontrées dans les approches habituelles. Si l'atome de lien était compris dans le calcul, une redistribution des charges MM à la frontière serait nécessaire.

La précision du calcul (3.18) repose sur la validité des charges partielles calculées pour la région MQ. L'obtention de charges atomiques précises pour ce calcul offrira par ailleurs un suivi précieux de l'état de charge des atomes MQ lors de processus réactifs.

3.2.4 La matrice de Fock

Afin de calculer l'énergie d'un système avec le potentiel hybride MQ/MM que nous proposons, il est nécessaire de connaître la forme analytique des éléments de la matrice de Fock.

La définition générale

Un résultat important de la chimie quantique est obtenu en dérivant l'expression (1.50) de l'énergie de l'état fondamental par rapport à un élément de la matrice densité :

$$\begin{aligned}
 \frac{dE_0}{dP_{xy}} &= \left(\frac{\partial E_0}{\partial P_{xy}} \right) + \sum_{\mu=1}^L \sum_{\nu=1}^L \left(\frac{\partial E_0}{\partial G_{\nu\mu}} \right) \left(\frac{\partial G_{\nu\mu}}{\partial P_{xy}} \right) \\
 &= \left(H_{yx}^{core} + \frac{1}{2} G_{yx} \right) + \sum_{\mu=1}^L \sum_{\nu=1}^L \left(\frac{1}{2} P_{\mu\nu} \right) [(\nu\mu|yx) - \frac{1}{2}(\nu x|y\mu)] \\
 &= H_{yx}^{core} + \frac{1}{2} G_{yx} + \underbrace{\frac{1}{2} \sum_{\mu=1}^L \sum_{\nu=1}^L P_{\mu\nu} [(yx|\nu\mu) - \frac{1}{2}(y\mu|\nu x)]}_{G_{yx}} \quad (3.19)
 \end{aligned}$$

Du fait que la matrice de Fock est symétrique, nous pouvons conclure :

$$F_{xy} = \frac{dE_0}{dP_{xy}} \quad (3.20)$$

Les interactions électrostatiques

La propriété (3.20) permet de définir l'opérateur de Fock associé à l'énergie caractéristique (3.18) de notre approche. L'application du nouveau traitement s'est limitée aux potentiels hybrides semi-empiriques et nous définissons ainsi le terme associé comme :

$$F_{xy} = \frac{dE_{MQ/MM}^{elec}}{dP'_{xy}} \quad (3.21)$$

Il est judicieux de reformuler l'expression de l'énergie (3.18), comme :

$$E_{MQ/MM}^{elec} = \sum_{Q=1}^{M_Q} V_Q q_Q \quad , \quad avec \quad V_Q = \sum_{M=1}^{M_M} \frac{q_M}{|\vec{R}_M - \vec{R}_Q|} \quad (3.22)$$

où V_Q est le potentiel crée par l'ensemble des atomes MM sur l'atome Q.

Du fait que le potentiel V_Q est indépendant de la matrice densité, l'opérateur de Fock associé à l'énergie (3.22) s'écrit :

$$F_{xy} = \sum_{Q=1}^{M_Q} V_Q \left(\frac{dq_Q}{dP'_{xy}} \right) \quad , \quad avec \quad V_Q = \sum_{M=1}^{M_M} \frac{q_M}{|\vec{R}_M - \vec{R}_Q|} \quad (3.23)$$

Partant de la relation (3.23), nous avons développé des potentiels hybrides MQ/MM associés aux différentes méthodes de calcul de charge partielles précédemment testées (cf chap.2). Leur implémentation a été réalisée en Fortran90 dans la librairie *fDynamo*.

La méthode de Löwdin

La dérivation d'une charge de Löwdin (2.40) permet d'obtenir l'opérateur associé à l'utilisation de cette méthode, tel que :

$$F_{xy} = - \sum_{Q=1}^{M_Q} V_Q \times \begin{cases} \delta_{xy} & \text{si } (x, y) \in Q \\ 0 & \text{si } (x, y) \notin Q \end{cases} \quad (3.24)$$

où δ_{xy} est un delta de Kronecker.

La méthode de Mulliken

En dérivant la charge de Mulliken (2.41) d'un atome Q par rapport à un élément de la matrice densité $\underline{\underline{P'}}$, nous obtenons :

$$\frac{dq_Q}{dP'_{xy}} = - \sum_{\mu \in Q}^L X_{x\mu} X_{\mu y}^{-1} \quad (3.25)$$

Afin de conserver la symétrie de la matrice de Fock, nous définissons ses éléments comme :

$$F_{xy} = F_{yx} = -\frac{1}{2} \sum_{Q=1}^{M_Q} V_Q \sum_{\mu \in Q}^L (X_{x\mu} X_{\mu y}^{-1} + X_{y\mu} X_{\mu x}^{-1}) \quad (3.26)$$

Les méthodes de partition

En dérivant la charge atomique (2.42) d'une méthode de partition par rapport à un élément de la matrice densité \underline{P}' , nous obtenons :

$$\frac{dq_Q}{dP'_{xy}} = - \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{g=1}^K \sum_{h=1}^K X_{y\nu}^\dagger D_{\nu g}^\dagger w_Q^{gh} S'_{gh} D_{h\mu} X_{\mu x} \quad (3.27)$$

Afin de conserver la symétrie de la matrice de Fock, nous définissons ses éléments comme :

$$F_{xy} = F_{yx} = -\frac{1}{2} \sum_{Q=1}^{M_Q} \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{g=1}^K \sum_{h=1}^K (X_{y\nu}^\dagger X_{\mu x} + X_{x\nu}^\dagger X_{\mu y}) D_{\nu g}^\dagger V_Q w_Q^{gh} S'_{gh} D_{h\mu} \quad (3.28)$$

3.2.5 Le gradient de l'énergie

Afin d'appliquer nos potentiels hybrides MQ/MM à des méthodes de simulation, il est nécessaire d'obtenir la forme analytique du gradient de l'énergie.

La définition

En dérivant la définition de l'expression de l'énergie (3.22) par rapport à une position nucléaire, nous obtenons la forme générale de son gradient :

$$\frac{dE_{MQ/MM}^{elec}}{d\vec{R}_A} = \sum_{Q=1}^{M_Q} \left[\left(\frac{dV_Q}{d\vec{R}_A} \right) q_Q + V_Q \left(\frac{dq_Q}{d\vec{R}_A} \right) \right] \quad (3.29)$$

Le premier terme est obtenu par la dérivation du potentiel V_Q , tel que :

$$\frac{dV_Q}{d\vec{R}_A} = \sum_{M=1}^{M_M} \frac{q_M}{|\vec{R}_M - \vec{R}_Q|^3} \times \begin{cases} 0 & \text{si } A \neq M, Q \\ (\vec{R}_M - \vec{R}_Q) & \text{si } A = Q \\ (\vec{R}_Q - \vec{R}_M) & \text{si } A = M \end{cases} \quad (3.30)$$

Le second terme est différent suivant la méthode utilisée pour calculer les charges partielles.

La méthode de Löwdin

La dérivation des charges de Löwdin mène à :

$$\frac{dq_Q}{d\vec{R}_A} = - \sum_{\mu \in Q}^L \left(\frac{\partial P'_{\mu\mu}}{\partial \vec{R}_A} \right) \equiv 0 \quad (3.31)$$

Les termes impliquant la dérivation des éléments de la matrice densité par rapport aux coordonnées atomiques sont appelés termes de Pulay [87]. Dans le cas présent des calculs semi-empiriques, ces termes sont nuls.

La méthode de Mulliken

La dérivation des charges de Mulliken mène à :

$$\frac{dq_Q}{d\vec{R}_A} = - \sum_{\mu \in Q}^L \sum_{x=1}^L \sum_{y=1}^L P'_{xy} \left[X_{y\mu}^{-1} \left(\frac{\partial X_{\mu x}}{\partial \vec{R}_A} \right) + \left(\frac{\partial X_{y\mu}^{-1}}{\partial \vec{R}_A} \right) X_{\mu x} \right] \quad (3.32)$$

Nous avons appliqué une méthode de perturbations [88, 89] usuellement utilisée pour la dérivation des éléments de la matrice de transformation symétrique. Cette méthode mène au résultat suivant :

$$\left(\frac{\partial X_{\mu x}}{\partial \vec{R}_A} \right) = - \sum_{\alpha=1}^L \sum_{\beta=1}^L \sum_{\lambda=1}^L \sum_{\sigma=1}^L \frac{T_{\mu\alpha} T_{\alpha\lambda}^\dagger \left(\frac{\partial S_{\lambda\sigma}}{\partial \vec{R}_A} \right) T_{\sigma\beta} T_{\beta x}^\dagger}{s_\alpha^{1/2} s_\beta^{1/2} (s_\alpha^{1/2} + s_\beta^{1/2})} \quad (3.33)$$

Nous avons par ailleurs reproduit les développements de cette méthode perturbative afin de définir la dérivée des éléments de la matrice de transformation symétrique inverse. Ces développements ont abouti à l'expression suivante :

$$\left(\frac{\partial X_{y\mu}^{-1}}{\partial \vec{R}_A} \right) = \sum_{\alpha=1}^L \sum_{\beta=1}^L \sum_{\lambda=1}^L \sum_{\sigma=1}^L \frac{T_{y\alpha} T_{\alpha\lambda}^\dagger \left(\frac{\partial S_{\lambda\sigma}}{\partial \vec{R}_A} \right) T_{\sigma\beta} T_{\beta\mu}^\dagger}{s_\alpha^{1/2} + s_\beta^{1/2}} \quad (3.34)$$

Au final, nous obtenons :

$$\frac{dq_Q}{d\vec{R}_A} = \sum_{\lambda=1}^L \sum_{\sigma=1}^L \left(\frac{\partial S_{\lambda\sigma}}{\partial \vec{R}_A} \right) W_{\sigma\lambda} \quad (3.35)$$

$$\text{avec } W_{\sigma\lambda} = \sum_{\alpha=1}^L \sum_{\beta=1}^L \sum_{x=1}^L \sum_{y=1}^L \sum_{\mu \in Q}^L T_{\sigma\beta} \frac{T_{\beta x}^\dagger P'_{xy} X_{y\mu}^{-1} T_{\mu\alpha} - s_\beta^{1/2} T_{\beta\mu}^\dagger X_{\mu x} P'_{xy} T_{y\alpha} s_\alpha^{1/2}}{s_\alpha^{1/2} s_\beta^{1/2} (s_\alpha^{1/2} + s_\beta^{1/2})} T_{\alpha\lambda}^\dagger$$

Les méthodes de partition

La dérivation des charges de nos méthodes de partition mène à :

$$\frac{dq_Q}{d\vec{R}_A} = - \sum_{\mu=1}^L \sum_{\nu=1}^L \sum_{x=1}^L \sum_{y=1}^L \sum_{g=1}^K \sum_{h=1}^K P'_{xy} D_{\nu g}^\dagger D_{h\mu} \times$$

$$\left(\frac{\partial X_{\mu x}}{\partial \vec{R}_A} X_{y\nu}^\dagger w_Q^{gh} S'_{gh} + X_{\mu x} \frac{\partial X_{y\nu}^\dagger}{\partial \vec{R}_A} w_Q^{gh} S'_{gh} + \right.$$

$$\left. X_{\mu x} X_{y\nu}^\dagger \frac{\partial w_Q^{gh}}{\partial \vec{R}_A} S'_{gh} + X_{\mu x} X_{y\nu}^\dagger w_Q^{gh} \frac{\partial S'_{gh}}{\partial \vec{R}_A} \right) \quad (3.36)$$

Les deux premiers termes sont obtenus à partir de la définition (3.33). Le troisième terme diffère suivant le schéma de partition : la dérivée des coefficients w_Q^{gh} est quelque peu fastidieuse et ne sera pas décrite ici. Le dernier terme est obtenu par dérivation des éléments de la matrice de recouvrement des primitives.

L'implémentation d'une telle expression nécessite quelques précautions. La validité du gradient analytique implémenté est testée par comparaison au gradient numérique. Ce dernier est calculé à partir de plusieurs énergies par la méthode des différences finies. En adoptant cette procédure, chaque terme de l'expression peut être vérifié séparément. Cela permet d'aller pas à pas vers l'implémentation exacte du gradient analytique.

3.3 Résultats et discussion

Afin de tester la qualité du traitement des interactions électrostatiques par les différents potentiels MQ/MM développés, nous avons calculé un ensemble d'énergies d'interactions intermoléculaires.

3.3.1 Calcul d'énergies de formation de dimères

L'énergie de formation de plusieurs dimères de molécules organiques a été calculée par différents potentiels. Cette énergie correspond à l'interaction favorable créée entre deux monomères, tel que :

$$E_{formation} = E_{dimere} - (E_{monomereA} + E_{monomereB}) \quad (3.37)$$

Les trois termes sont obtenus séparément par minimisation de l'énergie des structures du dimère, du premier et du second monomère respectivement. Dans le cas d'un cal-

cul hybride, le dimère est traité par un potentiel MQ/MM, le premier monomère par MQ et le second par MM.

La liste des 62 dimères utilisés ainsi que leur énergie de formation HF/6-31G* est reportée dans le tableau A.5. Pour 29 des 44 molécules présentes dans cette série de dimères, les paramètres OPLS n'étaient pas disponibles. Par conséquent, 107 énergies de formation de dimères ont pu être testées en MQ/MM. Les erreurs absolues moyennes par rapport aux énergies HF/6-31G* sont reportées dans le tableau 3.1 pour les principales méthodes. Le tableau A.6 rassemble les résultats de toutes les méthodes.

Potentiel	DEN	LDW	MLK	LNR	DHB3	DXB3	MHB3	MXB3
MNDO/MM	9,14	12,12	9,68	8,10	7,25	7,16	7,55	7,71
PM3/MM	8,14	11,66	10,28	7,41	7,05	8,59	7,31	8,68
AM1/MM	7,75	9,63	8,53	7,18	7,06	9,71	7,20	9,04
PDDG/MM	6,82	10,94	10,41	8,55	8,84	11,69	8,56	10,54
RM1/MM	6,25	9,98	8,57	7,26	7,53	10,70	7,49	9,48

TAB. 3.1: Erreur absolue moyenne dans le calcul des 107 énergies de formation de dimères (en kJ/mol) par rapport aux valeurs HF/6-31G* pour les différents potentiels hybrides MQ/MM (le champs de force MM est toujours OPLS).

Les potentiels hybrides traditionnels utilisent la densité électronique (DEN) pour le calcul des interactions électrostatiques MQ/MM. L'objectif principal de ce test est de voir si nos nouveaux potentiels hybrides peuvent reproduire des résultats de qualité similaire aux potentiels actuels. Les potentiels basés sur les charges de Mulliken donnent de meilleures énergies d'interaction que ceux basés sur les charges de Löwdin. Toutefois, ces deux potentiels sont beaucoup moins précis que les potentiels traditionnels.

Comme déjà observé pour le test des méthodes de calcul de charges partielles (cf § 2.3.4), les schémas de partition multiatomiques ont peu d'utilité pour l'analyse des fonctions d'onde semi-empiriques. Par ailleurs, le traitement hétéronucléaire donne toujours de moins bons résultats que le traitement homonucléaire.

Nos schémas di-homoatomiques LNR et DHB3 donnent à nouveau des résultats satisfaisants en reproduisant le niveau de précision des potentiels traditionnels. Avec

les méthodes semi-empiriques MNDO, PM3 et AM1, les potentiels LNR et DHB3 donnent même des énergies d'interaction plus proches des valeurs de référence (HF/6-31G*). Il doit toutefois être noté que ces bons résultats sont obtenus pour des dimères de molécules organiques. Des systèmes possédant davantage d'atomes diffus pourraient notablement altérer la qualité de ces résultats (cf § 2.3.4).

Conclusion et perspectives

Nous venons de présenter une nouvelle approche pour le traitement des interactions électrostatiques avec des potentiels hybrides MQ/MM. Les charges partielles des atomes MM n'interagissent plus avec la densité électronique mais avec des charges partielles des atomes MQ. Notre modèle se veut plus cohérent et contourne par la même occasion certains problèmes du modèle traditionnel. Pour le calcul des charges partielles dans la région MQ, nous avons développé une série de méthodes de partition de la densité électronique. Nous les avons ensuite appliqué au traitement des interactions électrostatiques avec des potentiels hybrides MQ/MM semi-empiriques.

Les schémas de partition les plus simples donnent les meilleurs résultats. Ceci assure tout d'abord une certaine efficacité dans le calcul des interactions électrostatiques. Les meilleures méthodes reproduisent les résultats des potentiels hybrides traditionnels pour des systèmes simples. Ces méthodes doivent désormais être appliquées aux potentiels hybrides *ab initio* et DFT. Les résultats obtenus lors des tests des charges partielles sont déjà prometteurs notamment pour l'utilisation de fonctions de base diffuses. Ce cas de figure représente un enjeu important, car il fait intervenir les problèmes de surpolarisation de la densité électronique à la frontière MQ/MM pour les potentiels traditionnels. Egalement, l'utilisation de potentiels hybrides de haut niveau de calcul deviendra probablement de plus en plus routinière dans les années futures. Notre approche offre des perspectives d'efficacité importantes pour l'application de tels potentiels aux protéines.

Nous souhaitons par ailleurs trouver des méthodes de calculs de charges partielles alternatives aux schémas de partition. Une méthode de type numérique pourrait donner des résultats de précision meilleure et définitivement bannir la dépendance des charges partielles aux jeux de fonctions de base. Une autre perspective pourrait être l'ajout d'un champ de force polarisable dans la région MM. Cela permettrait tout

d'abord de finaliser la cohérence de notre modèle avec un même degré de description des densités de charges dans les régions MQ et MM. Ensuite, la polarisation deviendrait mutuelle entre les deux régions et non plus limitée à la polarisation de la région MQ par la partie MM. En parallèle de l'amélioration constante des ressources informatiques, ce type de développement méthodologique devrait permettre de poursuivre les progrès de la modélisation moléculaire pour l'étude des systèmes biologiques.

Partie B - Etude par simulation numérique de protéines fluorescentes

Introduction

Les protéines fluorescentes sont désormais des outils incontournables pour l'imagerie cellulaire [90]. Cependant, ces marqueurs biologiques possèdent des limitations comme notamment la perte de fluorescence après un temps limité d'illumination (cf. annexe B). Un enjeu important est l'amélioration de leurs propriétés photophysiques et potentiellement la création de nouvelles [91]. Pour étudier et développer les protéines fluorescentes, une approche pragmatique est d'associer différents domaines de compétence scientifique allant de la biologie moléculaire aux techniques de cristallographie et de spectroscopie. En complément de ces méthodes expérimentales, la modélisation moléculaire est d'une grande utilité pour comprendre la chimie caractéristique de ces macromolécules biologiques.

Les propriétés intrinsèques d'un chromophore peuvent être étudiées *in vacuo* par les méthodes de la chimie quantique. Cela permet de comprendre les principales modifications qu'il subit lors de son excitation [92, 93, 94, 95, 96, 97, 98, 99]. Dans la protéine, l'environnement du chromophore modifie de façon importante ses propriétés intrinsèques. Cette complexe biochimie peut être abordée précisément avec les potentiels hybrides (cf § 3.1.3) de la mécanique quantique et de la mécanique moléculaire [100, 101, 102, 103, 104]. Les simulations de dynamique moléculaire classique permettent également d'obtenir de précieuses informations sur l'évolution de la structure biologique à son état fondamental [105, 106, 107] ou à ses états excités [108, 109, 110]. Lors de cette thèse, nous avons utilisé plusieurs de ces techniques de simulation numérique pour l'étude de protéines fluorescentes.

Dans cette seconde partie du manuscrit, nous présentons ces études dans un ordre historique. Depuis une quinzaine d'année, l'intérêt autour de la protéine *aequorea victoria* GFP a suscité le développement de nombreux mutants [111, 112, 113]. Le chapitre 4 présente une étude de la dynamique structurale de deux de ces mutants,

l'ECFP (*Enhanced Cyan Fluorescent Protein*) et la Cerulean. Des protéines homologues à *avGFP* ont été découvertes ultérieurement dans les coraux [114]. Au chapitre 5, nous présentons une étude du mécanisme de photoconversion de l'une d'entre elles, la protéine fluorescente photoactivable EosFP issue du corail *Lobophyllia Hemprichii*. Enfin, nos collaborateurs ont récemment développé un mutant d'EosFP nommé IrisFP qui présente de multiples propriétés de phototransformations. Le chapitre 6 revient sur l'étude de la stabilité thermodynamique des différents états d'IrisFP.

Chapitre 4

La dynamique structurale des protéines ECFP et Cerulean

Parmi les nombreux mutants de l'*avGFP*, les protéines fluorescentes cyans possèdent des spécificités importantes. Tout d'abord, elles sont largement utilisées avec la protéine YFP (*Yellow Fluorescent Protein*) dans les expériences de FRET (*Fluorescence Resonance Energy Transfer*) pour l'étude des interactions protéine-protéine [113]. Concernant les aspects de biophysique, la mutation de la tyrosine du chromophore de la GFP par un tryptophane élimine la possibilité d'un ESPT (Excited State Proton Transfer) [115]. Cette mutation fait également apparaître une complexité supplémentaire dans les propriétés de fluorescence [116]. En particulier, les rendements quantiques de fluorescence sont drastiquement affectés [117]. Les protéines fluorescentes cyans constituent ainsi des systèmes modèles pour identifier les facteurs qui influencent les propriétés photophysiques d'un chromophore.

Nous avons ici tenté de comprendre les différences qui existent entre les propriétés de fluorescence de l'ECFP et celles de l'un de ses mutants, la Cerulean. Le travail de dynamique moléculaire s'est appuyé sur des données structurales et spectroscopiques de haute précision. L'article ci-dessous relate l'ensemble de ce travail.

Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield

*Mickaël Lelimosin[‡], Marjolaine Noirclerc-Savoye[‡], Christelle Lazareno-Saez[‡], Bernhard Paetzold[§],
Sophie Le Vor[‡], Richard Chazal[‡], Pauline Macheboeuf[‡], Martin J. Field[‡], Dominique Bourgeois^{‡,§}, and
Antoine Royant^{‡,§,*}*

[‡]Institut de Biologie Structurale Jean Pierre Ebel, UMR5075 CEA-CNRS-Université Joseph Fourier, 41,
rue Jules Horowitz, 38027 Grenoble Cedex 1, [§]European Synchrotron Radiation Facility, 6 Rue Jules
Horowitz, BP 220, 38043 Grenoble Cedex, France

AUTHOR EMAIL ADDRESS * antoine.royant@ibs.fr

RECEIVED DATE

TITLE RUNNING HEAD Structural dynamics of Cerulean fluorescent protein

ABSTRACT. Enhanced Cyan Fluorescent Protein (ECFP) and its variant Cerulean are genetically-encoded fluorophores widely used as donors in FRET-based cell imaging experiments. To explain the improvement in the fluorescence properties of Cerulean when compared to ECFP, we have determined the high-resolution crystal structures of these two proteins at physiological pH, and performed molecular dynamics simulations. In both proteins, the N-terminal half of the seventh strand undergoes conformational changes on the nanosecond timescale, resulting in a complex set of van der Waals interactions with the chromophore. The Y145A and H148D mutations in Cerulean stabilize these interactions and allow the chromophore to be more planar and less prone to collisional quenching from Ile146, albeit only intermittently. As a consequence, the probability of non-radiative decay is

significantly decreased. In addition, we solved the structure of the T65A mutant of Cerulean. It displays two conformations of the chromophore, one planar and one twisted, thereby explaining the intermediate value of its fluorescence quantum yield. Our results highlight the considerable dynamical flexibility that exists in the vicinity of the tryptophan-based chromophore of these engineered fluorescent proteins, and provide insights which should allow the design of mutants with enhanced optical properties.

KEYWORDS Fluorescent proteins, ECFP, Cerulean, X-ray crystallography, molecular dynamics, *in crystallo* spectroscopy, tryptophan-based chromophore, van der Waals interactions

INTRODUCTION

The potential of Green Fluorescent Protein (GFP) to be a genetically-encoded fluorescent reporter was demonstrated as early as 1994.¹ The chromophore results from the autocyclization of three consecutive amino acid residues Serine-Tyrosine-Glycine located inside the β -barrel shaped core of the protein. Efforts to tune the color of its emission by random and site-directed mutagenesis enlarged its palette of colors to blue, cyan and yellow.² Mutation of the chromophore's tyrosine into a tryptophan was crucial to obtain the cyan-fluorescent protein ECFP (**Figure 1a**) and yellow fluorescence was achieved in EYFP by positioning a second tyrosine in a π -stacking interaction with the tyrosine of the chromophore. The spectral overlap between the ECFP and EYFP pair has made it an attractive system for the development of genetically-encoded FRET sensors.³ As a result, ECFP has become a popular reporter but with non-ideal spectroscopic properties: its brightness is significantly lower than most other widely-used fluorescent proteins,⁴ its excitation and emission fluorescence spectra are two-peaked and broad, and its fluorescence lifetime behavior is multi-exponential.⁵

Several groups have tried to optimize the spectroscopic properties of ECFP. CyPet was obtained by an evolutionary strategy, with the sole constraint of optimizing the FRET signal to a yellow FP.⁶ Rational design has only been possible after the crystal structure of ECFP has been determined.⁷ Inspection of the structure revealed the presence of two alternate conformations of the [144-149] strand, one very

close to the one adopted in GFP with the Tyr145 side chain in the vicinity of the chromophore (Conformation A or *gfp*, PDB code 1OXD, green trace of **Figure 1b**), and a new conformation in which Tyr145 points toward the bulk solvent (Conformation B or *ecfp*, PDB code 1OXE, cyan trace on **Figure 1b**). This supports the observation from ^{19}F NMR experiments that the chromophore interconverts between two conformations on the millisecond timescale.⁸ As Tyr 145 and His148 interact either with the solvent or with the chromophore in each conformation it was postulated that these residues modulate the fluorescence properties of the protein.⁹ The triple mutant S72A/Y145A/H148D, called Cerulean, was shown to exhibit a significant increase in brightness and a reduced heterogeneity of fluorescence lifetimes.^{9, 10} The structure of Cerulean was recently solved at pH 5.0, showing an unanticipated *trans* configuration of the $\text{C}_\beta\text{-C}_\gamma$ bond of the chromophore.¹¹ However, the differences in experimental conditions, most importantly in pH, make the identification of the reasons for fluorescence improvement very difficult, underscoring the need for a structure of Cerulean at a higher pH.

In this work, we report the crystal structures of ECFP and Cerulean under the same physiological pH crystallization conditions at 1.37 and 1.15 Å resolution, respectively. We have also performed fluorescence measurements (emission spectra and lifetimes) on crystals to ensure that the spectroscopic properties were preserved in the crystal, and that the structures reflect the properties they have in solution. The high similarity between the structures of ECFP and Cerulean prompted us to undertake a molecular dynamics simulations study to analyze the differences in chromophore interactions, in particular with the [143-150] strand. We were able to identify a structural determinant for the improved quantum yield (QY) of Cerulean. In addition, we solved the 1.27 Å resolution crystal structure of the T65A mutant of Cerulean in order to investigate the role of residue 65's side chain on the stabilization of the chromophore. Altogether our results give a highly dynamic picture of the interaction between the chromophore and its protein matrix.

MATERIALS AND METHODS

Molecular biology. ECFP cDNA in the pQE expression vector (Qiagen) was a kind gift from Prof. N. Budisa. The different mutants were obtained sequentially from ECFP by Quick change mutagenesis. ECFP-Y145A/H148D was obtained from ECFP with the following primers: Y145A/H148D_For, 5'-CTGGAGTACAACGCCATCAGCGACAACGTCTATATCACCG-3', and Y145A/H148D_Rev 5'-CGGTGATATAGACGTTGTCGCTGATGGCGTTGTACTIONCCAG-3'; Cerulean was obtained from ECFP-Y145A/H148D with the following primers : S72A_For, 5'-CGTGCAGTGCTTCGCCCGCTACCCCGAC-3', S72A_Rev, 5'-GTCGGGGTAGCGGGCGAAGC ACTGCACG-3'; Cerulean-T65A was obtained from Cerulean with the following primers : T65A_For, 5'-GTGACCACGCTAGCCTGGGGCGTGC-3', T65A_Rev, 5'-GCACGCCCCAGGCTAGC GTGGT CAC-3'. All constructs were confirmed by DNA sequencing.

Expression, purification. His-tagged recombinant proteins were expressed in E. coli BL21 CodonPlus (DE3) RIL cells (Stratagene) in auto-induction media,¹² at 27°C during 24 h. Cells were lysed by sonication in presence of 20 mM Tris pH 8, 500 mM NaCl with EDTA-free protease inhibitors (Complete, Roche). His-tagged proteins were purified on Ni-NTA Superflow column (Qiagen) and eluted with 100 mM imidazole in the above buffer. Fractions containing purified proteins were pooled, dialyzed against 20 mM Tris pH 8.0 and concentrated to 25 -70 mg/ml.

Thermal shift assay. Assays were performed using an IQ5 96-well format real-time PCR detection system (Bio-Rad) over a temperature range starting from 20°C up to 100°C, in increments of 0.2°C. Assay samples were done with 23µL of protein (2mg/ml) and 2 µl of 250x SYPRO Orange dilution (Molecular Probes) into a 96-well thin-wall PCR plate (BioRad). The plates were heated and changes in fluorescence for the environmentally sensitive dye SYPRO Orange were monitored through a charge-coupled device (CCD) camera.¹³ The melting temperature (T_m) values of each protein were determined

graphically from the fluorescence first derivative of the melting curves. The wavelengths for excitation and emission were 470 nm and 570 nm, respectively.

Crystallisation. We obtained diffraction-quality crystals in conditions based on those published for ECFP and EGFP.^{7, 14} Crystals were grown from protein solution at 12-20 mg/mL in 0.1 M Hepes pH 6.7-7.5, 14-20% PEG8000, 0-0.1 M MgCl₂ in 1-1.5 μ l hanging drops using the vapor diffusion method. Crystals were transferred to a cryoprotectant solution consisting of the mother liquor supplemented with glycerol 20% v/v.

Structure Determination. X-ray diffraction data were collected at 100 K on various macromolecular crystallography beamlines of the European Synchrotron Radiation Facility (Grenoble, France), and processed using either XDS¹⁵ or MOSFLM.¹⁶ All crystals belong to the P2₁2₁2₁ space group. Solvent content is approximately 30%. A 5% random set of reflections was selected for R-free flagging for the first ECFP dataset, and kept for each subsequent dataset. Structure refinement was carried out with Refmac.¹⁷ Structures have been deposited in the Protein Data Bank with PDB IDs XXXX for ECFP, YYYY for Cerulean and ZZZZ for Cerulean-T65A. Crystallographic data statistics can be found in **Table S1**.

Spectroscopic characterization. Absorption spectra of solution samples were measured on a UV-visible spectrophotometer (Cary 3E, Varian). Extinction coefficients were obtained by a method derived from the one used for tyrosine-containing GFP chromophores.¹⁸ Protein samples with an Optical Density not higher than 0.7 were denatured by mixing with an equal volume of NaOH 1 M. The method relies on knowledge of the absorption properties of a tryptophan-based chromophore, free from its protein environment, which is missing in the literature. We used 30 mM⁻¹.cm⁻¹ as an average of the published values of EC₄₃₄ of ECFP. EC for the denatured chromophore was then deduced to be 46 mM⁻¹.cm⁻¹ at 462 nm (**Figure S1**). Fluorescence emission spectra of protein samples were measured on a

fluorimeter (Fluorolog 3, HORIBA Jobin Yvon). Fluorescence quantum yields were calculated on protein samples with an optical density not higher than 0.05, using 9-amino acridine as a standard (QY= 0.97 in water). Finally, the overall level of fluorescence can be quantified by brightness, the product of the extinction coefficient and the quantum yield.

Absorption and fluorescence emission spectra for crystals were recorded at the Cryobench laboratory of the European Synchrotron Radiation Facility.¹⁹ Fluorescence lifetime measurements were performed using a time-correlated single-photon-counting (TCSPC) setup adapted for crystals and solutions at either cryogenic or room temperature, with an excitation wavelength of 440 nm and a 470 nm interferential emission filter.¹⁹

Molecular dynamics simulations. MD simulations were performed on the *ecfp* conformation of ECFP and the *ecfp* and *cerulean* conformations of Cerulean that were obtained by X-ray crystallography. The fDynamo library²⁰ was used to set up the systems and to analyze the MD trajectories. Simulations were carried out with the AMBER 9 program (University of California San Francisco), applying the OPLS-AA force field to the protein²¹ and the TIP3P water model to the solvent.²² Parameters for the chromophore were deduced from equivalent groups in the OPLS-AA force field and were similar to those previously used in the literature.²³ Protonation states for the proteins at pH 7 were estimated using pK_a calculations.²⁴ The resulting structures were then solvated in cubic boxes of water molecules and sufficient cations were added to ensure a total system charge of zero using the SODIUM program (Beckman Institute). Preparation of the systems for simulation consisted of an initial geometry optimization of the proteins to remove unfavorable steric interactions with solvent molecules, followed by molecular dynamics simulations of the full systems at constant volume and temperature (300 K) for 100 ps with time-decreasing harmonic constraints on heavy atom positions. Dynamics simulations of 39 ns duration were performed for each of the systems at constant pressure and temperature. The equations of motion were integrated using a leapfrog algorithm and a 2 fs time step. The temperature was maintained at 300 K by a Langevin thermostat with a 2 ps⁻¹ collision frequency. The average pressure

was held at 1 atm by a Berendsen coupling algorithm²⁵ with a relaxation time of 2 ps and SHAKE constraints²⁶ were applied to bonds involving hydrogen atoms. Long-range electrostatic interactions were calculated with the particle mesh Ewald method,²⁷ using 8 Å as cut-off distance. Coordinates were saved every 2 ps, resulting in 19500 structures per simulation, and all results presented in this article are based on this sampling. We observed an equilibration of the temperature in less than 2 ps, while the properties of volume, density, pressure and potential energy were equilibrated in less than 20 ps. Snapshots of the molecular dynamics simulations superimposed on the X-ray structures, as well as root mean square deviations on C α positions between simulated and X-ray structures, are reported as supplementary information for illustration purposes (**Figure S2**).

RESULTS

The aim of our study is to provide an explanation for the improvement of fluorescence properties in Cerulean compared to ECFP. Given the significant discrepancy in reported values,^{9, 28} we have redetermined molar extinction coefficients (EC) and quantum yields (QY) for ECFP, ECFP-Y145A/H148D and Cerulean (**Table 1**) to allow for an accurate comparison. ECs are indistinguishable, whereas the QYs of ECFP-Y145A/H148D and Cerulean are ~50% higher than that of ECFP. Fluorescence lifetime measurement experiments show that each protein displays two lifetimes, and the average fluorescence lifetime increases by ~20 % between ECFP and the other two mutants, reflecting the fluorescence QY improvement. Overall, this shows that the S72A mutation has no effect on the spectroscopic properties and our structural study will focus on the Y145A and H148D mutations to explain QY modulation.

We recorded absorption and fluorescence spectra on all mutant crystals, and measured their fluorescence lifetimes, to check whether the crystal environment affects the spectroscopic properties of the protein at room temperature. Fluorescence emission spectra and TCSPC histograms can be found as supporting information (**Figures S3 and S4**). The position of the absorption and emission peaks are

conserved, whereas the fluorescence lifetimes show a ~25 % decrease. This can be explained by the difference in the refractive index of the medium surrounding the protein.²⁹ The refractive index (RI) of water is 1.33, whereas a salt crystal has a RI of 1.54 and a lysozyme crystal can have a RI as high as 1.45.³⁰ Overall, this indicates that spectroscopic properties are well conserved on going from dilute solution to the concentrated arrangement of molecules in the crystal at room temperature and that the information derived from our structural analyses is relevant in explaining differences between mutants.

We wanted to check whether the mutations of Cerulean were beneficial to the thermal stability of the protein. Using the thermal shift assay, we indeed observed a gradual increase of the melting temperature (T_m) from ECFP to Cerulean with the sequential additions of Y145A/H148D and S72A double and single mutations (**Table 1**). Interestingly, we found a similar correlation with the resistance to denaturation in NaOH during EC measurements. Denaturation was immediate for ECFP and ECFP-Y145A/H148D, and complete after 3min for Cerulean and Cerulean-T65A, highlighting the stabilizing role of the S72A mutation.

Because we anticipated that accurate determination of the [143-150] strand structure would be fundamental to our study,⁷ we devised a specific procedure for the first steps of the ECFP structure determination process. Phases were obtained by molecular replacement using a partial model of the published structure of ECFP based on a common core identical to the *gfp* and *ecfp* conformations.⁷ This model excluded the chromophore, the complete strand from residues 143 to 150, all water molecules, and the side chains of Ser205, Ile220 and Glu222. After initial positional refinement, good quality electron density was interpretable for the chromophore, all side chains and most of the [143-150] strand. At first glance, the strand appeared to be in a single conformation that closely resembles the *ecfp* conformation. However, the electron density was significantly weaker between main chain atoms C and C_α of Asn144 (**Figure S5a**). Because cleavage of the main chain is unlikely, we investigated the possibility of dual conformations at this position and indeed found that putting residues 143 to 146 in

alternate conformations explained the weakened electron density at this position (**Figure S5a**, **Figure 2a**). The maximum deviation between C_{α} positions is 1.02 Å for Asn144. In all, we examined ten datasets of ECFP crystals, and repeated the procedure. Each time, we observed weak electron density on the main chain at position Asn144 that could be explained by two conformations with variable occupancies. The two conformations of ECFP in our crystal structure were denoted *ecfp* and *ecfp'*. Both clearly differ from the previously-observed *gfp* conformation (**Figure 1b**),⁷ which we were unable to unambiguously identify in our crystals. We hypothesize that its absence originates from the significantly higher amount of PEG in our crystallization conditions, resulting in a more viscous mother liquor.

We used the same protocol to solve the structure of Cerulean, which is very close to that of ECFP. The main difference is found on the [143-150] strand, which could also be rebuilt in two alternate conformations (**Figure S5b**, **Figure 3a**), one closely resembling the *ecfp* conformation ('Outside' conformation of the main chain), and a new one in which the main chain is closer to the chromophore ('Inside' conformation of the main chain) (**Figure 1b**). These two conformations are called *ecfp* and *cerulean*, respectively. In the latter conformation, the side chain of Ile146 adopts a new conformation (**Figure S5d**) that is closer to the chromophore than in the *ecfp* conformation (**Figure 4**, **Figure S5c**).

In both ECFP and Cerulean, the chromophore is in the *cis, cis* configuration at physiological pH when the two partial double-bonds $C_{\alpha 2}-C_{\beta}$ and $C_{\beta}-C_{\gamma}$ are considered (**Figures S5c and d**). Its conformation is controlled by the surrounding amino acid residues, mostly through van der Waals (VdW) interactions (**Figure 4**). In particular, the indole ring of the chromophore interacts mainly with Thr203, but also with Val150 and Glu222 on one side of the chromophore plane. The other side of the chromophore is within VdW distance of Val61, Thr62, Phe 165 and Ile167 and, to a lesser extent, Ile146. There are also some side-on interactions, with His/Asp148, and one H-bond between the $N_{\epsilon 1}$ atom and O_{γ} from Ser205. The orientation of residue 148's side chain differs between ECFP and Cerulean. Asp148 is exposed to the solvent in Cerulean, whereas His148's side chain is parallel to the surface of the protein, which allows it

to be within VdW interaction distance of Ile146. The latter residue has a peculiar position in the *cerulean* conformation of Cerulean (**Figure 3a**), as it interacts preferentially with Val61, Ile167 and Thr203, thereby producing a tighter packing (**Figure 5**). The stabilization of the chromophore is ensured primarily by the pair of threonines 62 and 203, whose positions are remarkably symmetric with respect to the indole ring. Val150 and Phe165 have a similar cooperative role on the side of the chromophore. These two ‘clamps’ are largely responsible for preventing rotation of the indole ring with respect to the imidazolinone. The degree of chromophore planarity can be estimated by inspection of the values of the torsion angles Φ_1 and Φ_2 about the $C_{\alpha 2}$ - C_{β} and C_{β} - C_{γ} bonds, both of which have partial double-bond character (**Figure 1a**). As a result of the clamping, the chromophore is roughly planar, albeit only approximately when compared to the parent GFP and DsRed,³¹ and is slightly more planar in Cerulean than in ECFP (**Table 2**).

Although we identified a slightly different interaction between the protein and the chromophore in the crystal structures of ECFP and Cerulean, we could only hypothesize that the change in orientation of Ile146 and residue 148 would explain the improvement in QY. We postulated that molecular dynamics simulations based on our structures could unveil subtle details of the interactions between the chromophore and the protein matrix. Three molecular dynamics simulations of 39 ns duration were performed starting from the X-ray structures of ECFP in the *ecfp* conformation and Cerulean in both the *ecfp* and *cerulean* conformations. We did not expect the conversion of any of our structures to the *gfp* conformation, since the *ecfp* to *gfp* conversion should occur on the millisecond timescale.⁸ During the ECFP simulation, frequent switching between the *ecfp* and *ecfp'* conformations is observed (**Figure 2b**), but this local variation of dihedral angles between C_{α} atoms of Asn144 and Tyr145 is not correlated with significant changes in other residue interactions. Unexpectedly, a new conformation *ecfp''* is transiently observed (**Figure 3b, Movie S1**), in which Ile146 adopts its *cerulean* conformation, and the main chain of residues [143-150] its *ecfp* (or ‘Outside’) conformation (**Table 3**). For Cerulean, the *cerulean* conformation is preserved throughout the 39 ns simulation, whereas the *ecfp* conformation

converts to the *cerulean* one after only 9 ns (**Figure 3c, Movie S2**)) thereby suggesting that the *cerulean* conformation is the more stable one. Averaging energy terms between pairs of residues 66, 146 and 148 over the course of the simulation demonstrates that these three residues are engaged, on average, in cooperative attractive VdW interactions in ECFP (**Table 3**). The situation is markedly different in Cerulean as such a 3-partner interaction does not exist: there is a repulsive electrostatic energy term between Trp66 and Asp148, hardly any interaction between Ile146 and Asp148, and a strong VdW interaction between Trp66 and Ile146.

Monitoring of torsion angles during the molecular dynamics simulations shows that the chromophore retains a slightly twisted conformation in ECFP, and an almost perfectly planar conformation in Cerulean (**Table 3**). We have observed that the probability of a positive Trp66-Ile146 interaction energy is much higher in the *ecfp* conformation than in the *ecfp''/cerulean* conformations, *i.e.* it occurs when Ile146 interacts *via* its C_{γ2} atom with the chromophore (**Table 3, Figures 3d and 3e**). VdW interactions of positive energy correspond to a molecular collision event in which atoms that are too close repel each other. This implies that in the *ecfp* conformation, ECFP and Cerulean have a higher probability of undergoing collisional fluorescence quenching. In contrast, the conformation of Ile146 in *ecfp''* and *cerulean* favors a cohesive interaction with the chromophore (**Table 3, Figures 3d and 3e**). Finally, we have analyzed the interaction of Ile146 with the residues forming the binding pocket on one side of the chromophore's indole ring, namely Val61, Ile167 and Thr203 (**Figure 5**). These residues engage in VdW interactions that are 2kJ.mol⁻¹ more stabilizing in Cerulean than in ECFP when equivalent conformations are compared. In addition, Ile146 interacts more strongly with either Trp66 or Val161-Ile167-Thr203 in the *ecfp''* and *cerulean* conformations, suggesting a cooperative effect (**Table 3**).

To complete our study of the relationship between spectroscopic properties and structure of tryptophan-based chromophore fluorescent proteins, we investigated the role of Thr65, the first residue of the chromophore. The orientation of Thr65 can be related to the protonation state of Glu222 in ECFP

and Cerulean. In our structures, Thr65 is consistently H-bonded to Glu222, in contrast to the structure of S65T GFP mutants,¹⁴ and to the published ECFP structure,⁷ in which the hydroxyl group of Thr65 interacts with the carbonyl group of Val61. Conversely, Thr65 is rotated away from Glu222 in the low-pH structure of Cerulean.¹¹ For GFP it has been proposed that the interaction of the hydroxyl group of residue 65 is controlled by the protonation state of Glu222.³² If Glu222 is deprotonated, H-bonding to residue 65 will contribute to a decrease of the net charge of the carboxylate group, whereas if it is protonated, residue 65 will prefer to interact with Val61. Following this argument, Glu222 should be deprotonated in ECFP and Cerulean at physiological pH.

To extend our analysis of the role of Thr65, we have mutated it into alanine in Cerulean. The absorption and fluorescence emission peaks of Cerulean-T65A are slightly red-shifted (**Table 1, Figure S3**) whereas the EC is slightly higher than in ECFP and Cerulean, showing that the photon absorption cross-section is hardly affected by the mutation (**Table 1**). The QY is decreased to 0.37, which is intermediate between the QYs of ECFP and Cerulean, and the average fluorescence lifetime shows a similar trend (**Table 1**). We also wanted to see how a mutation located away from the Trp66/Ile146/Asp148 triad could affect the QY. Overall, the structure of Cerulean-T65A is similar to that of Cerulean with, most notably, retention of the *cerulean* conformation of the [143-150] strand that yields the strong interaction of Ile146 with the chromophore. We were also able to identify the *ecfp* conformation. In addition, the presence of strong, but discontinuous, residual density suggests that there is at least one more conformation, resembling the *gfp* one, but which we were unable to model unambiguously because its occupancy is low. Surprisingly, there are two conformations of the Gly67-Val68 peptide bond (**Figures 6 and S6**) that result in the coexistence of two chromophore conformations, one of which exhibits a ~ 0.7 Å in-plane displacement compared to the one adopted in Cerulean. The number of conformations of the [143-150] strand and of the chromophore, as well as the result of occupancy refinement, indicate that conformations of the chromophore and the [143-150] strand are not strongly correlated, and so we will not attempt to describe the nature and strength of

interactions about the indole ring but focus instead on the interactions with the imidazolinone ring. The absence of an H-bond donor atom in residue 65 forces Glu222 to adopt new conformations relative to the chromophore; in conformation **1**, Glu222 interacts directly with the conjugated system through an H-bond to atom N₂ of the imidazolinone ring (**Figure 6a**), and indirectly in conformation **2** with H-bonds to Ser205 and a water molecule (**Figure 6b**). As a consequence of these interactions, the chromophore is markedly tilted in conformation **1** and almost perfectly planar in conformation **2** (**Table 2**).

DISCUSSION

In this study we have attempted to relate the structures of tryptophan-based chromophore fluorescent proteins to their spectroscopic properties. ECFP is well-known for having two peaks in its absorption, fluorescence excitation and emission spectra and several hypotheses have been put forward to explain this. In particular it was proposed that it could be due to different structures of the protein,⁷ but it has been more recently suggested that this is in fact an intrinsic property of tryptophan-based chromophores.¹¹ The latter hypothesis is in agreement with the spectroscopic characteristics of Honeydew, a yellow-green fluorescent protein derived from the red FP DsRed, which harbors a tryptophan-based chromophore in a very different amino acid residue environment than ECFP.^{31, 33} In addition, we observe in our NaOH-driven denaturation experiments that the chromophore, most likely freed from its specific interactions with the protein, retains two peaks in its absorption spectrum (**Figure S1**). This constitutes an additional piece of evidence that this is a feature of tryptophan-based chromophores, which will not be eliminated by any set of mutations.

The critical mutation that gives ECFP its cyan light emission property is Y66W as this mutation alone suffices to turn Enhanced Green Fluorescent Protein (EGFP) into a CFP, although with a low level of fluorescence. The optimized ECFP contains the three additional mutations N146I, M153T and V163A, all found close to the chromophore, and presumably necessary for accommodating the increased bulk of

the indole ring when compared to the tyrosyl group of GFP.³⁴ Since M153T and V163A changes are present in other well-known green and yellow mutants such as Emerald, GFPuv or Venus, it is apparent that the replacement of a hydrophilic by a hydrophobic residue in the N146I mutation plays a pivotal role. In wild-type GFP, Asn146's side chain is completely exposed to the solvent and Tyr145 is buried in the vicinity of the chromophore, locked in place by a H-bond to a water molecule. If Ile146's side chain occupied the same position in ECFP as Asn146 in GFP, the protein would be destabilized due to a less favorable solvation. Energy minimization during protein folding potentially leads to a twist of the seventh β -strand that buries Ile146 in the interior of the protein, whereas Tyr145, thanks to its partial hydrophilic character, reorients toward the bulk solvent (**Figure 7a**) and is stabilized by H-bonds to two waters of the protein's first hydration shell. The new conformation of residues [143-150] modifies the number and nature of the interactions of the chromophore with the protein; in particular Ile146 and His148 are brought within VdW interaction distance of the indole ring of Trp66 from the chromophore. Indeed, our molecular dynamics simulations show that Trp66, Ile146 and His148 are engaged in VdW interactions and are likely to play a fundamental role in the fluorescence properties of ECFP (**Table 3**).

When considering the improvement of Cerulean vs. ECFP, we first checked what could be the importance of the S72A mutation that is known to improve folding efficiency.² We did not observe any detectable difference in the spectroscopic properties of ECFP-Y145A/H148D when compared to Cerulean (**Table 1**) and their structures are almost identical (data not shown). However, we observed a clear effect of the mutation in our unfolding experiments aimed at measuring T_m and EC. The T_m of Cerulean is increased by 3.3°C when compared to ECFP-Y145A/H148D and, whereas the latter is almost immediately unfolded by 0.5 M NaOH, it takes 3 minutes to get a fully solvent-exposed chromophore in Cerulean. This result is counterintuitive because the mutation essentially results in the replacement of a ~ 20 kJ.mol⁻¹ H-bond between Ser72 and Glu69, which is located on the helix bearing the chromophore, by a set of VdW interactions in Cerulean involving Ala72, Ile201 and Val224 and with a total energy of only ~ 5 kJ.mol⁻¹. Nevertheless, examination of the MD trajectories shows that

this H-bond is transiently disrupted, suggesting that it does not have a strong stabilizing role on the central helix. Consequently, it appears likely that the S72A mutation only plays a role in the conformational energy landscape during protein folding and unfolding, events that we cannot probe in our experiments.

The comparison of the [143-150] strand position in ECFP and Cerulean highlights the role of the Y145A mutation. In ECFP, we have observed two conformations *ecfp* and *ecfp'* that only differ by four residues; in Cerulean, we observe the *ecfp* conformation, and a new one, *cerulean*, in which the strand penetrates deeper into the protein. During the MD simulation, ECFP adopts a transient new conformation *ecfp''*, which shares similarities with the *ecfp* (for the position of the strand main chain), and *cerulean* conformations (for the orientation of Ile146's side chain). Examination of the trajectory shows that the hydroxyl group of Tyr145 is constantly H-bonded to water molecules of the solvent. In Cerulean, Ala145 cannot be solvated in the same way, hence the Y145A mutation likely favors the presence of the stable *cerulean* conformation which brings the main chain closer to the chromophore.

It has already been postulated that structural heterogeneity of the [143-150] strand has implications for the complex spectroscopic behavior of ECFP.⁷ We have now excluded that it is responsible for the two peaks in the absorption and fluorescence spectra. However, it is very likely that this structural heterogeneity, which is conserved in ECFP and Cerulean, is responsible for their complex fluorescence lifetime behaviors. Although Cerulean has primarily been assumed to be much improved in terms of monoexponential fluorescence decay, it has essentially two components, which is already an improvement when compared to ECFP.^{10, 28} Given that the [143-150] strand can significantly alter its position on the 1 to 10 ns timescale (**Figures 2 and 3**), it is likely that it can interact with the excited state of the chromophore, for instance through collisional quenching from Ile146, as previously mentioned. In the extreme case, it is possible that a molecule from the bulk solvent could occasionally enter the chromophore cavity and quench its fluorescence as well. Our structural analysis provides a

framework for analyzing the complexity of the fluorescence lifetime behavior of these tryptophan-based fluorescent proteins.

The low-pH structure of Cerulean raised questions about the real conformation of the chromophore at physiological pH.¹¹ Our results are unambiguous and uncover the structural response of Cerulean to a neutral-to-low pH transition (**Figures S7 and S8**). *Cis/trans* isomerization of the chromophore around the C_β-C_γ bond probably happens at physiological pH as an infrequent event. Upon acidification, the solvent-exposed residue Asp148 becomes protonated and provides a stabilizing interaction to the N_{ε1} atom of the isomerized chromophore. The dynamics of the [143-150] strand is altered, and it adopts a new conformation, intermediate between the *ecfp* and *cerulean* conformations. Due to the different environment, residues from the chromophore cavity must reorganize. This constitutes a pH switch, by which the protein can maintain fluorescence properties for a wide range of pH. The long timescale (hours) of the phenomenon underlines the infrequency of the isomerization.

After having identified structural differences between ECFP and Cerulean, we can explain why Cerulean exhibits an improved QY compared to ECFP. First, interactions between the chromophore and the surrounding residues control the planarity of the chromophore. It is widely accepted that everything else being equal, a more planar chromophore will have a better QY (**Table 2**).³¹ The combination of our crystallographic and molecular dynamics data shows that the chromophore is indeed more planar in Cerulean, although the difference of planarity between the two proteins is higher in the simulation than in the experimental structures. We postulate that this difference originates from the glassy ice environment in the crystal at 100K exerting a constraint on the [143-150] strand that is stronger than in the liquid environment at 300K modeled in our simulation. Indeed we observe that distances between this strand and the chromophore are consistently 0.2 Å shorter in the X-ray structures for this strand. In ECFP, Ile146 interacts with His148, which brings the latter residue within interaction distance of Trp66 and leads to the twisting of the indole ring compared to the rest of the chromophore. In contrast, Ile146

hardly interacts with Asp148 in Cerulean and the energy term between Asp148 and Trp66 is highly repulsive (**Table 3**). This demonstrates that the H148D mutation in Cerulean removes the interaction between residue 148 and the chromophore, allowing it to adopt a fully planar conformation.

Stabilization of the chromophore within the protein matrix favors radiative de-excitation pathways. The precise packing of amino acids against the chromophore plays a significant role in that regard. The only difference in chromophore environment between ECFP and Cerulean resides in the closer positioning of Ile146 in one of the Cerulean conformers. In this conformation, Ile146 can contribute to the stabilization of the chromophore through VdW interactions *via* atoms C_{γ2} and C_{δ1} (**Figure 3**). Ile146 rarely adopts that conformation in ECFP as our molecular dynamics simulation shows that Ile146 interacts with His148 during most of the simulation, thus hindering reorientation of its side chain toward the chromophore (**Movie S1**). This interaction is absent in Cerulean. Moreover, in the *cerulean* conformation, Ile146 establishes VdW interactions with Val61, Thr203 and Ile167, thereby filling a void in the vicinity of the chromophore (**Figure 5**). Calculation of the VdW interaction energy shows that the interaction between the four residues is favorable in that conformation (**Table 3**) and that they form a layer of atoms on one side of the chromophore that rigidifies its environment. We checked whether this would affect fluctuations of the Φ_i angles, thereby suggesting restriction of rotational movements of the chromophore as a reason for improvement of the QY.³⁵ Since we did not observe noticeable differences in these fluctuations, we specifically looked at the interaction between Ile146 and the chromophore and observed that a strong and stabilizing interaction was favored in the *cerulean* conformation of Cerulean, as well as in the infrequent *ecfp''* conformation of ECFP. By contrast, this interaction is repeatedly destabilizing ($E > 0$) in the *ecfp* conformation of both proteins (**Figures 3d** and **3e**). This can be interpreted as follows: in the *cerulean/ecfp''* conformation, Ile146 ‘breathes’ in a concerted manner with the chromophore, thanks to its strong interaction with Val61, Thr203 and Ile167. The change to the *ecfp* conformation drastically weakens the coupling between its movements and those of the chromophore, and Ile146 becomes a quencher of the chromophore’s excited state through

collision events. Evidently, the orientation of Ile146's side chain is controlled by the dynamics of the [143-150] strand, which implies that the improvement is only intermittent.

A more planar and better packed chromophore experiencing less collisional quenching constitutes a reasonable explanation for the better QY of Cerulean compared to ECFP (**Figure 7b**). Obviously, the QY could be higher if the [143-150] strand did not transiently adopt the *ecfp* conformation. It should be realized that the chromophore cavity has been optimized by nature for a tyrosine-based chromophore, which is symmetric, and stabilized by H-bonds to the hydroxyl group on the symmetry axis. For the tryptophan-based chromophore, the polar atom nitrogen is off the symmetry axis of the molecule, and can only form an H-bond with Ser205, leaving a large surface that can only be stabilized through VdW interactions. As a consequence, optimizing CFPs further might appear a daunting task, but the information that we present here concerning the specific interaction between the chromophore and the [143-150] strand should provide guidelines for mutagenesis.

The structure of Cerulean-T65A at neutral pH provides the picture of a tryptophan chromophore with two conformations of its terminal carbonyl group flipped 180° with respect to each other (**Figure 6**). Such a movement is possible only because residue 67 is a glycine in all fluorescent proteins. This is reminiscent of E2GFP, the T203Y mutant of EGFP, whose structure has been solved both at alkaline and acidic pH.^{36, 37} At pH 9.0, Gly67 adopts alternatively the two conformations, and the absorption spectrum of the protein shows two peaks corresponding to the protonated (A form) and non-protonated (B form) form of the chromophore.³⁷ At pH 5.0, it only adopts the unusual conformation and the absorption spectrum shows only the A-form peak.³⁶ Thus the equilibrium between the conformations appears to be controlled by the strongly-coupled pKa's of Glu222 and Tyr66 in the chromophore. In Cerulean-T65A, the chromophore cannot be ionized, and Glu222's pKa is probably the main parameter that controls the conformational equilibrium. In conformation **1** (**Figure 6a**), Glu222 must be protonated because it interacts directly with the N₂ atom of the imidazolinone ring, which is not expected to be

protonated.³⁸ In conformation **2** (**Figure 6b**), Glu222 is deprotonated, and cannot engage in an H-bond with the imidazolinone ring. Its negative charge is stabilized by H-bonds to Ser205 and a water molecule that is displaced by 1.8 Å from the position it occupies in conformation **1**. Reorganization of the H-bond network is caused by 180° rotation of Gly67's carbonyl group. The resulting conformational equilibrium constitutes the protein's response to loss of the stabilizing hydroxyl group of residue 65. As a consequence, two chromophore populations coexist. In the first conformation, Glu222 exerts a torque on the imidazolinone ring *via* a H-bond and the chromophore is significantly twisted. In the second, the torque is canceled, and the chromophore has an almost planar conformation. The combination of the two conformations, one planar and one twisted, provides a rationalization to why the QY of Cerulean-T65A shows an intermediate value between those of ECFP and Cerulean at physiological pH. This structure highlights the flexibility of the ensemble composed by the tryptophan-based chromophore and its protein matrix, which should clearly not be thought of as a rigid structure when rationally designing mutants.

CONCLUSION

We have shown that the tryptophan-based chromophore forms a very dynamic ensemble with the [143-150] strand in the fluorescent proteins ECFP and Cerulean. VdW interactions between Ile146, residue 148 and the indole ring of the chromophore control its planarity and stabilization, thus modulating its fluorescence quantum yield. The latter can also be affected by interactions with the imidazolinone ring of the chromophore, as exemplified by Cerulean-T65A. These data provides new insights to further improve the fluorescence properties of tryptophan-based fluorescent proteins. An ideal system would contain mutations that permanently stabilize Ile146 in contact with the chromophore.

ACKNOWLEDGMENT We thank Stephen Adams, Philippe Carpentier, Juan Fontecilla-Camps, Dominique Madern, Eve de Rosny, Xiakun Shu, Roger Tsien and Martin Weik for fruitful discussions and/or technical help. We also thank the ESRF for providing beamtime, members of the ESRF

macromolecular crystallography group for their continuous support, and the staff at CCRT-CEA for technical assistance. Support by the Agence Nationale de la Recherche (ANR-07-BLAN-0107-01) is acknowledged.

ABBREVIATIONS FRET: Förster Resonance Energy Transfer; MD: Molecular Dynamics; QY: fluorescence quantum yield; VdW: van der Waals; ECFP: Enhanced Cyan Fluorescent Protein; EGFP: Enhanced Green Fluorescent Protein; CFP: Cyan Fluorescent Protein; GFP: Green Fluorescent Protein; H-bond: Hydrogen bond; FP: fluorescent protein; T_m : melting temperature; rms: root mean square; conf.: conformation; PEG: polyethylene glycol; RI: refractive index; TCSPC: time-correlated single-photon-counting; EC: Extinction coefficient.

SUPPORTING INFORMATION AVAILABLE.

Crystallography data table, additional figures and MD simulation movies.

FIGURE CAPTIONS

Figure 1. a. Schematic representation of the chromophore in ECFP and Cerulean with atom numbering. R and R' are the normal continuation of the protein main chain through Leu64 and Val68, respectively. Red lines indicate individual amino-acid residues within the chromophore. Dihedral angles Φ_1 and Φ_2 around the aryl-alkene bond are indicated in blue. **b.** Representation of ECFP's three-dimensional structure in gold as determined in this study. The mainchain trace of the [141-151] strand of its *ecfp* conformation is represented in cyan. For comparison purposes, the *gfp* and *cerulean* conformations of the strand are represented in green and deep blue, respectively, as they are observed in the structures of EGFP¹⁴ and Cerulean (this study). All figures were prepared with PyMol (DeLano Scientific).

Figure 2. Representation of residues 143 to 146 in the *ecfp* (solid cyan) and *ecfp'* (transparent cyan) conformations of the crystal structure of ECFP. Dihedral angles Ψ and Φ are defined on each side of the Asn144-Tyr145, peptide bond. $(\Psi, \Phi) = (162.1^\circ, -48.5^\circ)$ and $(-78.6^\circ, -160.2^\circ)$ for the *ecfp* and *ecfp'*

conformations, respectively. **b.** Evolution of Ψ (red trace) and Φ (green trace) during the molecular dynamics simulation starting from the *ecfp* conformation of ECFP showing that it transiently adopts the *ecfp'* conformation.

Figure 3. a. Representation of residues 66 and 144 to 148 in the *cerulean* (solid blue) and *ecfp* (transparent blue) conformations of the crystal structure of Cerulean. Distances d_1 and d_2 between Trp66 atom $C_{\gamma 2}$ and either atom $C_{\gamma 2}$ or atom $C_{\delta 1}$ of Ile146 are depicted in red and green, respectively. Evolution of d_1 (green) and d_2 (red) during molecular dynamics simulations started from the *ecfp* x-ray conformations of ECFP (**b**), and Cerulean (**c**). These distances served to monitor the *ecfp* \leftrightarrow *ecfp''* and *ecfp* \leftrightarrow *cerulean* transitions in ECFP and Cerulean, respectively. $(d_1, d_2) = (3.6 \text{ \AA}, 5.4 \text{ \AA})$ for the *ecfp* conformation in the crystal structure of ECFP; $(d_1, d_2) = (3.8 \text{ \AA}, 6.0 \text{ \AA})$ and $(6.0 \text{ \AA}, 3.8 \text{ \AA})$ for the *ecfp* and *cerulean* conformations in the crystal structure of Cerulean. Interaction energy terms between Tyr66 and Ile146's side chain for ECFP (**d**), and Cerulean (**e**). This shows that the *ecfp''/cerulean* conformation of Ile146 is in a better interaction with the chromophore in both ECFP and Cerulean, resulting in a significantly lower probability of collisional quenching.

Figure 4. Environment of the indole ring of the chromophore in ECFP and Cerulean. The main two H-bonds stabilizing the chromophore are depicted in orange. Distances shorter than 4.0 \AA illustrating potential VdW interactions are depicted in black. The critical set of VdW interactions between Trp46, Ile146, His148 is depicted in purple. The two residues in Cerulean showing a different interaction with the chromophore, Ile146 and Asp148, are depicted in deep blue.

Figure 5. Comparison of chromophore environment around Ile146 in ECFP and Cerulean. Atoms are depicted as spheres with their VdW radii. **a.** In ECFP, Ile146 is clearly distant from Thr62, leaving an empty cavity in the immediate vicinity of the chromophore. **b.** Ile146 in the *cerulean* conformation of Cerulean. The complementarity of the Ile146 side chain to those of Val61, Thr62 and Ile167 indicates that these four residues form VdW interactions and provide an enhanced stabilization to the chromophore, which is supported by energy calculations.

Figure 6. Dual conformations of the chromophore in Cerulean-T65A. **a.** standard conformation of the carbonyl group of Gly67, and direct interaction between Glu222 and the imidazolinone ring through H-bonding. The black arrow indicates the torsion of the imidazolinone ring. **b.** Alternate conformation in which the carbonyl group is flipped 180° from its usual position, Glu222 is not in direct interaction with the chromophore, and the chromophore is translated by ~0.7 Å towards the solvent.

Figure 7. a. Comparison of the interaction of the chromophore with the [143-150] strand in EGFP, ECFP and Cerulean. Side chains are truncated to their C_β positions, except for residues 145 and 146. Names of hydrophobic, and hydrophilic, residues are drawn in black and blue, respectively. Tyrosines, due to their partial hydrophilic character, have their name in purple. The chromophore is depicted in black. Hydrophobic residues are all directed toward the interior of the protein. Hydrophilic residues are either parallel to, or pointing outside, the protein surface. This scheme illustrates the impact of the hydrophilic to hydrophobic N146I mutation between EGFP and ECFP/Cerulean and also shows that the strand enters into the protein's core in the *cerulean* conformation than in the *ecfp* one. **b.** Schematic representation of the interactions controlling the fluorescence properties of the chromophore in ECFP, Cerulean and Cerulean-T65A. The chromophore is stabilized by neighboring residues through van der Waals interactions (half-circles) and one H-bond (dashed line). Interactions potentially affecting chromophore planarity are represented as double-arrows. Potential collisional quenching of the chromophore's fluorescence by Ile146 is shown by a red star. Hydrophilic residues and polar interactions are depicted in blue; hydrophobic residues and VdW interactions are shown in black and brown respectively.

TABLES.

Table 1. Spectroscopic and biochemical characterization of ECFP mutants.

Protein	ECFP	ECFP-Y145A/H148D	Cerulean	Cerulean-T65A
Absorption maximum (nm)	435	434	434	437
Mol. extinct. coeff. ($\text{mM}^{-1}\text{cm}^{-1}$)	30.0 +/- 0.4 (n=6) [†]	30.2 +/- 0.3 (n=4)	30.7 +/- 0.3 (n=5)	32.2 +/- 0.1 (n=4)
Emission maximum (nm)	474	474	474	478
Quantum yield (-)	0.30 +/- 0.01 (n=6)	0.44 +/- 0.01 (n=10)	0.44 +/- 0.01 (n=8)	0.37 +/- 0.01 (n=7)
Brightness (-)	9.0	13.4	13.7	11.8
Average fluorescence lifetime in solution (ns)	3.2 +/- 0.1 (n=23)	3.7 +/- 0.2 (n=6)	3.8 +/- 0.2 (n=5)	3.6 +/- 0.2 (n=7)
Average fluorescence lifetime in crystal (ns)	2.2 +/- 0.2 (n=3)	2.8 +/- 0.3 (n=3)	2.7 +/- 0.4 (n=7)	2.7 +/- 0.3 (n=7)
Denaturation temperature ($^{\circ}\text{C}$)	75.7 +/- 0.1 (n=3)	82.0 +/- 0.4 (n=3)	85.3 +/- 0.4 (n=3)	84.8 +/- 0.2 (n=3)

[†]n is the number of individual measurements.

Table 2. Torsion angles of the chromophore in CFP and other fluorescent proteins.

Protein	ECFP	Cerulean	Cerulean-T65A		GFP	Citrine	DsRed	Rtms5
			conf. 1	conf. 2				
PSB Entry code	XXXX	YYYY	ZZZZ		1W7S	1HUY	1G7K	1MOV
QY (-)	0.30	0.44	0.37		0.67	0.71	0.68	0.00
Φ_1 ($^{\circ}$)	7.5	6.6	2.4	13.9	0.2	0.8	0.9	-11.5
Φ_2 ($^{\circ}$)	-7.4	-6.9	2.2	-7.0	-0.2	-0.4	0.1	42.4

(Φ_1 , Φ_2): chromophore torsion angles around bonds $\text{C}_{\alpha 2}\text{-C}_{\beta}$ and $\text{C}_{\beta}\text{-C}_{\gamma}$

Table 3. Statistics from MD simulations of the observed conformations of ECFP and Cerulean. The three sections refer to parameters relevant to: (i) the conformation of Ile146, (ii) the interaction between Ile146 and Trp66, (iii) chromophore planarity.

Protein Conformation	ECFP		Cerulean	
	<i>ecfp</i>	<i>ecfp''</i>	<i>ecfp</i>	<i>cerulean</i>
Conformation of [143-150] strand	Outside	Outside	Outside	Inside
$\langle E_{I146-res148} \rangle$ (kJ.mol ⁻¹) (type)	-1.6 (VdW)	-1.8 (VdW)	-0.5 (both)	-0.1 (both)
$\langle E_{I146-3AA} \rangle$ (kJ.mol ⁻¹) (type)	-8.3 (VdW)	-8.7 (VdW)	-9.5 (VdW)	-10.5 (VdW)
$\langle E_{I146-W66} \rangle$ (kJ.mol ⁻¹) (type)	-4.0 (VdW)	-6.0 (VdW)	-4.4 (VdW)	-6.4 (VdW)
Probability of $E_{I146-W66} > 0$ (‰)	21	3	28	1
$\langle E_{W66-res148} \rangle$ (kJ.mol ⁻¹) (type)	-1.5 (VdW)	-2.4 (VdW)	8.3 (elec)	8.0 (elec)
$\langle \Phi_1 \rangle$ (°)	-8.8	-11.1	-1.6	-0.7
$\langle \Phi_2 \rangle$ (°)	-1.4	-1.4	0.7	0.6

$\langle E_{res1-res2} \rangle$: Mean interaction energy between residues 1 and 2 - VdW: van der Waals interactions, elec: electrostatic interactions - 3AA: Val61, Ile167 and Thr203.

REFERENCES

1. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994**, 263, (5148), 802-5.
2. Cubitt, A. B.; Woollenweber, L. A.; Heim, R., Understanding structure-function relationships in the *Aequorea victoria* green fluorescent protein. *Methods Cell Biol* **1999**, 58, 19-30.
3. Tsien, R. Y., The green fluorescent protein. *Annu Rev Biochem* **1998**, 67, 509-44.
4. Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y., A guide to choosing fluorescent proteins. *Nat Methods* **2005**, 2, (12), 905-9.
5. Borst, J. W.; Hink, M. A.; van Hoek, A.; Visser, A. J., Effects of refractive index and viscosity on fluorescence and anisotropy decays of enhanced cyan and yellow fluorescent proteins. *Journal of Fluorescence* **2005**, 15, (2), 153-160.
6. Nguyen, A. W.; Daugherty, P. S., Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **2005**, 23, (3), 355-60.
7. Hyun Bae, J.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N., Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J Mol Biol* **2003**, 328, (5), 1071-81.
8. Seifert, M. H.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A., Slow exchange in the chromophore of a green fluorescent protein variant. *J Am Chem Soc* **2002**, 124, (27), 7932-42.
9. Rizzo, M. A.; Springer, G. H.; Granada, B.; Piston, D. W., An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol* **2004**, 22, (4), 445-9.
10. Villoing, A.; Ridhoir, M.; Cinquin, B.; Erard, M.; Alvarez, L.; Vallverdu, G.; Pernot, P.; Grailhe, R.; Merola, F.; Pasquier, H., Complex Fluorescence of the Cyan Fluorescent Protein: Comparisons with the H148D Variant and Consequences for Quantitative Cell Imaging. *Biochemistry* **2008**.
11. Malo, G. D.; Pouwels, L. J.; Wang, M.; Weichsel, A.; Montfort, W. R.; Rizzo, M. A.; Piston, D. W.; Wachter, R. M., X-ray structure of Cerulean GFP: a tryptophan-based chromophore useful for fluorescence lifetime imaging. *Biochemistry* **2007**, 46, (35), 9865-73.
12. Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W., Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **1990**, 185, 60-89.
13. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R., High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001**, 6, (6), 429-40.
14. Jain, R. K.; Ranganathan, R., Local complexity of amino acid interactions in a protein core. *Proc Natl Acad Sci U S A* **2004**, 101, (1), 111-6.
15. Kabsch, W., Automatic processing of rotation diffraction data from crystals of internally unknown symmetry and cell constants. *Journal of Applied Crystallography* **1993**, 26, 795-800.
16. Leslie, A. G., The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **2006**, 62, (Pt 1), 48-57.
17. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **1997**, 53, (Pt 3), 240-55.
18. Ward, W. W., In *Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical Applications*, DeLuca, M. a. M., D.W., Ed. Academic Press: New York, 1981; pp 235-242.
19. Royant, A.; Carpentier, P.; Ohana, J.; McGeehan, J.; Paetzold, B.; Noirclerc-Savoye, M.; Vernède, X.; Adam, V.; Bourgeois, D., Advances in spectroscopic methods for biological crystals. 1. Fluorescence lifetime measurements. *Journal of Applied Crystallography* **2007**, 40, 1105-12.
20. Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A., The Dynamo Library for Molecular Simulations using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *Journal of Computational Chemistry* **2000**, 21, 1088-1100.

21. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, 118, (45), 11225-11236.
22. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, 79, 926-935.
23. Demachy, I.; Ridard, J.; Laguitton-Pasquier, H.; Durnerin, E.; Vallverdu, G.; Archirel, P.; Levy, B., Cyan fluorescent protein: molecular dynamics, simulations, and electronic absorption spectrum. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys* **2005**, 109, (50), 24121-33.
24. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K., Prediction of pH-dependent properties of proteins. *J Mol Biol* **1994**, 238, (3), 415-36.
25. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, 81, 3684-3690.
26. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, 23, (3), 327-341.
27. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, 98, (12), 10089-10092.
28. Kremers, G. J.; Goedhart, J.; van Munster, E. B.; Gadella, T. W., Jr., Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Forster radius. *Biochemistry* **2006**, 45, (21), 6570-80.
29. Suhling, K.; Siegel, J.; Phillips, D.; French, P. M.; Leveque-Fort, S.; Webb, S. E.; Davis, D. M., Imaging the environment of green fluorescent protein. *Biophys J* **2002**, 83, (6), 3589-95.
30. Voros, J., The density and refractive index of adsorbing protein layers. *Biophys J* **2004**, 87, (1), 553-61.
31. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006**, 45, (32), 9639-47.
32. Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J., Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc Natl Acad Sci U S A* **1997**, 94, (6), 2306-11.
33. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* **2004**, 22, (12), 1567-72.
34. Heim, R.; Tsien, R. Y., Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* **1996**, 6, (2), 178-82.
35. Megley, C. M.; Dickson, L. A.; Maddalo, S. L.; Chandler, G. J.; Zimmer, M., Photophysics and dihedral freedom of the chromophore in yellow, blue, and green fluorescent protein. *J Phys Chem B* **2009**, 113, (1), 302-8.
36. Arosio, D.; Garau, G.; Ricci, F.; Marchetti, L.; Bizzarri, R.; Nifosi, R.; Beltram, F., Spectroscopic and structural study of proton and halide ion cooperative binding to gfp. *Biophys J* **2007**, 93, (1), 232-44.
37. Bizzarri, R.; Nifosi, R.; Abbruzzetti, S.; Rocchia, W.; Guidi, S.; Arosio, D.; Garau, G.; Campanini, B.; Grandi, E.; Ricci, F.; Viappiani, C.; Beltram, F., Green fluorescent protein ground states: the influence of a second protonation site near the chromophore. *Biochemistry* **2007**, 46, (18), 5494-504.
38. Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J., Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy. *Biochemistry* **2000**, 39, (15), 4423-31.

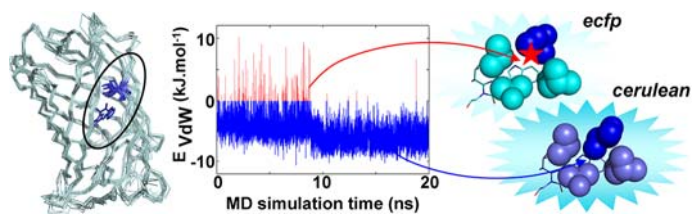


Figure 1.

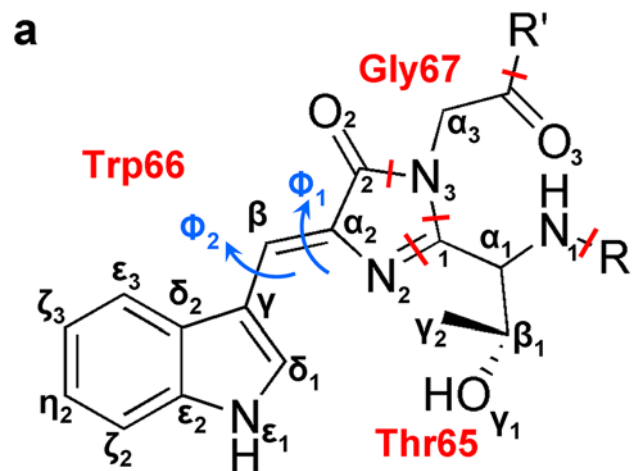


Figure 2.

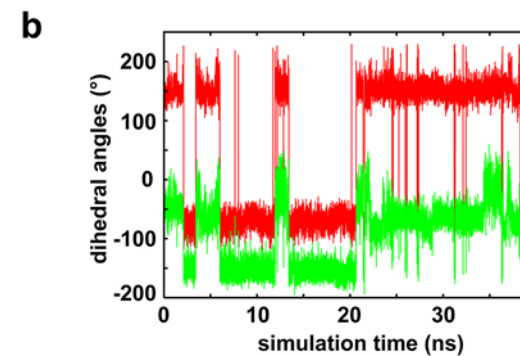
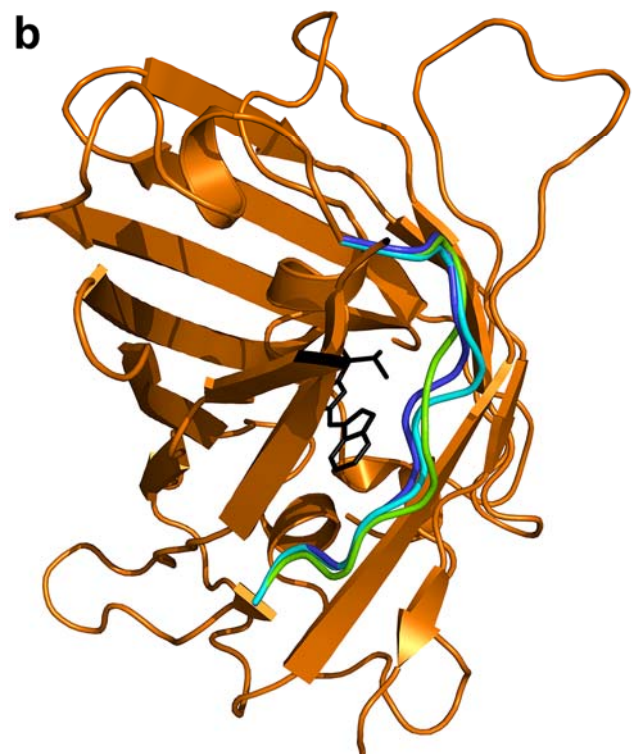
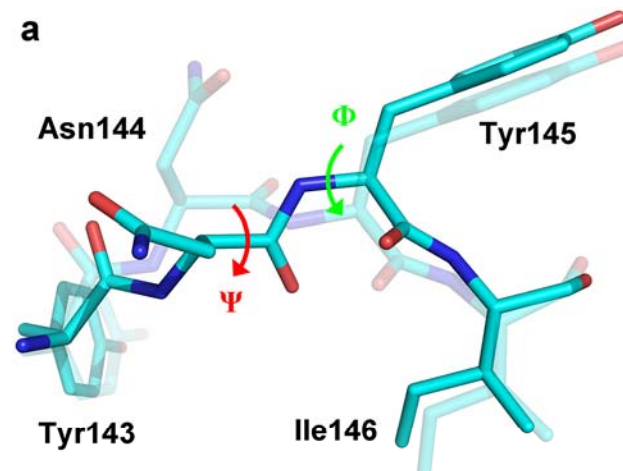


Figure 3.

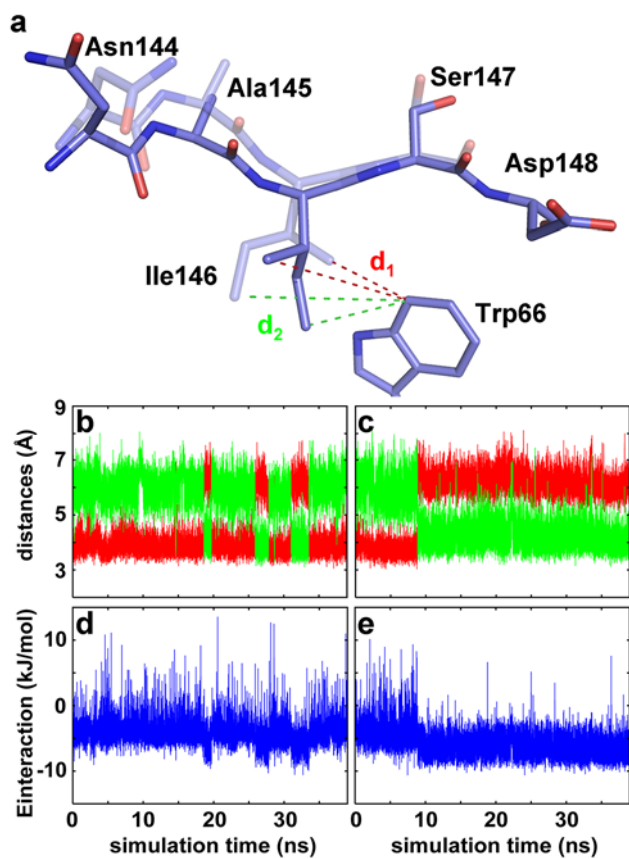


Figure 4.

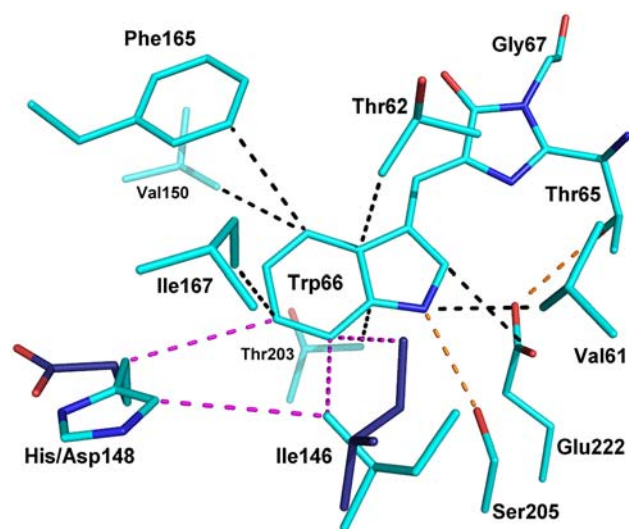


Figure 5.

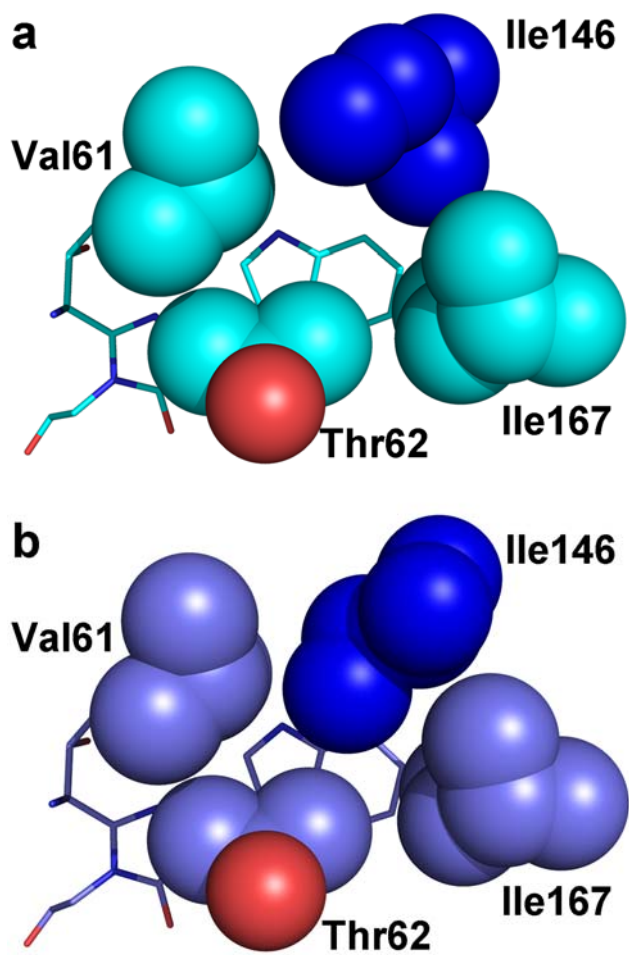


Figure 6.

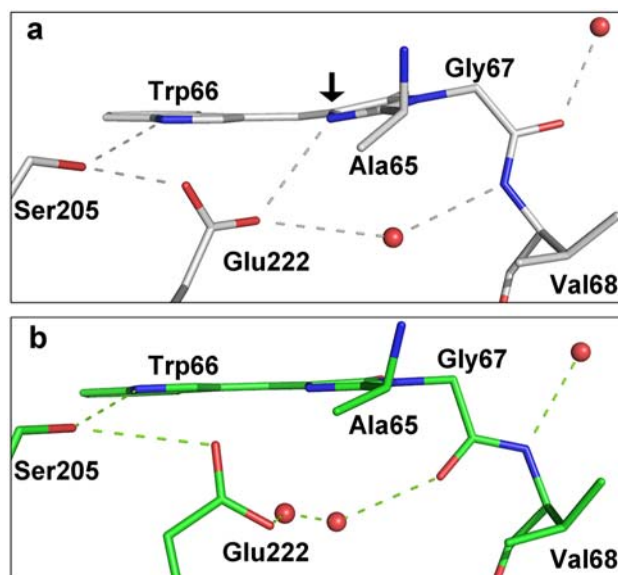
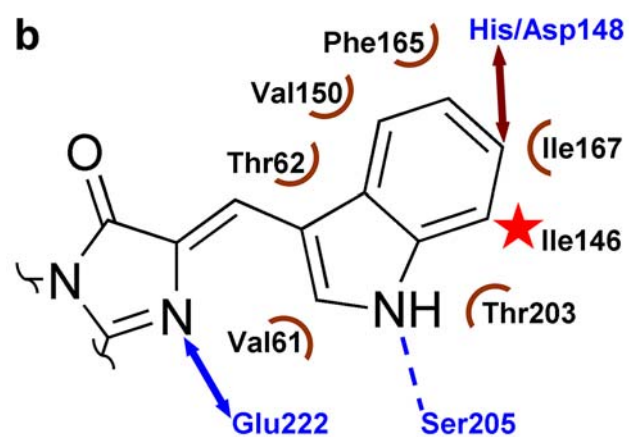
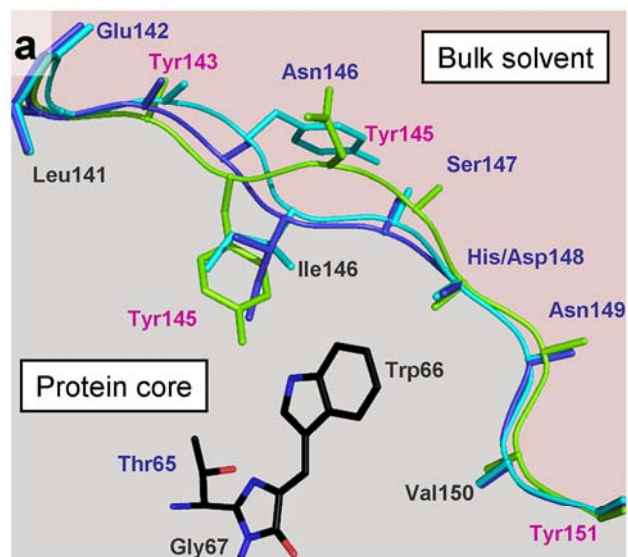


Figure 7.



SUPPORTING INFORMATION TO:

Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield

*Mickaël Lelimosin, Marjolaine Noirclerc-Savoie, Christelle Lazareno-Saez, Bernhard Paetzold,
Sophie Le Vot, Richard Chazal, Pauline Macheboeuf, Martin J. Field, Dominique Bourgeois, and
Antoine Royant*

Institut de Biologie Structurale Jean Pierre Ebel & European Synchrotron Radiation Facility, Grenoble,
France

Table S1

Figures S1 to S8

Captions for Movies S1 and S2

Table S1. Crystallographic data for ECFP, Cerulean and Cerulean-T65A structures.

	ECFP	Cerulean	Cerulean-T65A
Data collection			
Cell parameters	51.2, 62.7, 70.6	51.0, 62.8, 69.5	51.2, 62.9, 69.5
Beamline	ID14-eh1	ID29	ID23-eh2
Wavelength (Å)	0.933	0.984	0.873
Resolution limits (Å)	40-1.37	50-1.15	40-1.27
Completeness [‡] (%)	94.9 (70.4)	97.7 (92.1)	100.0 (100.0)
Multiplicity	4.7 (3.7)	3.6 (3.2)	3.6 (3.3)
<I/σ(I)>	22.3 (3.3)	20.4 (2.8)	11.3 (2.1)
R _{sym} [†] (%)	4.6 (41.4)	4.1 (45.4)	7.0 (52.8)
Unique reflections	45997	78022	60026
Structure refinement			
R _{work} [‡] (%)	15.6 (23.3)	16.6 (26.9)	16.2 (26.5)
R _{free} [§] (%)	18.5 (27.2)	19.0 (28.8)	18.9 (28.8)
Bond length rmsd (Å)	0.009	0.008	0.010
Bond angle rmsd (Å)	1.43	1.37	1.46
Average B-factor (Å ²)	9.5	8.8	9.7
Number of non-H atoms	2197	2242	2277

[‡]Values in parentheses refer to the highest resolution shell.

$$^{\dagger}R_{\text{sym}} = \frac{\sum_h \sum_j |I_{h,j} - \langle I_h \rangle|}{\sum_h \sum_j I_{h,j}}$$

$$^{\ddagger}R_{\text{work}} = \frac{\sum_h ||F_{\text{obs}}(\mathbf{h})| - |F_{\text{calc}}(\mathbf{h})||}{\sum_h |F_{\text{obs}}(\mathbf{h})|}$$

[§]R_{free} is calculated from a set of 5% randomly-selected reflections that were excluded from refinement.

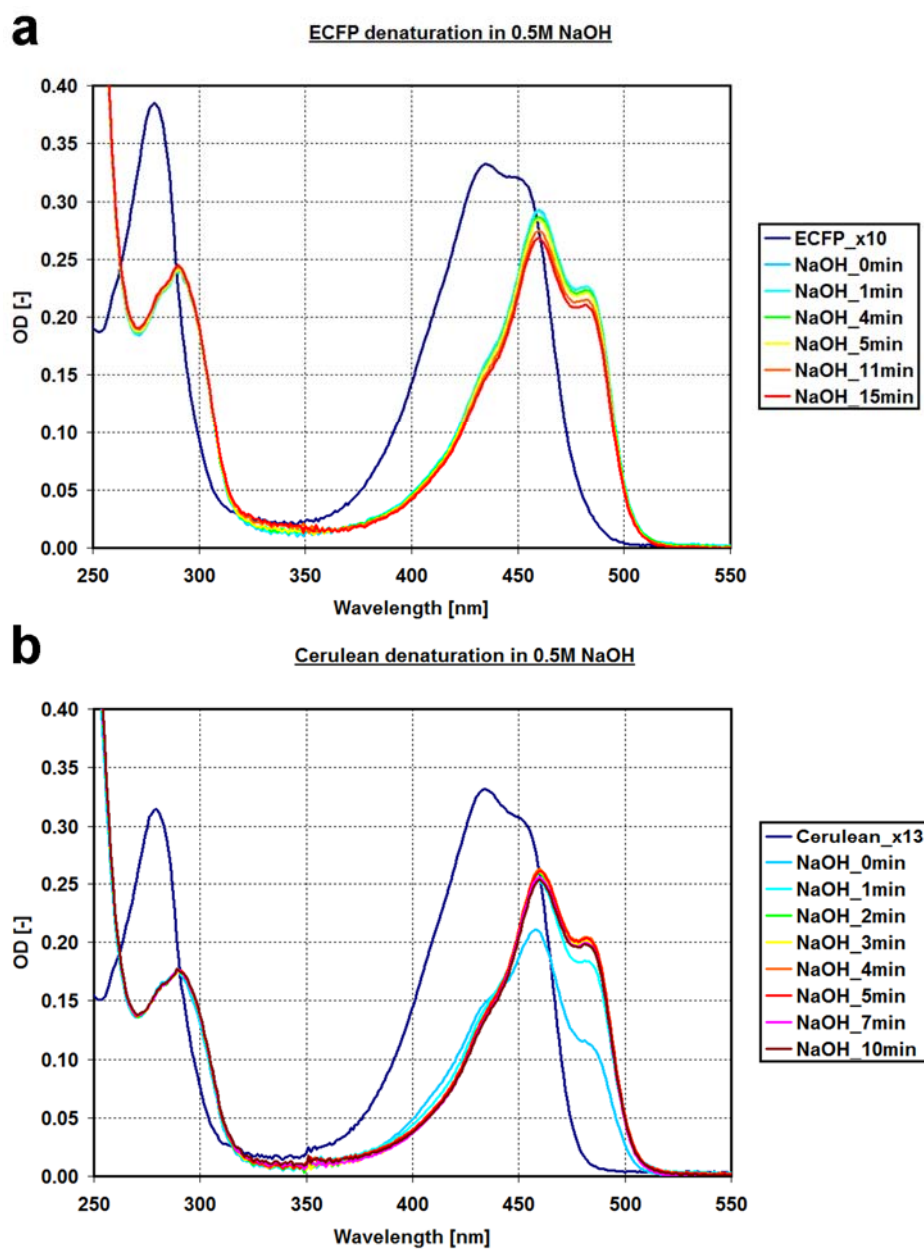


Figure S1. Determining molar extinction coefficients for the tryptophan-based fluorescent proteins ECFP (**a**) and Cerulean (**b**). After denaturation by NaOH, the chromophore freed from its interactions with the protein absorbs maximally at 460 nm, with a second peak at 482 nm. The maximal absorption is reached immediately for ECFP, and after 3 minutes for Cerulean.

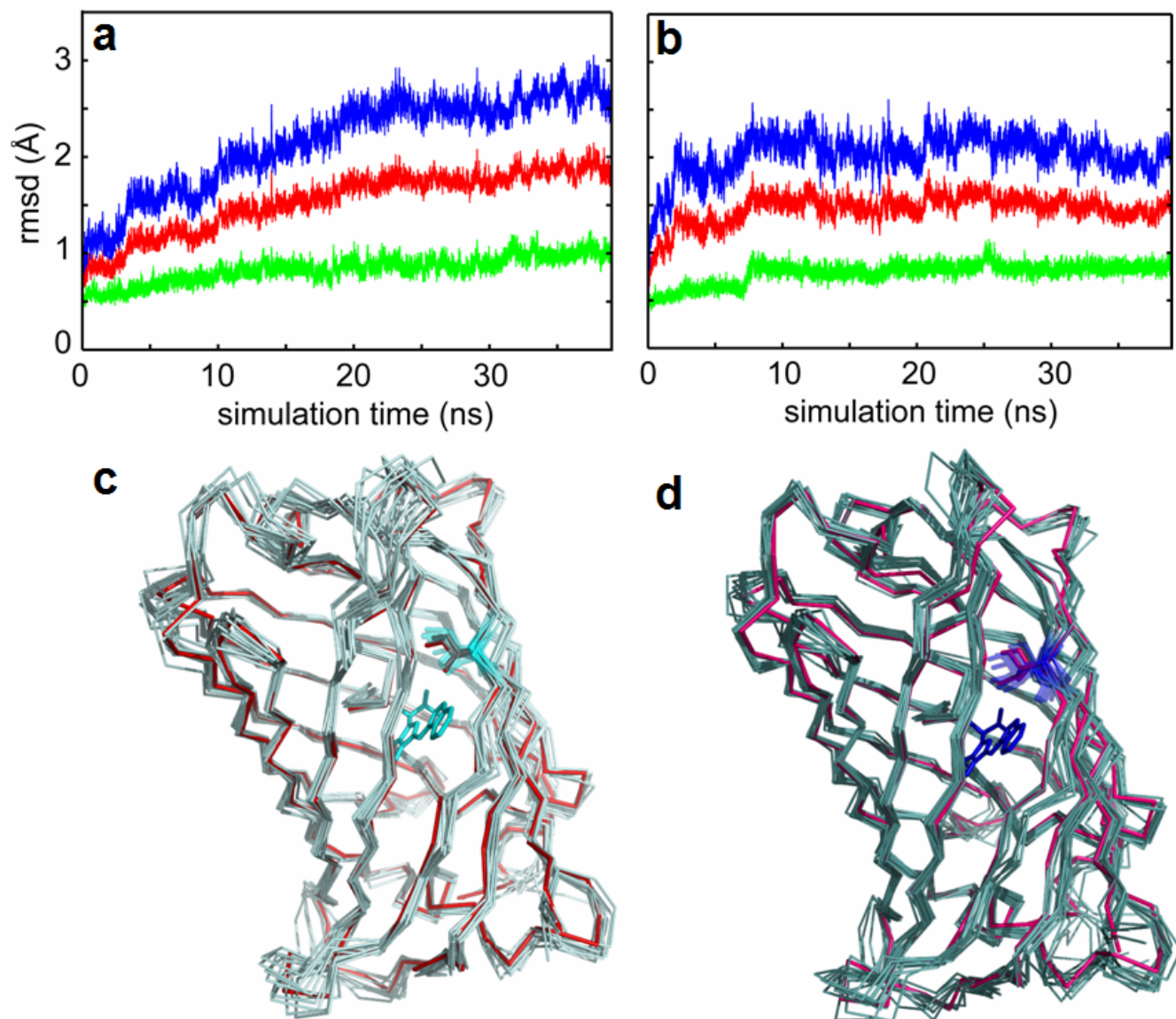


Figure S2. **a.** Root mean square deviation of C_{α} positions from X-ray structures during simulation of conformation *ecfp* of ECFP. Green line: C_{α} of central helix and β -barrel, Blue line: remaining C_{α} on loops, Red line: all C_{α} of the protein. **b.** *idem* for conformation *ecfp* of Cerulean. **c.** Superposition of 10 structures (light blue) extracted every 2 ns from the first simulation (from 0 to 18 ns) on the crystallographic structure of ECFP (red) highlighting the conformation of Ile146's side chain compared to the chromophore. **d.** Superposition of 10 structures (teal blue) extracted every 2 ns from the second simulation (from 0 to 18 ns) on the crystallographic structure of ECFP (purple).

ECFP, Cerulean and Cerulean-T65A spectra at RT

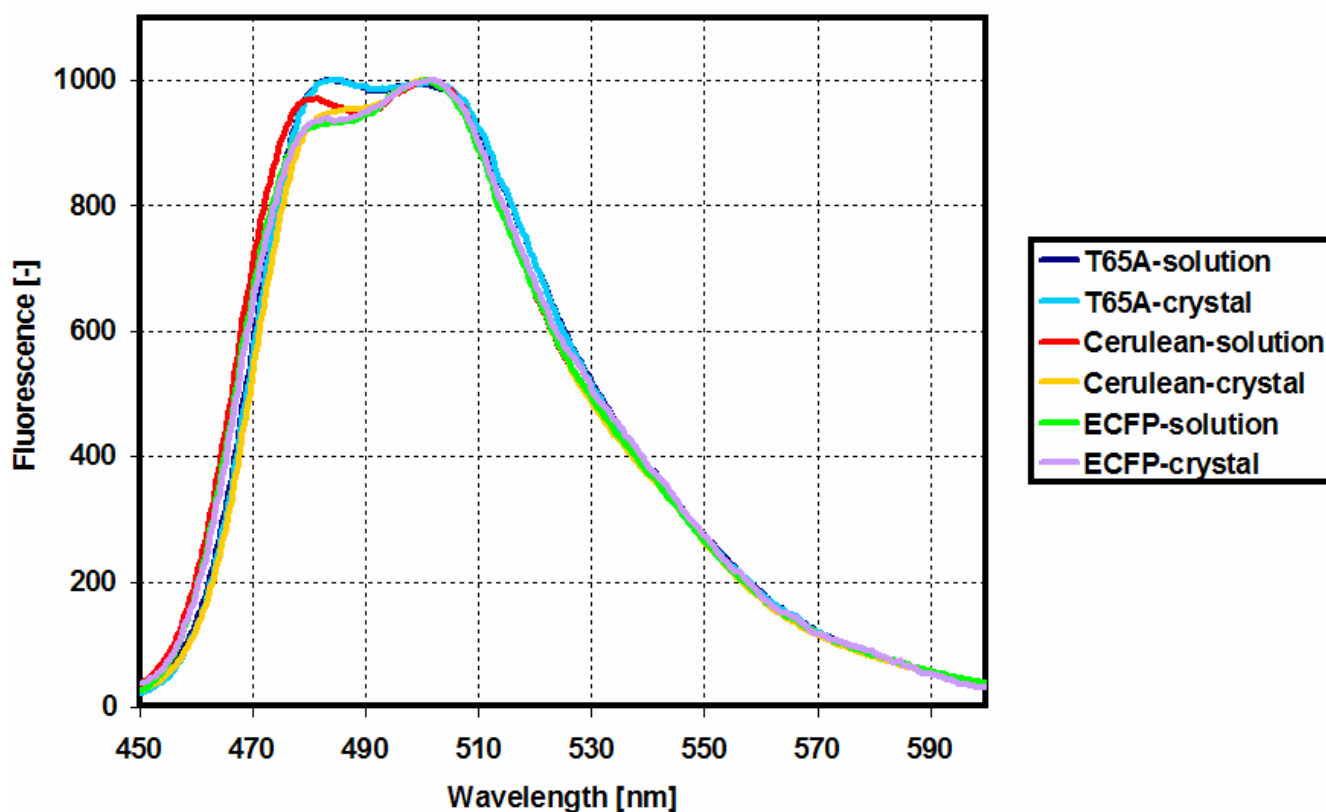


Figure S3. Fluorescence emission spectrum for ECFP, Cerulean and Cerulean-T65A in solution and in a crystal. The fluorescence properties of ECFP mutants are similar in the dilute solute state (samples with $OD < 0.05$, i.e. with less than $2 \mu\text{M}$ protein), and the very concentrated hydrated solid state (crystals of $\sim 30 \text{ mM}$). The difference in the blue part of the Cerulean spectra can be explained by auto-absorption of the fluorescence by the bulk of the massive Cerulean crystal.

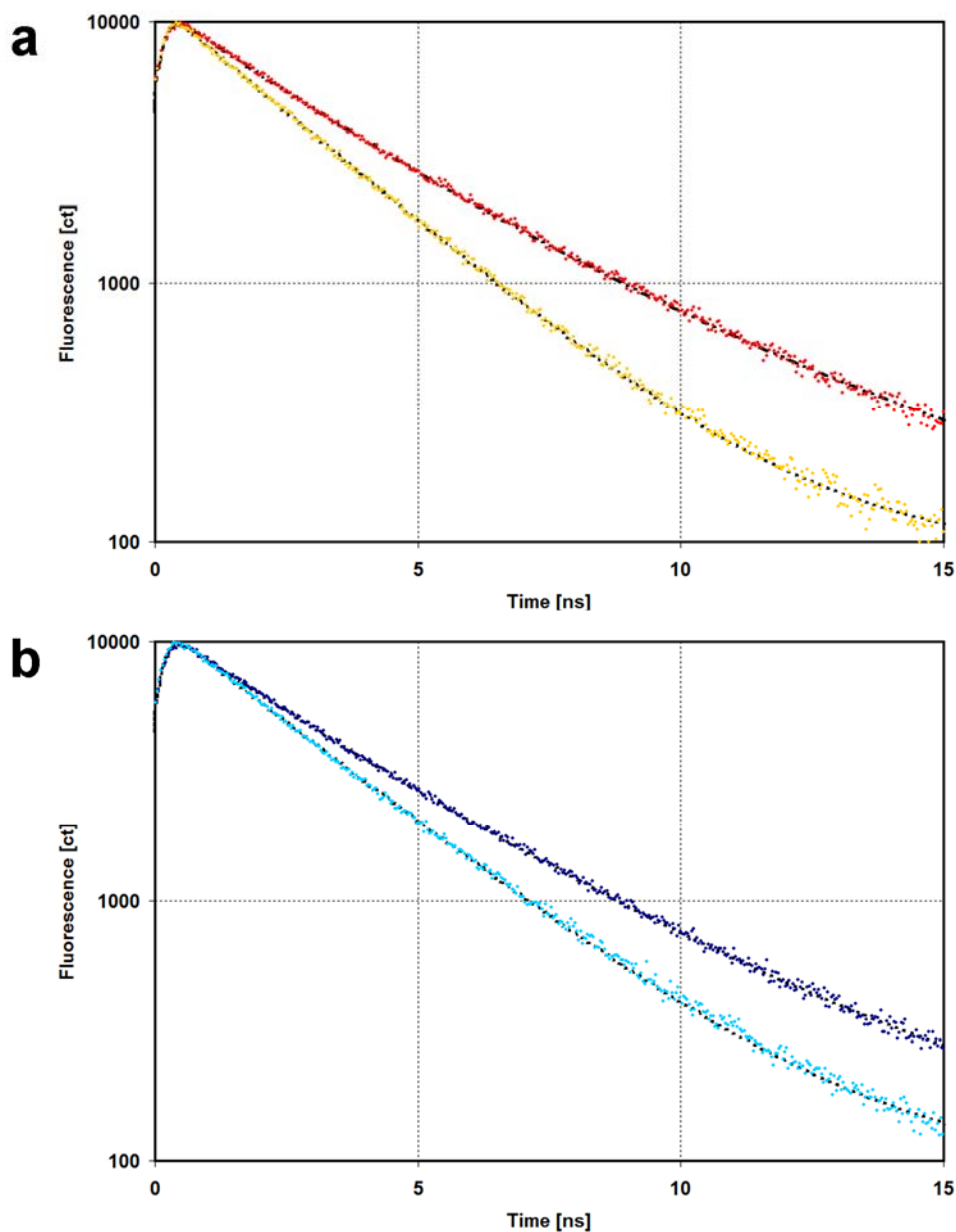


Figure S4. Fluorescence lifetime measurements for (a) Cerulean (solution in red, crystal in yellow) and (b) Cerulean-T65A (solution in dark blue, crystal in light-blue). The decrease of the fluorescence lifetime can be explained by the increase of refractive index between pure water and the crystal. Together with the steady-state fluorescence emission spectra, these results show that our fluorescent proteins have essentially retained their fluorescence properties in the crystalline state.

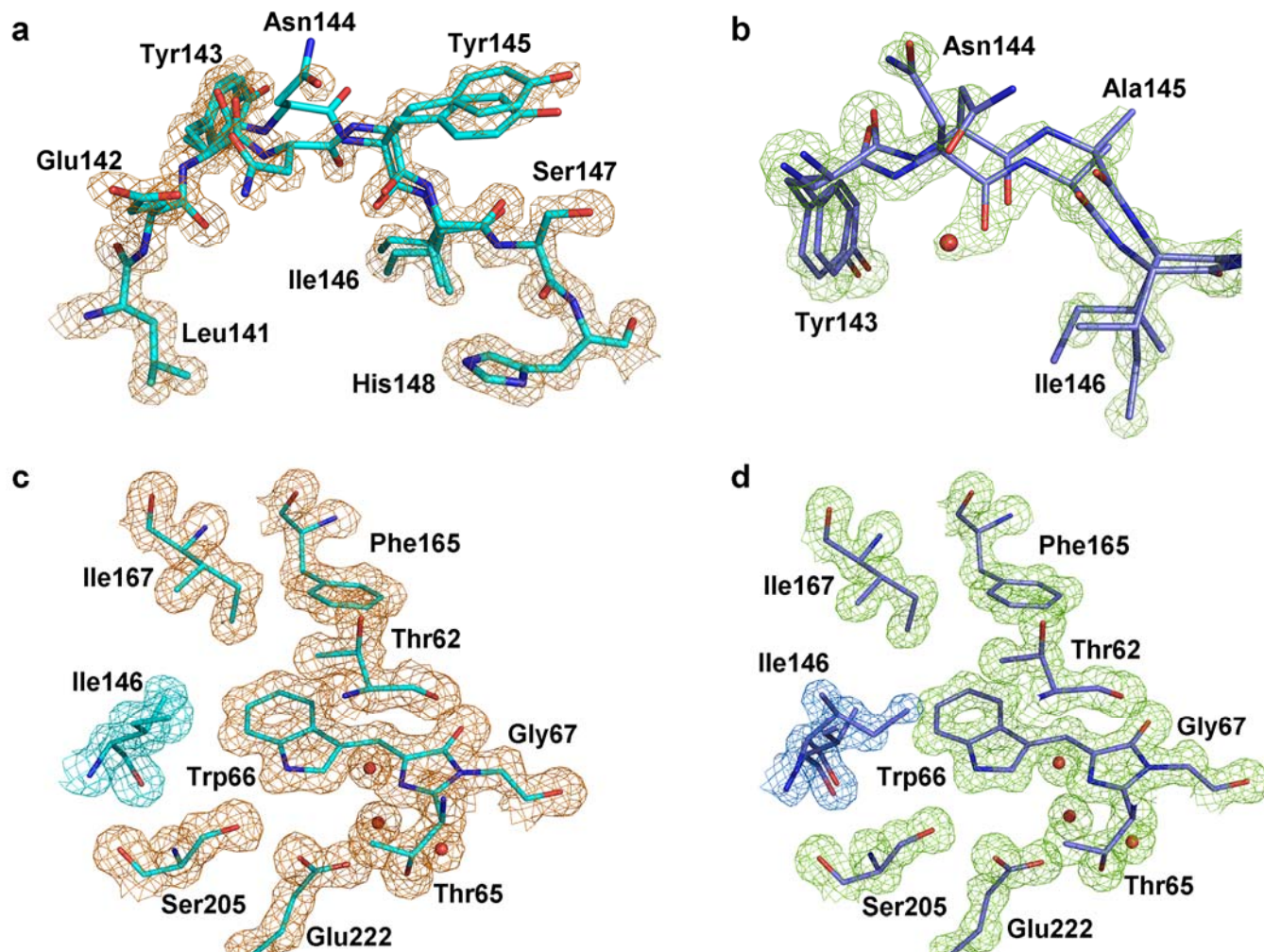


Figure S5. Final 2Fo-Fc electron density for ECFP and Cerulean. **a.** Electron density in ECFP represented at a 1.5σ level superimposed on residues 141 to 148. There is a discontinuity of the main chain density between atoms C and C_α of Asn144, suggesting alternate conformation of the residue. **b.** Electron density in Cerulean represented at a 1.5σ level superimposed on residues 141 to 146, showing the same discontinuity, and suggesting two conformers for residues 144 to 146. **c.** Electron density in ECFP represented at a 1.2σ level superimposed on the chromophore and its neighboring residues. The density of Ile146 is highlighted in cyan. **d.** Electron density in Cerulean represented at a 1.2σ level superimposed on the chromophore and its neighboring residues. The density of Ile146 is highlighted in blue. The weaker density shows that Ile146 adopts alternate conformers.

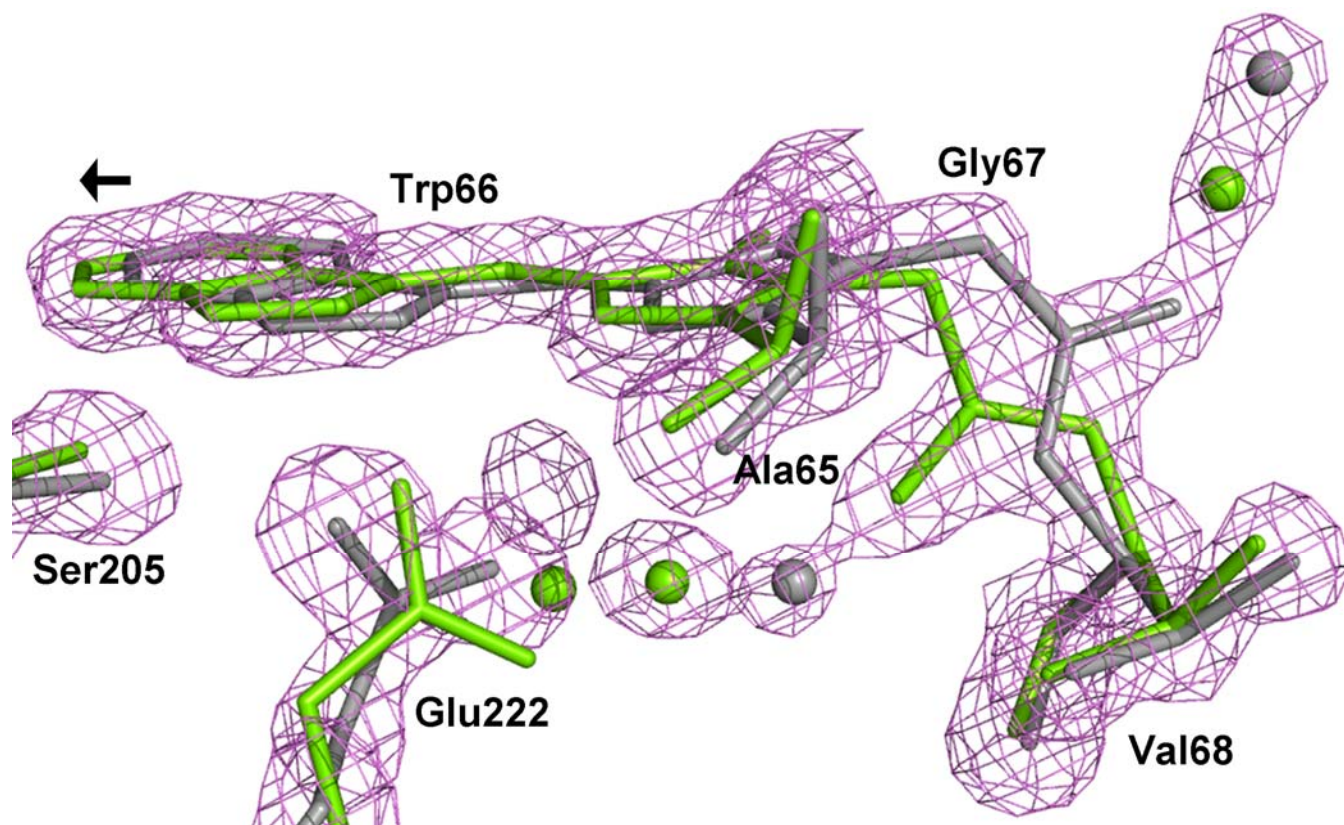


Figure S6. Electron density in Cerulean-T65A represented at a 1.5σ level on the chromophore showing the double conformations of the carbonyl group of Gly67. The black arrow figures the in-plane translation of the conjugated system moiety.

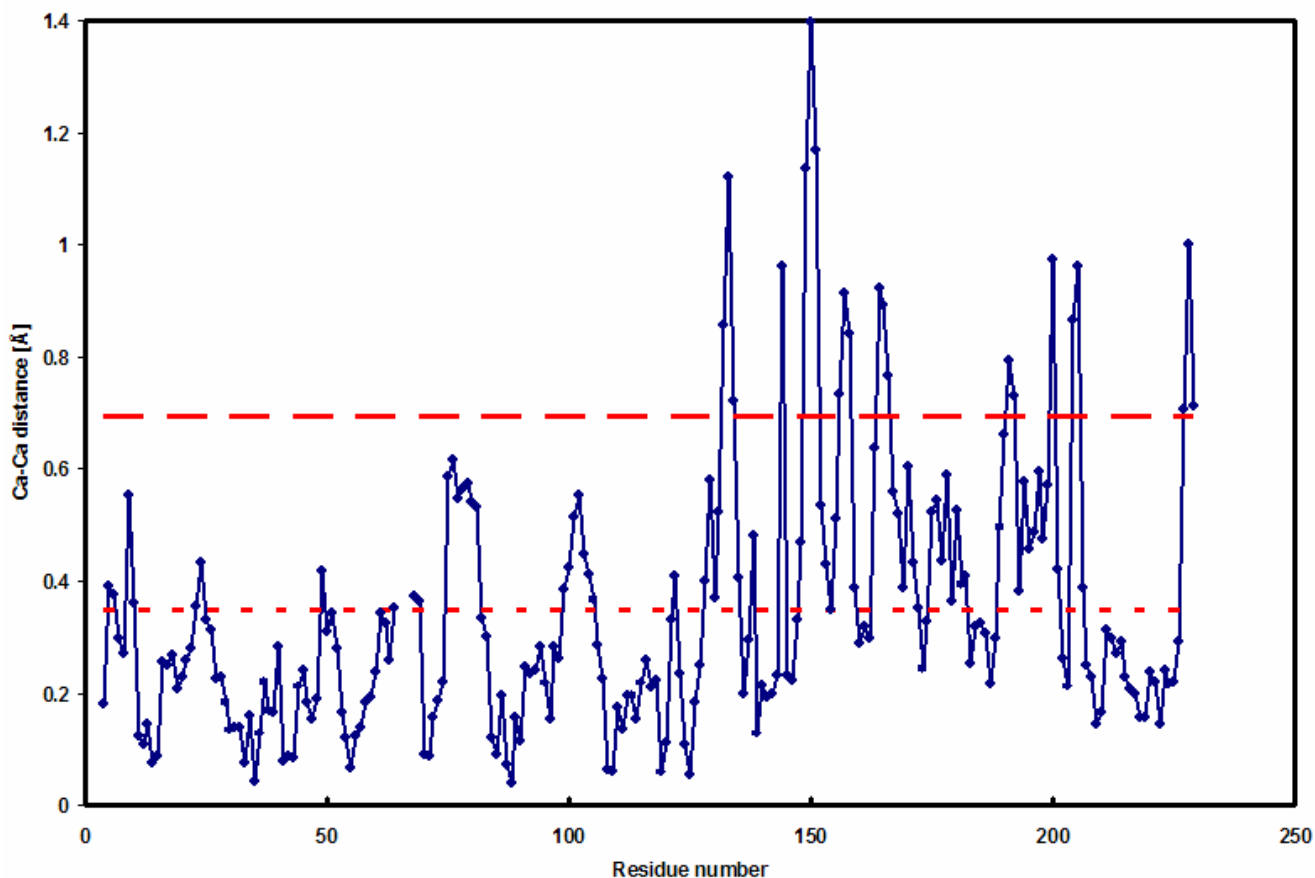


Figure S7. Plot of C_{α} positional differences between the low and neutral pH (*cerulean* conformation) structures of Cerulean. The lower red line corresponds to the average difference (0.35 Å), and the upper one to twice the average difference. Only differences above the upper line are interpreted. There are only nine locations, all in the C-terminal half of the protein, where the main chains differ significantly. Four of them are the very C-terminal part (Residues [227-229]), two regions where side chains are poorly defined (Residues [132-134] and [156-158]) and a crystal contact (Residues [191-192]). In each of these regions, there is at least a solvent-exposed carboxylate group whose partial neutralization of their negative charge at pH 5.0 is likely to be responsible for these changes. The remaining five differences are located around the chromophore (**Figure S8**).

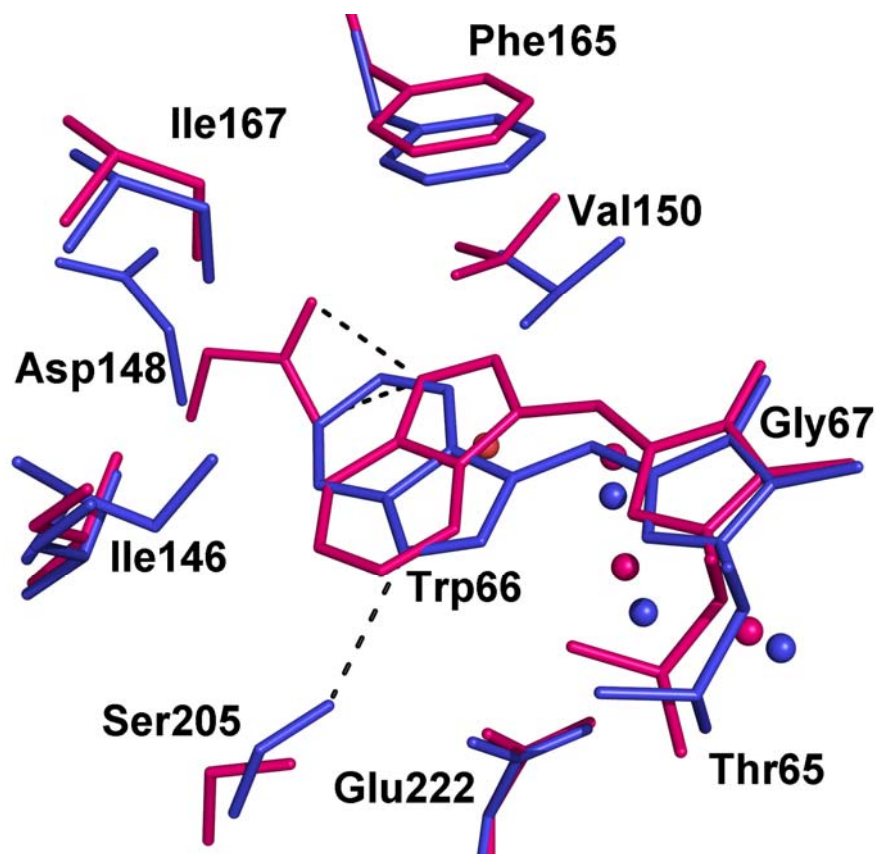


Figure S8. Superposition of the low- and neutral pH structures of Cerulean, represented in purple and blue, respectively. H-bonds to the N_{e1} atom of the chromophore are drawn in black. There are five regions in the vicinity of the chromophore that can be directly related to the *cis/trans* isomerization of the chromophore around the C_{β} - C_{γ} bond. Residues [149-151] and [164-166] are affected because the Val150/Phe165 dyad clamping the chromophore is displaced. Loss of the Hydrogen-bond between Ser205 and the chromophore induces repulsion of the Serine side chain, which in turns affects the main chain of residues [204-206]. The last difference concerns the [143-150] strand which, in the low-pH structure, adopts an intermediate position between the *ecfp* and *cerulean* conformations, most likely because Asp148 is H-bonded to the chromophore. In addition, the new conformation allows Glu222 to be connected to the bulk solvent through a chain of four water molecules, which must participate in the modulation of its pKa. Finally, displacement of Tyr200 could be an indirect effect of Val150 movement, since Tyr151 undergoes a reorientation of its side chain, which is in interaction with Tyr200 side chain.

Movie captions

Movie S1. Period 18-20 ns of the molecular dynamics simulation on the *ecfp* conformation of ECFP showing the influence of His148 on the conformation of Ile146. Rotation of the His148 side chain toward the solvent cancels the VdW interaction with Ile146, allowing the latter to adopt transiently a stronger interaction with the chromophore (*ecfp''* conformation). VdW interaction between the chromophore and His148 is maintained throughout the simulation.

Movie S2. Period 0-12 ns of the molecular dynamics simulation on the *ecfp* conformation of Cerulean showing the gradual conversion to the *cerulean* conformation. The initial position of the [143-147] strand, including the Ile146 side chain, is permanently depicted in magenta as a reference. It is apparent that once the strand is close enough to the chromophore, Ile146 adopts a conformation that allows it to interact stronger with the indole ring.

Chapitre 5

Le mécanisme de photoconversion d'EosFP

La découverte dans les coraux des protéines fluorescentes homologues à l'*avGFP* a permis d'étendre la palette des marqueurs biologiques pour l'imagerie cellulaire [114]. Il a notamment été décelé chez certaines d'entre elles des propriétés de photoactivation [118]. Il est par exemple possible de photoconvertir du vert au rouge la fluorescence de certaines de ces protéines photoactivables. C'est le cas de la protéine fluorescente EosFP et d'une dizaine d'autres protéines de la même famille. Cette propriété offre un moyen précieux de suivre les processus dynamiques dans les cellules vivantes [119].

Nous avons étudié le mécanisme de photoconversion d'EosFP par une exploration individuelle de ses surfaces d'énergie potentielle. Dans ce type de traitement adiabatique, des intersections coniques entre ces surfaces peuvent suggérer des voies réactionnelles possibles. Une approche alternative est l'utilisation de méthodes de dynamique moléculaire qui permettent au système de sauter d'une surface à une autre (*surface hopping*) lorsque ces dernières deviennent très proches [120]. Ces méthodes sont généralement applicables à des processus photochimiques très rapides tels que ceux de photo-isomérisations [121, 122]. L'article ci-dessous présente notre étude du mécanisme de photoconversion qui possède une cinétique de réaction beaucoup plus lente que ce type de phénomènes.

Photoconversion of the fluorescent protein EosFP: a hybrid potential simulation study reveals intersystem crossings

*Mickaël Lelimosin[‡], Virgile Adam[§], G. Ulrich Nienhaus[#], Dominique Bourgeois^{‡,§}, and
Martin J. Field^{*,‡}*

[‡]Institut de Biologie Structurale Jean Pierre Ebel, UMR5075 CEA-CNRS-Université Joseph Fourier, 41, rue Jules Horowitz, 38027 Grenoble Cedex 1, [§]European Synchrotron Radiation Facility, 6 Rue Jules Horowitz, BP 220, 38043 Grenoble Cedex, France, [#] Institute of Biophysics, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany, and Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 6180

AUTHOR EMAIL ADDRESS * mjfield@ibs.fr

RECEIVED DATE

CORRESPONDING AUTHOR FOOTNOTE (Word Style “FA_Corresponding_Author_Footer”). Clarify all corresponding authors’ addresses by accompanying footnotes if they are not apparent from the affiliation line. Telephone numbers, fax numbers, and e-mail addresses may all be included in the corresponding author footnotes. If an author’s address is different than the one given in the affiliation line, this information may also be included here.

ABSTRACT Fluorescent proteins undergoing green to red photoconversion have proved to be essential tools in cell biology, notably in photoactivation-based nanoscopy approaches. However, the

exact mechanism governing photoconversion, which overall involves irreversible cleavage of the protein backbone and elongation of the chromophore π -conjugation, remains uncertain. In this paper we present a theoretical investigation of the photoconversion reaction in the fluorescent protein EosFP, using excited-state hybrid quantum mechanical and molecular mechanical potentials, in conjunction with reaction-path-finding techniques. Our results reveal a mechanism in which excitation of the green form of EosFP to the first singlet excited state is followed by two intersystem crossing events, first to a triplet state and then back to the ground state singlet surface. From there, a number of rearrangements occur in the ground state and lead to the red form. Analyses of the structures and of the energetic of the intermediates along the reaction path enable us to identify the critical role of the chromophore environment in promoting photo-induced backbone cleavage. Possible ways in which photoconvertible fluorescent proteins can be engineered to facilitate photoconversion are suggested.

KEYWORDS Photoconvertible fluorescent proteins, backbone cleavage, excited-state QM/MM potentials, Nudged Elastic Band method, intersystem crossings, protein catalysis, electrostatic interactions

INTRODUCTION

During the last fifteen years, fluorescent proteins (FPs) have become indispensable tools in cell biology, as their use as markers has revolutionized understanding of cellular processes.¹ Some of these proteins can change their photophysical properties on exposure to light. Such modifications in the fluorescence of photoactivatable fluorescent proteins (PaFPs) include reversible switching between a fluorescent on state and a non-fluorescent off state, and irreversible photoconversion between a green and a red-emitting state.² The structures of the green and red forms of several photoconvertible fluorescent proteins (PcFPs) have been determined by X-ray crystallography, including EosFP,³ KikGR,⁴ Kaede,⁵ and IrisFP⁶. Because of the utility of these proteins for investigating movements in live cells² or designing super-resolution imaging schemes based on the PALM approach,⁷ considerable work has

gone into optimizing PcFPs. This led to the recent development of the monomeric variants Dendra2,⁸ mEos2,⁹ and mKikGR¹⁰ which proved more suitable for designing functional fusion constructs.

A conserved triad of residues, His62-Tyr63-Gly64 characterizes the chromophore of PcFPs. As for all members of the GFP family, the chromophore is enclosed in the center of an 11 stranded β -barrel structure (**Figure 1a**) and is held rigid by an environmental network of hydrogen bonds and van der Waals interactions (**Figure 1b**). The green form of PcFPs arises via autocatalytic maturation in the dark, yielding a 4-(*p*-hydroxybenzylidene)-5-imidazolinone π -conjugated system. Illumination with UV light causes cleavage of the amide bond between the N atom of residue 61 and the C atom of His62, formation of a double bond on the His62 sidechain, and extension of the conjugated system to a 2-[(1*E*)-2-(5-imidazolyl)-ethenyl]-4-(*p*-hydroxybenzylidene)-5-imidazolinone moiety,^{3,11} leading to the red-emitting form. Despite these changes, the three-dimensional structure of the protein is almost completely preserved in the red form (**Figure 1c**). However, a water molecule (W1) that is seen in the structures of green EosFP, IrisFP and Kaede, is not observed in the red forms. This water molecule is absent from both the green and red forms of KikGR, which is characterized by an additional isomerization of the His62 sidechain.

The basic (anionic) form of PcFP chromophores absorbs around 500 nm and the acidic (neutral) form around 390 nm. Although it has been reported that very intense illumination at 488 nm converts Dendra2,⁸ possibly in the anionic form of the chromophore, it is generally accepted that photoconversion only occurs in the neutral (phenolic) form.^{2,10,12,13} The reaction is irreversible as neither dark exposure nor illumination at any wavelength can restore the green state.¹² Measurements of rate constants for the photoconversion of PcFPs revealed much slower kinetics than for other photo-induced reactions such as photoisomerization.^{10,14} As a consequence, the quantum yield of photoconversion in PcFPs is low ($\sim 10^{-3}$ - 10^{-4}).^{4,6,10,12} Photoconversion can be abolished by substitution

of His62 with any other amino acid,^{11,13} thereby asserting this residue's critical role, and by various other mutations in the chromophore's environment.^{11,15}

To date, the exact mechanism of green to red photoconversion in PcFPs remains unknown, although several hypotheses have been proposed. Using NMR structures of Kaede, Mizuno *et al.* suggested initially that the imidazole of His62 could become biprotonated and that this would facilitate backbone cleavage via β -elimination.¹¹ Based on the X-ray crystallographic structures of EosFP in both its green and red states, Nienhaus *et al.* followed up this idea by proposing an excited state proton transfer (ESPT) from the hydroxyl group of Tyr63 to the N_e of His62, followed by a β -elimination step in which Glu212 acts as a proton acceptor.³ Finally, Hayashi *et al.* proposed a water-assisted mechanism to explain the loss of the water molecule W1 in the red form of Kaede.⁵

In fact, a detailed experimental investigation of photoconversion is difficult, due to the low quantum yield and the irreversibility of the process. In contrast, molecular simulation can provide important insights into the photoconversion mechanism. Especially appropriate tools in this endeavor are hybrid quantum chemical (QC)/molecular mechanical (MM) potentials because they are well adapted for studying phenomena that require a quantum mechanical description but which occur in large systems. Most reported QC/MM studies of photochemical reactions in proteins concern ultrafast events.¹⁶ Examples include photoisomerization of chromophores in rhodopsins,¹⁷ photoactive yellow protein¹⁸ and in other fluorescent proteins.¹⁹ In this paper, we investigate theoretically the mechanism of the green-to-red photoconversion in EosFP which is a typical representative of the PcFP family. By employing hybrid QC/MM potentials in conjunction with reaction-path finding techniques, we propose that two intersystem crossing (ISC) events to the triplet state T₁ and back to the ground state S₀ are critical for forming the red-emitting state. Structural determinants of the EosFP chromophore pocket for efficient photoconversion are also investigated, and general conclusions about PcFPs are drawn.

METHODS

All simulations were performed with the fDynamo library.²⁰ The starting structures for the simulations were those of the tetramers determined by X-ray crystallography for the green (PDB id: 1ZUX) and red (PDB id: 2BTJ) forms of EosFP.³ The positions of the hydrogen atoms were determined after the protonation states of the residues in the protein had been estimated at a pH of 7.0, using standard Poisson-Boltzmann calculations.²¹ All residues were determined to have their expected protonation, with the exception of His194, for which the pKa calculation and examination of the X-ray crystallographic structures indicated a biprotonation at pH 7.0. The resulting protein structures were geometry-optimized in vacuum using the OPLS-AA force field.²² The optimized tetrameric structures were then solvated and optimized in a cubic water box of the appropriate size, employing the TIP3P MM water model.²³ Counter-ions were added to ensure charge neutrality of the overall system. The final systems comprised approximately 43100 atoms.

QC/MM simulations were performed starting from the two prepared MM systems. Unless explicitly stated in the text, the QC region consisted of 65 atoms and contained the chromophore and the side chains of two amino-acid residues: Glu212, which has been proposed to play an essential role in photoconversion; and His194, which is involved in π -stacking with the hydroxybenzylidene moiety (**Figures 1b and 1c**). Atoms in the MM region were treated with the OPLS-AA force field and those in the QC region with either the AM1²⁴ or the PDDG-PM3²⁵ semi-empirical methods. The results obtained with the two potentials are qualitatively and, in most cases, quantitatively similar, so that we only show the PDDG-PM3 results in the main paper. The AM1 results are given as Supporting Information (SI).

To calculate the energy and atomic forces of the system in both ground and excited states, a standard configuration interaction (CI) method was incorporated into the fDynamo library. The CI method itself permits calculations with a number of possible configuration sets, either all configurations within a given active space (full CI) or a mixture of single and double excitations (CIS, CID and CISD) from a

closed-shell singlet reference state. Gradients of the energy are calculated with the Z-matrix technique.^{26,27} Our tests showed that a ten-electron/nine-orbital or [10-9]-CISD approximation gave the best compromise between accuracy and computational time. Sample results for our systems are given in **Table S1**. The accuracy obtained is similar to that of alternative, often much more computationally expensive approaches. To simplify the discussion, we refer to the [10,9]-CISD(PDDG-PM3)/OPLS and [10,9]-CISD(AM1)/OPLS hybrid potentials as the PDDG and AM1 potentials, respectively.

To investigate the photoconversion reaction, we employed the nudged-elastic-band (NEB) method²⁸ to determine minimum energy paths (MEPs) between various, selected reactant and product structures. Although MEPs do not directly give time-scale or thermodynamic information, they can provide useful information about possible mechanisms, including the identity of intermediates, the order in which events occur and estimates of barrier heights.

All NEB calculations were done using the A monomer of the prepared systems. To reduce the computational expense of the optimization, only the ~840 atoms within 10 Å of the C_α atom of His62 were allowed to move. The remaining atoms were fixed although their interactions with the mobile atoms were taken into account. Reaction paths were determined between geometry-optimized start and end structures, on either the ground or excited states, and using at least 11 structures along the pathway. Whenever suggested by intersystem crossings, results were refined by recalculating parts of the reaction path between specific intermediates at the appropriate electronic state.

RESULTS AND DISCUSSION

Previous mechanistic proposals

Scheme 1 provides a summary of the steps putatively involved in the previously proposed photoconversion mechanism of EosFP -an ESPT from the hydroxybenzylidene group of the

chromophore to the N_ϵ of His62, followed by a β -elimination reaction in which Glu212 acts as a proton acceptor.³ The (arbitrary) notation that we use for the various intermediate states throughout this work is also provided in **Scheme 1**.

Our initial simulations were performed to test the mechanism of **Scheme 1** by calculating NEB pathways on the excited state singlet (S1) for the first step and on the ground state singlet (S0) surface for all following steps. The energy profiles obtained for these latter are shown in **Figure 2 and S1** for the PDDG and AM1 hybrid potentials, respectively. Overall we consider this mechanism unlikely for the reasons given below.

Construction or minimization of NEB pathways between the G and A1 states failed or gave unreasonably high energy barriers (not shown). The occurrence of an ESPT from Tyr63 to His62 in this step was anticipated from the known strong decrease of the hydroxybenzylidene pKa in S1.²⁹ In avGFP, the proton is transferred to Glu222 (corresponding to Glu212 in EosFP) via a hydrogen-bonding network.³⁰ Such a network between Tyr63 and His62 does not appear to exist in EosFP, based on the crystallographic structures in ground state S0, and it is difficult to identify a realistic path that could transiently form in S1, given the ~ 11 Å distance that separates the hydroxyl group of Tyr63 and the N_ϵ atom of His62. Moreover, the N_ϵ atom of His62 resides in a non-polar environment unfavorable for protonation.

The next step in the mechanism of **Scheme 1** is the cleavage of the C_α -N bond from the imidazolium-based A1 state. This step appears unlikely to occur as it requires crossing an energy barrier approaching 150 kJ/mol. It is followed by an intra-residue proton transfer within Phe61 (A2 to A3) and by the second step of the β -elimination reaction, which involves proton transfer from C_β of His62 to the carboxylate group of Glu212 (A3 to A4). Together these steps have small energy barriers and result in a large decrease in energy. By contrast, the last step (A4 to RA) again requires crossing a substantial

energy barrier of ~ 75 kJ/mol. This step involves proton transfer from N_ϵ to N_δ of the imidazole via an intermediate characterized by a sp^3 -hybridized C_ϵ . This rearrangement is necessary to restore the hydrogen bond between the carbonyl moiety of Phe61 and the N_δ atom of His62.

A new proposal

The previously discussed mechanism starts with an ESPT from the hydroxybenzylidene moiety of the chromophore. Thus, the backbone cleavage occurs from an intermediate with a deprotonated chromophore (A1 in **Scheme 1**). This hypothesis appears inconsistent with the fact that photoconversion, which is mainly characterized by the backbone cleavage, is not possible from the anionic form of the chromophore. To counter this, Mizuno *et al.* suggested that the hydrogen-bond network around the anionic form of the chromophore could be different from that in the intermediate deprotonated state.¹¹ Their proposition was based on the well-known three-state model depicting avGFP photophysics.²⁹ However, it is not evident how such changes could occur during the reaction as the environment of the chromophore, and thus its hydrogen-bond network, remains highly conserved after photoconversion (**Figure 1b and 1c**).

As an alternative we propose that the chromophore remains protonated during the entire photoconversion process and that backbone cleavage is a direct consequence of excitation of the neutral chromophore. To investigate this hypothesis we tested the mechanism shown in **Scheme 2** in which Glu212 acts as the proton acceptor in the β -elimination. We started off by optimizing NEB pathways between the G and B4 species on the first singlet excited state (S1) surface. The resulting energy profiles are shown in **Figures 3a and S2a**.

The first event along the optimized NEB path is an ESPT from the N_δ atom of His62 to the carbonyl moiety of Phe61. The energy of this proton transfer was determined to be ~ 25 kJ/mol with the PDDG potential. A conical intersection between the S1 and first excited state triplet (T1) surfaces appears at

the B1 intermediate. On the S1 excited state surface, the activation energies for breaking the C α -N bond is ~110 kJ/mol. This barrier is too high to be crossed, considering the short lifetime of the S1 state, which is estimated to be of the order of a nanosecond. Consequently, to reach B4 from G, there must be an intersystem crossing from the S1 to T1 surfaces, which we denote ISC1.

Making this assumption, we re-optimized NEB pathways for the transition between the B1 and B4 structures of **Scheme 2** on the T1 surface (**Figure 3b and S2b**). On going from the B1 to B2 intermediates, ~15 kJ/mol are required to activate the C α -N bond cleavage. At B2, the T1 and S0 surfaces become degenerate and remain so until formation of the B3 intermediate, after which ground S0 state become lower in energy. This suggests a second intersystem crossing, ISC2, somewhere between the B2 and B3 intermediates. Thus, proton transfer from C β of His62 to the Glu212 carboxylate, leading to B4, occurs on the S0 surface.

To check whether photoconversion could be initiated in the T1 state (rather than in the S1 state), we optimized NEB pathways on this surface between the G and B4 structures of **Scheme 2**. The profiles are shown in **Figures 4a and S3a**. In contrast to what was observed for the S1-optimized pathways (**Figures 3a and S2a**), cleavage of the C α -N bond is concerted with proton transfer from the N δ atom of His62. Thus, no separate B1 intermediate is observed. However, an activation energy of ~135 kJ/mol is required to reach B2, suggesting that photoconversion is much more likely to occur *via* the ISC1-path.

Contrary to red FP's such as DsRed, which mature to their red forms in the dark, light is necessary to convert PcFPs. To check whether our calculations are in agreement with this fact, we optimized NEB pathways on the S0 surface between the G and B4 structures. The profiles shown on **Figures 4b and S3b**, although they display characteristics similar to those obtained in the T1 state (**Figures 4a and S3a**), involves unrealistically high activation energies for the cleavage step, which reach ~300 kJ/mol. Thus, the results of our calculations are consistent with the inability of PcFPs to convert in the dark.

Activation with violet light (~390nm) of chromophores in PcFPs in their anionic form does not induce photoconversion. To check that our calculations are also able to predict this experimental observation, we calculated NEB pathways for cleavage on the S1 surface but using the anionic form of the chromophore. The energy profiles are shown in **Figures 4c and S3c** and are qualitatively similar to those obtained for the S0 and T1 states with a neutral chromophore. The C $_{\alpha}$ -N bond cleavage is concerted with proton transfer from the N $_{\delta}$ atom of His62, and the activation energy is very high, at ~170 kJ/mol. This is in agreement with the impossibility to photoconvert anionic forms of PcFPs.

Role of Glu212 in the photoconversion

Glu212 is known to be essential for photoconversion of EosFP, as mutation of this residue to a glutamine (E212Q) abolishes the reaction.^{3,15} Nienhaus *et al.* therefore suggested that the glutamate may act as a proton acceptor,³ whereas Mizuno *et al.* proposed that it participates in the stabilization of the transition state.⁵

In our **Scheme 2**, we found that proton transfer to Glu212 occurs between intermediates B3 and B4. To investigate the back proton transfer from Glu212 to the N $_{\delta}$ atom of His62, we optimized NEB pathways on the S0 surface starting from B4 and going to the final red structure R. The optimized profiles are shown in **Figures 3c and S2c**. Overall the process is complex and involves a return of the proton to the C $_{\beta}$ atom of His62 in a B3-like state, in which the S0 and T1 states are degenerate, and with an activation barrier of ~55 kJ/mol. This suggests that B4 may not be involved in the actual photoconversion pathway.

To check if another mechanism is possible, without protonation of Glu212, we optimized NEB pathways from the second intersystem crossing ISC2 to the final red structure on the ground state

surface S0. The full mechanistic pathway is shown in **Scheme 3** and the optimized NEB profiles in **Figures 5 and S4**. Simulations of the mechanism can be visualized with **Movies S1 and S2**.

First of all, the B2 intermediate has a carbocation on the C_α atom of His62. This makes the C_β atom very acidic and delocalized electrons on the imidazolide make the C_γ atom a strong base, much more so than the carboxylate of Glu212. Therefore, in the step between B2 and C3, the proton on C_β is spontaneously transferred to the C_γ of the imidazolide moiety and a double bond is formed between C_α and C_β . The same proton is then transferred to the N_δ of His62 and the imidazole ring is reformed. In the last step, the proton on the Phe61 carbonyl rearranges to form the carboxamide. Overall, the complete process is energetically downhill and the highest energy barrier for any of the proton transfers is ~ 25 kJ/mol.

KikGR specificities in the PcFP family

In the PcFP family, KikGR differs from the other members, in that the His62 sidechain isomerizes along the photoconversion pathway. The mechanism depicted in **Scheme 3** accounts for this specificity. We propose that isomerization of the His62 side chain occurs in the B2 intermediate state of **Scheme 3**, by rotation about the C_α - C_β single bond, taking advantage of the long lifetime of the triplet state before ISC2 takes place. After rotation, formation of the double bond may follow and the final proton transfers lead to the red (isomerized) state. In EosFP and Kaede, residue Ile196, in the vicinity of His62, and maintained by hydrophobic interactions with Leu210, does not leave enough space for isomerization of the His62 sidechain (**Figure 6a**). The situation is different in KikGR, where Ile196 and Met40 of EosFP are swapped to Met198 and Val40 (**Figure 6b**). As a result, Leu212 maintains a hydrophobic interaction with Val40 and leaves Met198 free of interaction. This provides conformational freedom to Met198, which allows His62 isomerization to occur (**Figure 6b**).

Another specificity of KikGR in the PcFP family is the absence of water molecule W1 in the green form (**Figure 6b**). For Kaede and EosFP, Hayashi *et al.* proposed that W1, which disappears in the red form

of these proteins (**Figure 6a**), may assist in photoconversion.⁵ In our study, we were unable to find reasons for such involvement of W1. The absence of the water molecule in the crystal structure of KikGR also suggests that it is not essential for the reaction. In our models of EosFP green, W1 is hydrogen-bonded to the side chain oxygen atom of Gln38, to the carbonyl group of Ser39 and, to a lesser extent, to the C_ε atom of His62. The latter two interactions are disrupted during our simulations of the photoconversion and formation of the imidazolidine moiety on His62 induces their replacement by hydrogen-bonds to the C_{δ1} atom of Ile10, and, to a lesser extent, to the C_ε atom of Met40 (**Movies S3 and S4**). In our model of red EosFP, because our methods require keeping the same number of atoms throughout the simulations, the hydrogen-bonds observed in the green form are restored. We suggest that, in the real mechanism, rearrangements that occur during photoconversion induce transient weakening of the interactions maintained by W1, which eventually leads to the exit of this water molecule from the protein.

Electronic structure of intermediates

Figure 7 is a schematic that illustrates the electronic structures of the various intermediates along the photoconversion pathway of **Scheme 3** obtained with the PDDG potential. For each intermediate, cartoons of the most important electronic configurations in each electronic state are given, together with images of the orbitals that are involved.

Overall the majority of the intermediates' electronic states are dominated by a single electronic configuration of the appropriate type. The G electronic states are of simple type except that excitation involves the HOMO-1 rather than the HOMO. As these orbitals are localized on the hydroxybenzylidene-imidazole and the His62 imidazole moieties, respectively, the S1 and T1 states retain charge density on His62. By contrast, for the B1 intermediate, the excitation to S1 and T1 comes from the HOMO even though this orbital and the LUMO have very similar character to those of G. Thus, there is charge transfer between the His62 imidazolidine moiety (where the HOMO is localized)

and the hydroxybenzylidene-imidazole group (where the LUMO is localized), which probably explains why the S1 and T1 states are degenerate, as the two unpaired electrons in the HOMO and the LUMO are in different regions of space and weakly interact.

At the B2 intermediate, T1 retains its simple form although the orbitals involved, the HOMO and LUMO, have density concentrated mainly between the imidazolidine and imidazolinone moieties, but little on the phenol group. S0 and S1 are more complicated. S1 has no compact description whereas S0 has significant contributions from the two closed-shell configurations that are possible by doubly-occupying the HOMO and LUMO. These are the same orbitals that are singly-occupied in T1 and may indicate why the S0 and T1 states are degenerate for this intermediate.

After B2, the mechanism continues on the S0 surface. The remaining electronic states are all of simple type, except for the S1 and T1 states of the red form R, which involve a mixture of two configurations obtained by excitation from the HOMO-1 and the HOMO, respectively.

Anionic form of PcFPs

To better understand the role of the chromophore protonation state in photoconversion, we calculated energy profiles for the anionic chromophore using suitably modified structures from the NEB pathways optimized with the neutral chromophore. The profiles are shown in **Figures 8a and S5**.

Clearly, the presence of the extra negative charge on the chromophore is likely to stabilize the starting S1 state of G due to electrostatic interactions between Tyr63 and the cationic His194, while perhaps destabilizing the S1 state of B1 because formation of the negatively-charged imidazolidine is less favorable. The trends observed in the profile suggest that the NEB pathways optimized for cleavage does not go through a B1-like structure with an anionic chromophore, because the activation barrier, approaching 200 kJ/mol, is larger than those found for a concerted reaction (**Figures 4c and S3c**).

Cartoons of the electronic structures of the anionic G and B1 intermediates are shown in **Figure 8b**. The G states are similar to those of **Figure 7** although the charge density localized on the imidazole of His62 in the HOMO is no longer present. For B1, the S0 state resembles that found before whereas the S1 and T1 states have changed markedly. The T1 state of B1, like that of G, now involves excitation between the HOMO-1 and LUMO orbitals, both localized on the hydroxybenzylidene-imidazolinone moiety, whereas the S1 state involves excitation between the HOMO and LUMO+1 orbitals that are localized on the imidazolidine and the amide moieties, respectively. This is consistent with the observation that the S1 and T1 states are no longer degenerate for this intermediate. Overall, these results show probable obstacles for a photoconversion pathway involving an anionic chromophore in PcFPs.

Catalytic effects inside PcFPs

The influence of the protein on the reaction mechanism was investigated by recalculating the energies of the structures along the optimized pathways *in vacuo*, i.e. by excluding all MM atoms and the sidechains of His194 and Glu212. The results are shown in **Figures 9a and S6a** and indicate a clear stabilization of the S1 and T1 (degenerated) states of the B1 intermediate by the protein. The energy difference between G and B1 species at the S1 state, which can be approximate as the most critical activation barrier to be crossed due to the short lifetimes at this excited state, is shifted from ~25 kJ/mol in the protein to ~95kJ/mol in the vacuum. To more fully understand this catalytic effect, the energies of the structures along the optimized NEB pathways were recalculated with a new QC/MM partitioning in which Glu212 and His194 were excluded from the QC region. The profiles, which are shown in **Figure S7**, are indistinguishable when superimposed on the optimized ones, suggesting that the role of these residues can equally well be described with an MM potential.

To probe the specific effects of various residues on the photoconversion mechanism, the profile calculations were repeated by systematically setting the MM charges on each residue within the mobile

region to zero. This change has minimal effect on the profiles except in the case of Glu212 and the three positively charged residues, His194, Arg66 and Arg91, which are found closest to the hydroxybenzylidene-imidazolinone moiety (**Figure 1b**).

The profiles when Glu212 has no MM charge are displayed in **Figures 9b and S6b**. They indicate that removal of the residue's negative charge actually stabilizes the S1 and T1 excited states of B1 and has little effect on the remainder of the pathway. This result is at variance with the suggestion of Hayashi *et al.* for rationalizing the key role of this residue.⁵ The profiles for His194 are shown in **Figures 10a and S8**. In both there is a stabilization of the green form (structure G) and a destabilization of the B1 intermediate on the S1 surface. The energy difference between the G and B1 structures increases from ~25 to ~110 kJ/mol, equivalent to a factor of ~4.5. When compared to His194, the two arginines, Arg66 and Arg91, show similar, but less pronounced, catalytic effects on the photoconversion energy profiles (**Figure S9**). Clearly, these three positively charged residues, and particularly His194, might have a crucial stabilizing effect on the transition state leading to backbone cleavage.

Figure 10b shows cartoons of the electronic states of the G and B1 intermediates in the case where the charge of His194 has been nullified. The orbitals have very similar forms to their equivalents in **Figure 7**, except that, they are destabilized and the order of the HOMO and HOMO-1 is reversed. As a consequence, the S1 state now has contributions from two configurations, which seems to lead to a decrease in the energy. In contrast, in the B1 state, the LUMO, which is localized on the hydroxybenzylidene-imidazole group, is destabilized (of 1 eV) by the absence of the positive charge on His194, thereby leading to higher energies for S1 and T1.

The results discussed above can also account for the fact that the E212Q variant of EosFP does not undergo photoconversion, even though, according to our findings, Glu212 itself does not participate to the reaction. Glu212 hydrogen bonds directly to His194 and indirectly, *via* a water molecule (W3 in

Figure 1b), to Arg66. Hence, it is possible that in the E212Q mutant, His194 can no longer be bi-protonated or that, due to the disruption of the hydrogen-bond network around the chromophore, the critical Arg66 residue is displaced, thus impeding photoconversion. This possibility is substantiated by the observation that the E212Q mutation increases the pKa of Tyr63 by 0.8 unit (6.6 vs 5.8 in wtEosFP), suggesting substantial modifications in the hydroxybenzylidene-imidazole environment.¹⁵ The crucial role of Arg66 in shifting the Tyr63 pKa found in Dendra2³¹ also agrees with this hypothesis. Similar indirect effects could account for the loss of activity reported in other PcFP mutants, such as the A69S variant of Kaede.¹¹

Getting improved variants of PcFPs

Based on our findings, it could be speculated that the introduction of additional positively charged residues in the chromophore environment, or of mutations that would result in an improved network of positive charges, could facilitate photoconversion. However, whereas a lowering of the hydroxybenzylidene pKa, as observed in Dendra2 effectively achieves this goal by increasing the fraction of protonated chromophores at physiological pH,³¹ it is expected that the yield of photoconversion cannot be improved to an appreciable extent. In the mechanism that we propose in **Scheme 3**, photoconversion is inherently limited by the two intersystem crossings that lead to transitions between potential energy surfaces.

Overall, we assign the slow kinetics observed experimentally in PcFPs to the need for two low-yield intersystem crossings events. This also accounts for the somewhat surprising finding that such a low yield process is accompanied by only very small structural changes between the green and red states in PcFPs.

CONCLUSION

We have investigated the mechanism of photoconversion between the green and red forms of the fluorescent protein EosFP using hybrid QC/MM potentials in conjunction with reaction-path-finding techniques. The preferred pathway determined from our simulations involves excitation of the green protonated form to the S1 state, followed by two intersystem crossings (ISCs), first to the T1 state, and then to the S0 ground state. Proton transfer from His62 to the carbonyl group of Phe61 occurs on the S1 surface, forming an imidazolide ring on His62 sidechain. Cleavage of the amide bond between Phe61 and His62 occurs on the T1 surface. Both steps involve small energy barriers. Completion of the red form occurs on the S0 surface *via* a series of proton transfers that are downhill in energy, extend the π -conjugated system, and lead to reformation of the imidazole ring on His62 sidechain.

The occurrence of ISCs is a novel aspect of the photoconversion mechanism in Pcfp's and provides an explanation for the observed low QY of the process. The mechanism proposed in this work differs from previous hypothesis in a number of ways: (i) The phenol ring of the chromophore does not undergo an ESPT and remains protonated along the whole reaction pathway; (ii) The essential residue Glu212 remains deprotonated throughout and takes no part in proton transfer; (iii) Waters in the vicinity of the chromophore are not implicated. An analysis of the structures and energetics of the intermediates along the pathway shows that three cationic residues closest to the chromophore have an important catalytic role, whereas Glu212, which is itself anticatalytic, indirectly serves to maintain the integrity of the chromophore's electrostatic environment. These results suggest that engineering EosFP by reorganizing or introducing extra cationic residues around the chromophore could facilitate photoconversion to a limited extent.

ACKNOWLEDGMENT

We thank Antoine Royant for fruitful discussions. We also thank the Ministère de l'Enseignement Supérieur et de la Recherche, the Centre National de la Recherche Scientifique, and the Agence Nationale de la Recherche (ANR-07-BLAN-0107-01) for financial support.

SUPPORTING INFORMATION PARAGRAPH

Table of absorption and emission wavelength predicted by the QC/MM potentials used in this work, supplementary figures of energy profiles along the optimized NEB pathways, and movies of the proposed photoconversion mechanism.

ABBREVIATIONS

QC/MM: Quantum Chemistry/Molecular Mechanics; ESPT: Excited State Proton Transfer; QY: Quantum Yield; NEB: Nudged Elastic Band; MEP: Minimum Energy path; NMR: Nuclear Magnetic Resonance; CISD: Configuration Interaction Singly and Doubly (excited); ISC: InterSystem Crossing; HOMO: Highest Occupied Molecular Orbital; LUMO: Lowest Unoccupied Molecular Orbital; PaFP: Photoactivable Fluorescent Protein; PcFP: Photoconvertible Fluorescent Protein; PALM: PhotoActivated Localization Microscopy.

FIGURE CAPTIONS

Figure 1. (a) Tetrameric structure of wt-EosFP. Chromophores are shown as van der Waals spheres; (b) superposition of the chromophore and surrounding residues in green forms of EosFP, IrisFP, KikGR

and Kaede (hydrogen bonds as dotted lines); (c) superposition of the chromophore and surrounding residues in the red forms of EosFP, IrisFP, KikGR and Kaede.

Figure 2. Energy profile of optimized NEB paths from A1 to A4 and from A4 to RA at the ground state S0 calculated with a PDDG/OPLS hybrid potential.

Figure 3. Energy profiles of optimized NEB paths calculated with a [10,9]-CISD(PDDG)/OPLS hybrid potential: (a) from G to B4 at the excited state S1 (blue plus signs); (b) from B1 to B4 at the excited state T1 (green plus signs); and (c) from B2 to B4 and from B4 to R at the ground state S0 (red plus signs). Solid lines in each panel give the energies of the optimized NEB structures in the electronic states that were not used in the optimization.

Figure 4. Energy profile of optimized NEB paths calculated with a [10,9]-CISD(PDDG)/OPLS hybrid potential from G to B4: (a) at the excited state T1 (green plus signs); (b) at the ground state S0 (red plus signs); and (c) with a permanently anionic chromophore at the excited state S1 (blue plus signs). The solid lines give the energies of the optimized NEB structures but in the electronic states different from the one in which the optimization was performed.

Figure 5. Energy profile of optimized NEB paths calculated with a [10,9]-CISD(PDDG)/OPLS hybrid potential from G to B1 at the excited state S1 (blue plus signs), from B1 to B2 at the excited state T1 (green plus signs) and from B2 to R at the ground state S0 (red plus signs). The solid lines give the energies of the optimized NEB structures in electronic states that were not used during optimization.

Figure 6. Superposition of structures of EosFP with (a) Kaede and (b) KikGR in the chromophore region before (left) and after (right) photoconversion (hydrophobic interactions as dashed lines).

Figure 7. Schematic of the electronic states along the photoconversion path calculated with a [10,9]-CISD(PDDG)/OPLS hybrid potential. For convenience, triplets are represented by a single configuration (the $S=1$; $S_z = +1$ microstate), as are singlets with two unpaired electrons (electrons with opposing spins in two different orbitals). The weights of the states (indicated as %), however, include

contributions from all equivalent configurations. For each structure, only molecular orbitals involved in the excitation are shown.

Figure 8. (a) Energy profile of photoconversion calculated with a permanently anionic chromophore and a [10,9]-CISD(PDDG)/OPLS hybrid potential. Energies of structures extracted from the real path (permanent neutral chromophore) are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states with a deprotonated chromophore. Crosses show the energies of the optimized path structures with the anionic chromophore, and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states. (b) Schematic of the electronic states of the G and B1 anionic structures.

Figure 9. Energy profile of photoconversion calculated: (a) for the chromophore *in vacuo* with a [10,9]-CISD(PDDG) potential; (b) *in protein* with null atomic charges (MM) on Glu212 and with a [10,9]-CISD(PDDG)/OPLS hybrid potential. A new QC/MM partition is defined for the hybrid potential in which His194 and Glu212 are located in the MM region. Energies of structures extracted from the optimized NEB path are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states. Crosses show the energies of the optimized path structures with the modified potentials, and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states.

Figure 10. (a) Energy profile of photoconversion calculated with null atomic charges (MM) on His194 and a [10,9]-CISD(PDDG)/OPLS hybrid potential. The QC/MM partition is used in which His194 and Glu212 are located in the MM region. Energies of structures extracted from the real path (permanent neutral chromophore with initial QC/MM partitioning) are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states. Crosses show the energies of the optimized path structures with the modified hybrid potential and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states. (b) Schematic of the electronic states of the G and B1 structures with the modified hybrid potential.

SCHEME TITLES

Scheme 1. Hypothetic photoconversion mechanism inspired from Nienhaus *et al.*,³ with the intermediate structures derived from simulation. Arrows show the main reaction steps in red (color assigned to the ground state S0 throughout the paper).

Scheme 2. Suggested photoconversion mechanism with Glu212 as proton acceptor (B4) and intermediate structures derived from simulation. Arrows show the main reaction steps in the color of the electronic state involved: blue for S1, green for T1, red for S0.

Scheme 3. Photoconversion mechanism with the intermediate states (C3, C4) derived from simulation. Arrows show the main reaction steps in the color of the electronic state involved: blue for S1, green for T1, red for S0.

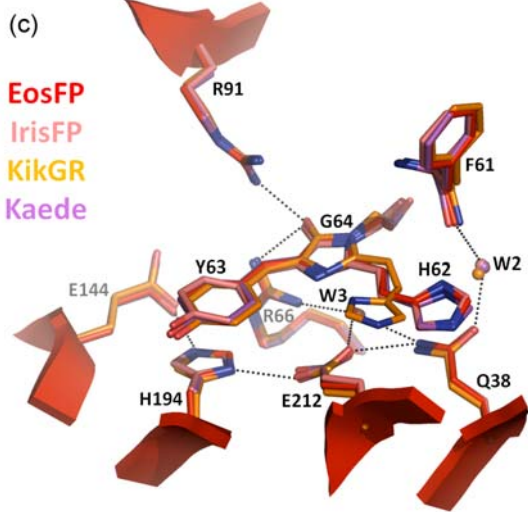
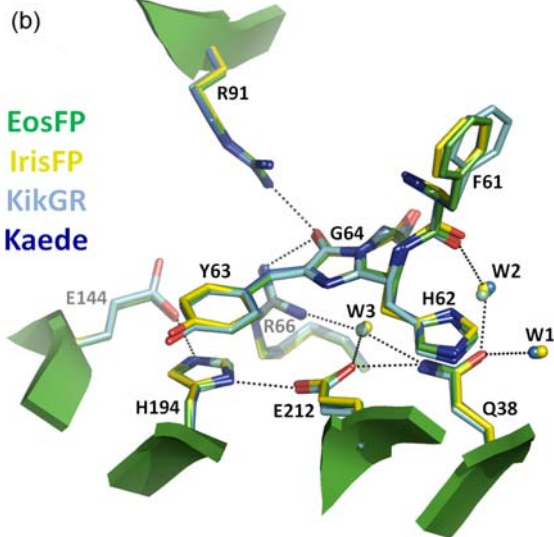
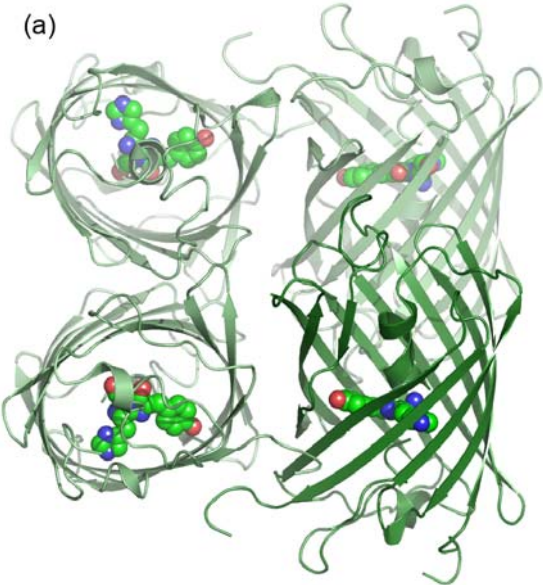
REFERENCES

SYNOPSIS TOC

- (1) Shaner, N. C.; Patterson, G. H.; Davidson, M. W. *J Cell Sci* **2007**, *120*, 4247-60.
- (2) Wiedenmann, J.; Nienhaus, G. U. *Expert Rev Proteomics* **2006**, *3*, 361-74.
- (3) Nienhaus, K.; Nienhaus, G. U.; Wiedenmann, J.; Nar, H. *Proc Natl Acad Sci U S A* **2005**, *102*, 9156-9.
- (4) Tsutsui, H.; Karasawa, S.; Shimizu, H.; Nukina, N.; Miyawaki, A. *EMBO Rep* **2005**, *6*, 233-8.
- (5) Hayashi, I.; Mizuno, H.; Tong, K. I.; Furuta, T.; Tanaka, F.; Yoshimura, M.; Miyawaki, A.; Ikura, M. *J Mol Biol* **2007**, *372*, 918-26.
- (6) Adam, V.; Lelimosin, M.; Boehme, S.; Desfonds, G.; Nienhaus, K.; Field, M. J.; Wiedenmann, J.; McSweeney, S.; Nienhaus, G. U.; Bourgeois, D. *Proc Natl Acad Sci U S A* **2008**, *105*, 18343-8.
- (7) Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F. *Science* **2006**, *313*, 1642-5.
- (8) Gurskaya, N. G.; Verkhusha, V. V.; Shcheglov, A. S.; Staroverov, D. B.; Chepurnykh, T. V.; Fradkov, A. F.; Lukyanov, S.; Lukyanov, K. A. *Nat Biotechnol* **2006**, *24*, 461-5.
- (9) McKinney, S. A.; Murphy, C. S.; Hazelwood, K. L.; Davidson, M. W.; Looger, L. L. *Nat Methods* **2009**, *6*, 131-3.

- (10) Habuchi, S.; Tsutsui, H.; Kochaniak, A. B.; Miyawaki, A.; van Oijen, A. M. *PLoS ONE* **2008**, *3*, e3944.
- (11) Mizuno, H.; Mal, T. K.; Tong, K. I.; Ando, R.; Furuta, T.; Ikura, M.; Miyawaki, A. *Mol Cell* **2003**, *12*, 1051-8.
- (12) Ando, R.; Hama, H.; Yamamoto-Hino, M.; Mizuno, H.; Miyawaki, A. *Proc Natl Acad Sci U S A* **2002**, *99*, 12651-6.
- (13) Wiedenmann, J.; Ivanchenko, S.; Oswald, F.; Schmitt, F.; Rocker, C.; Salih, A.; Spindler, K. D.; Nienhaus, G. U. *Proc Natl Acad Sci U S A* **2004**, *101*, 15905-10.
- (14) Dittrich, P. S.; Schafer, S. P.; Schuille, P. *Biophys J* **2005**, *89*, 3446-55.
- (15) Nienhaus, G. U.; Nienhaus, K.; Holzle, A.; Ivanchenko, S.; Renzi, F.; Oswald, F.; Wolff, M.; Schmitt, F.; Rocker, C.; Vallone, B.; Weidemann, W.; Heilker, R.; Nar, H.; Wiedenmann, J. *Photochem Photobiol* **2006**, *82*, 351-8.
- (16) Virshup, A. M.; Punwong, C.; Pogorelov, T. V.; Lindquist, B. A.; Ko, C.; Martinez, T. J. *J Phys Chem B* **2009**, *113*, 3280-91.
- (17) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys J* **2009**, *96*, 403-16.
- (18) Groenhof, G.; Schafer, L. V.; Boggio-Pasqua, M.; Grubmuller, H.; Robb, M. A. *J Am Chem Soc* **2008**, *130*, 3250-1.
- (19) Schafer, L. V.; Groenhof, G.; Boggio-Pasqua, M.; Robb, M. A.; Grubmuller, H. *PLoS Comput Biol* **2008**, *4*, e1000034.
- (20) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *Journal of Computational Chemistry* **2000**, *21*, 1088-1100.
- (21) Antosiewicz, J.; Mccammon, J. A.; Gilson, M. K. *Journal of Molecular Biology* **1994**, *238*, 415-436.
- (22) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225-11236.
- (23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926-935.
- (24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *Journal of the American Chemical Society* **1985**, *107*, 3902-3909.
- (25) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *Journal of Computational Chemistry* **2002**, *23*, 1601-1622.
- (26) Handy, N. C.; Schaefer, H. F. *Journal of Chemical Physics* **1984**, *81*, 5031-5033.
- (27) Patchkovskii, S.; Thiel, W. *Theoretical Chemistry Accounts* **1997**, *98*, 1-4.
- (28) Galvan, I. F.; Field, M. J. *Journal of Computational Chemistry* **2008**, *29*, 139-143.
- (29) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc Natl Acad Sci U S A* **1996**, *93*, 8362-7.
- (30) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. *Proc Natl Acad Sci U S A* **1997**, *94*, 2306-11.
- (31) Adam, V.; Nienhaus, K.; Bourgeois, D.; Nienhaus, G. U. *Biochemistry* **2009**, *48*, 4905-15.

Figure 1.



Scheme 1.

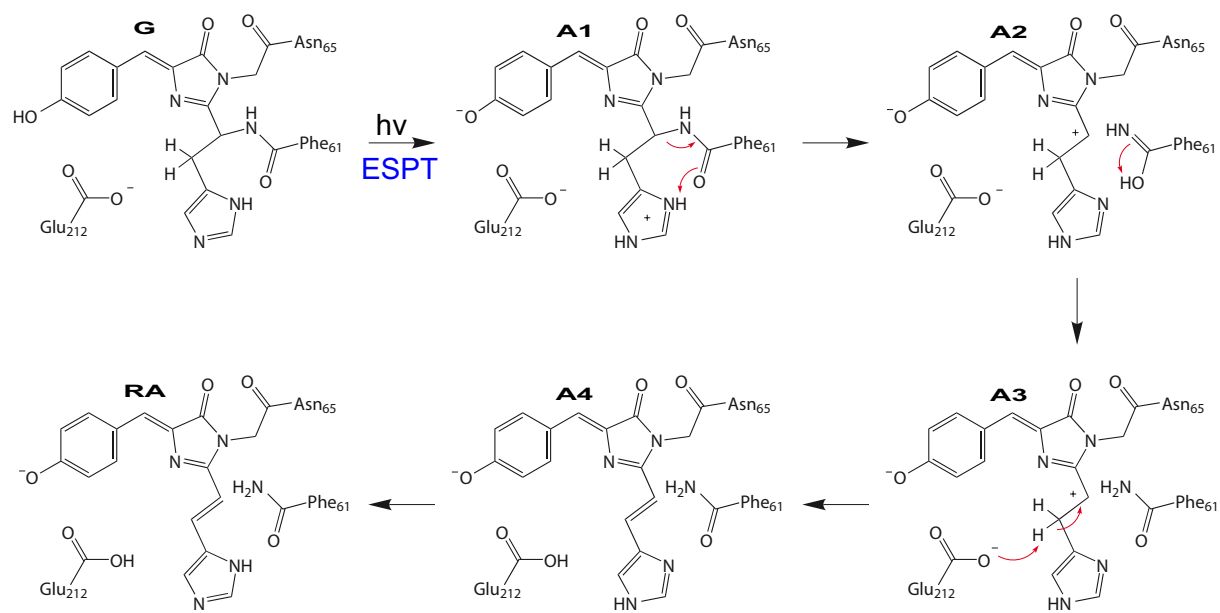
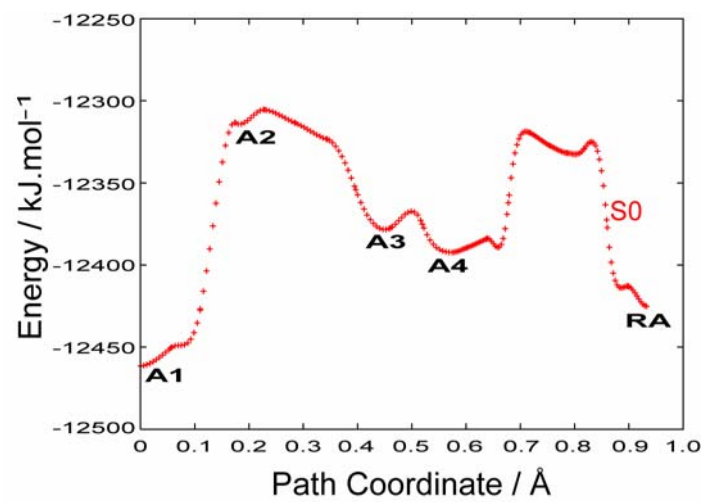


Figure 2.



Scheme 2.

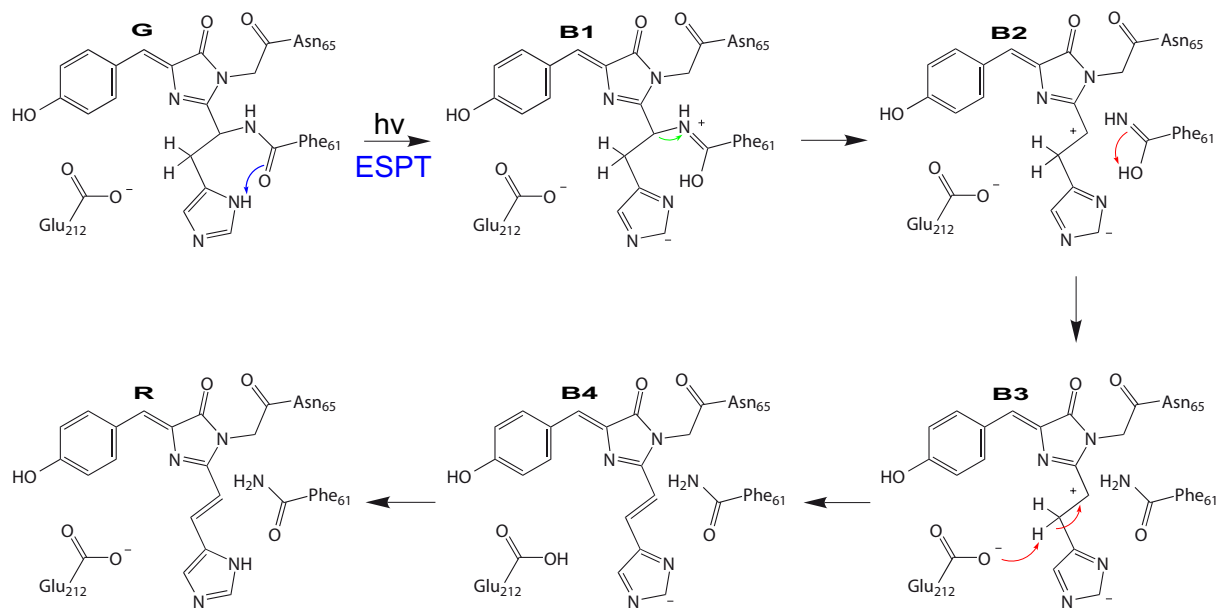


Figure 3.

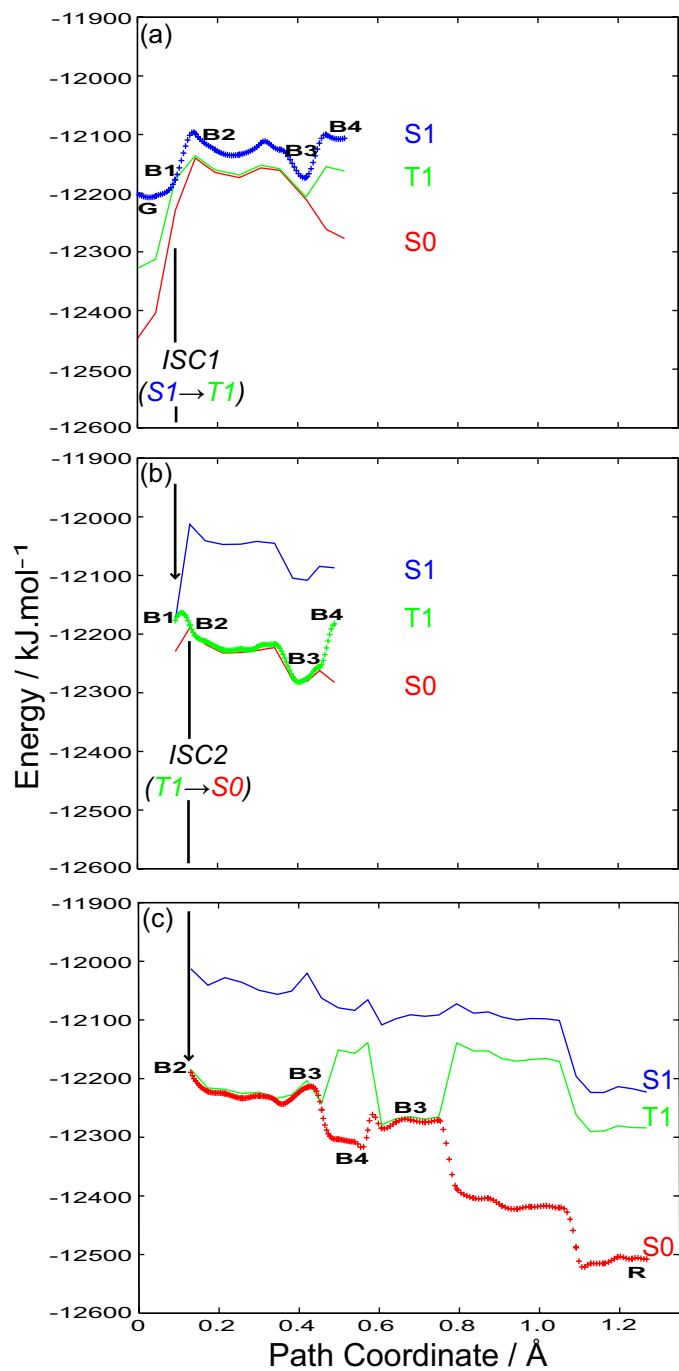
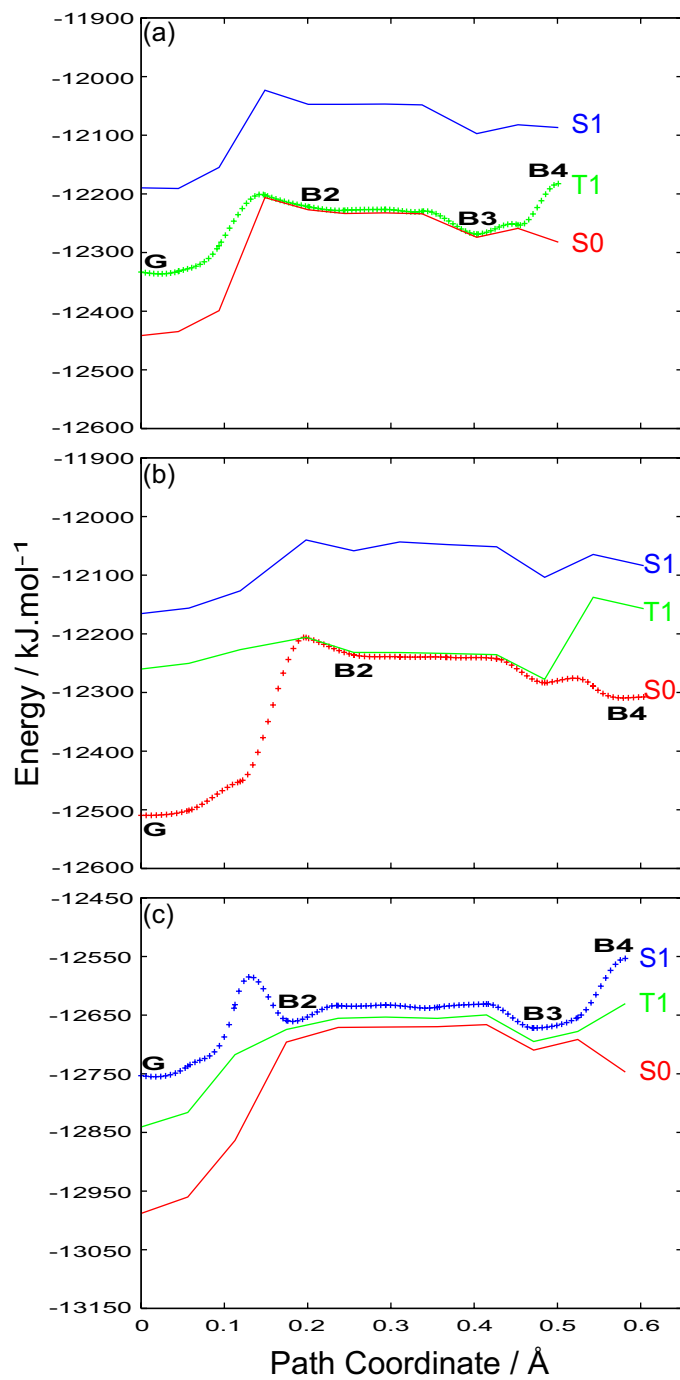


Figure 4.



Scheme 3.

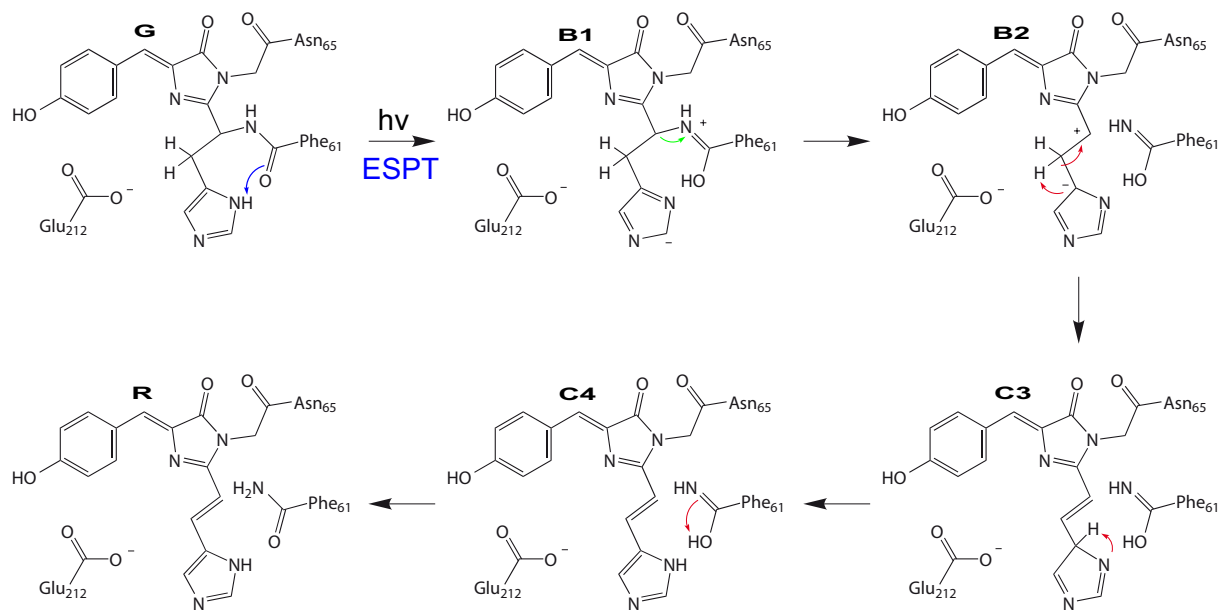


Figure 5.

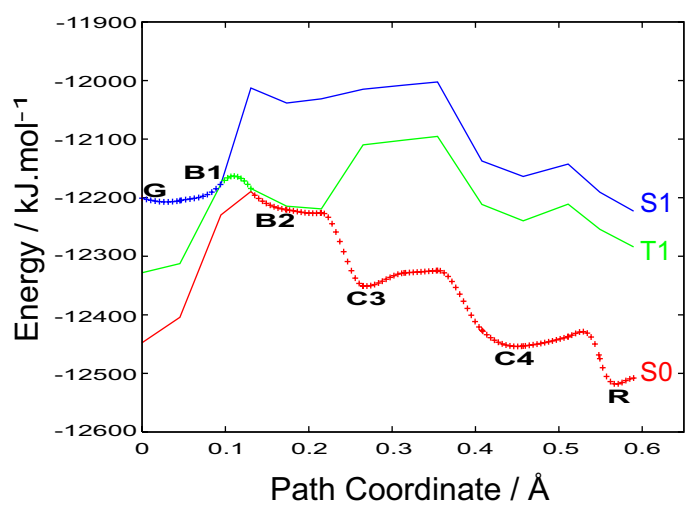


Figure 6.

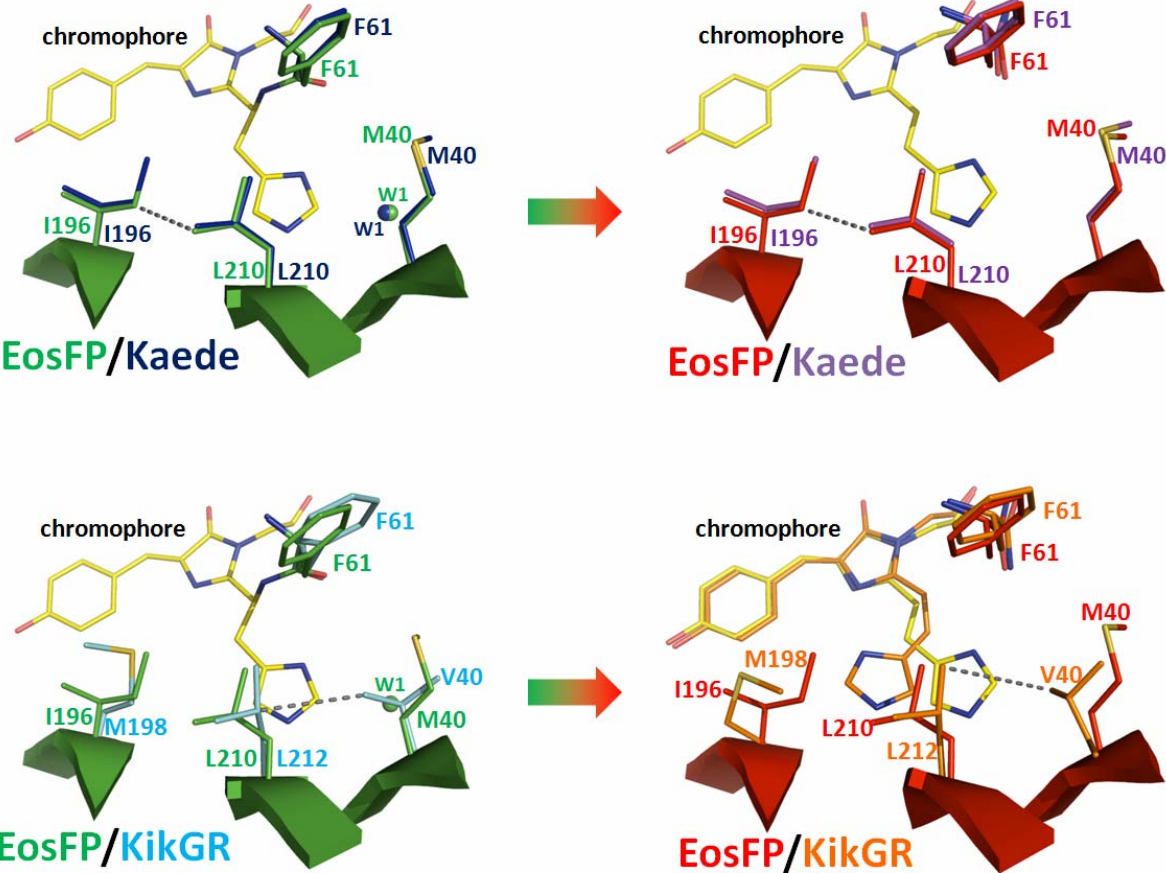


Figure 7.

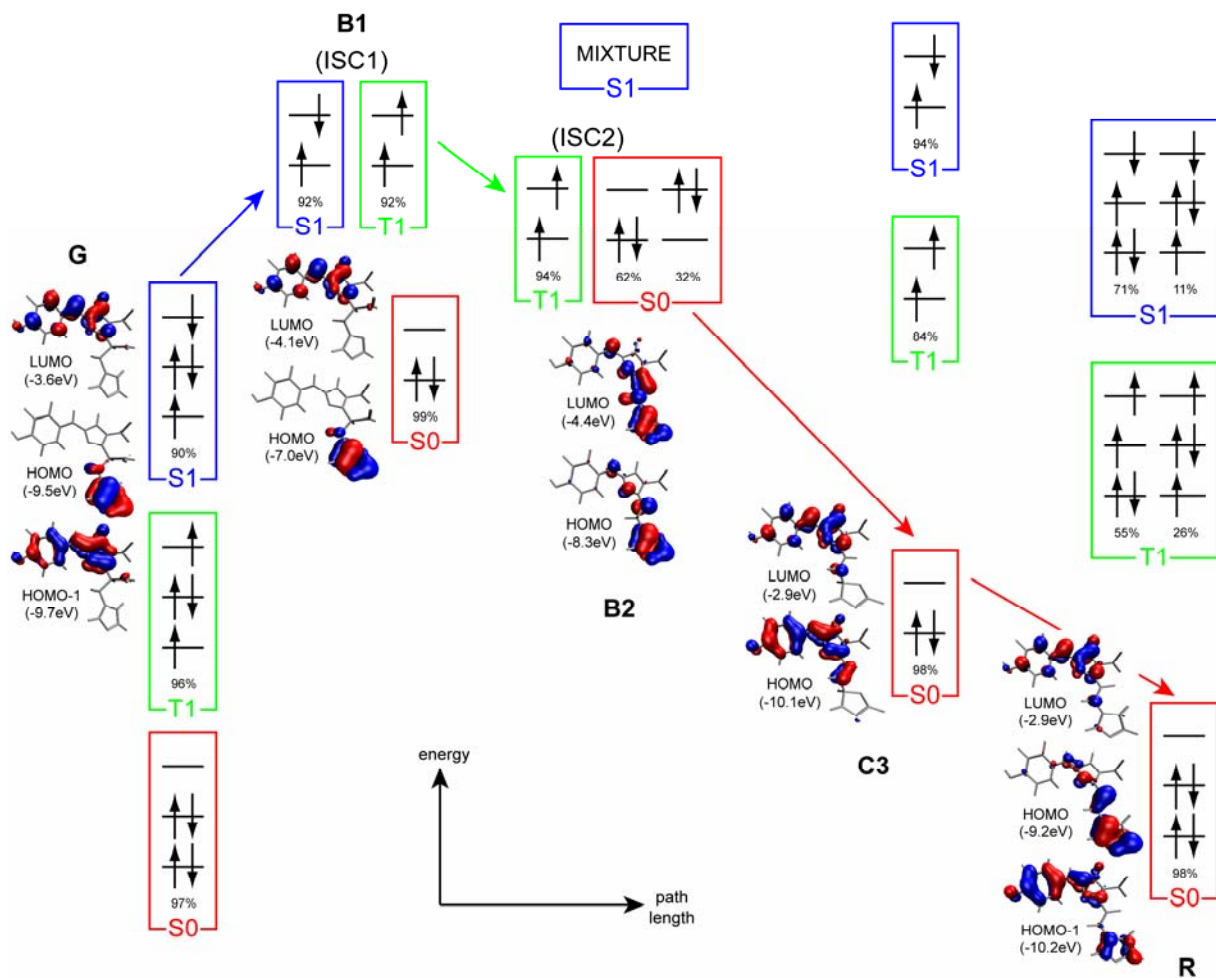


Figure 8.

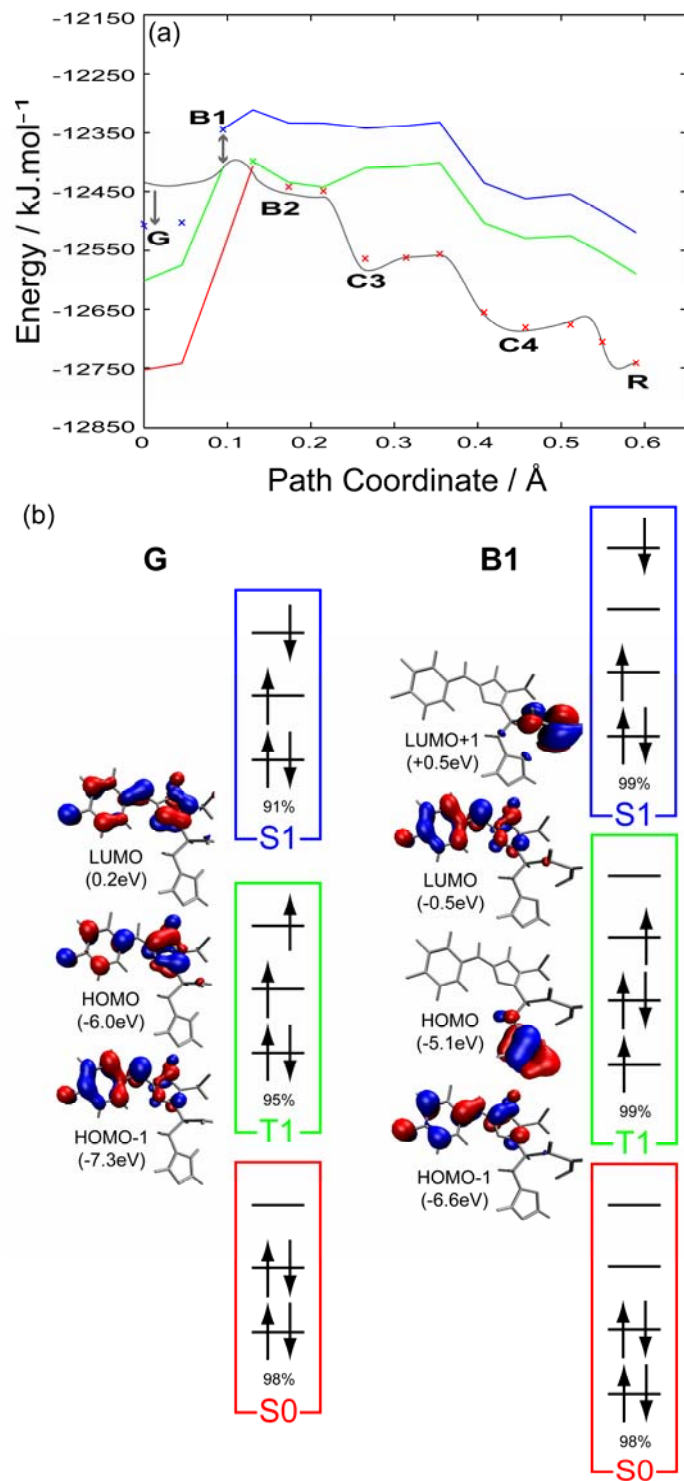


Figure 9.

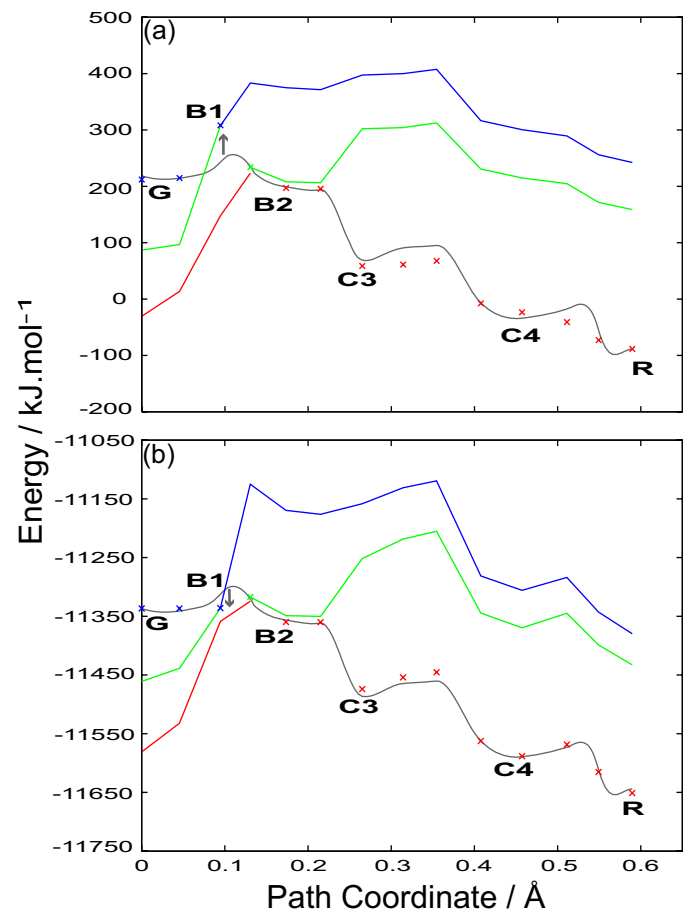
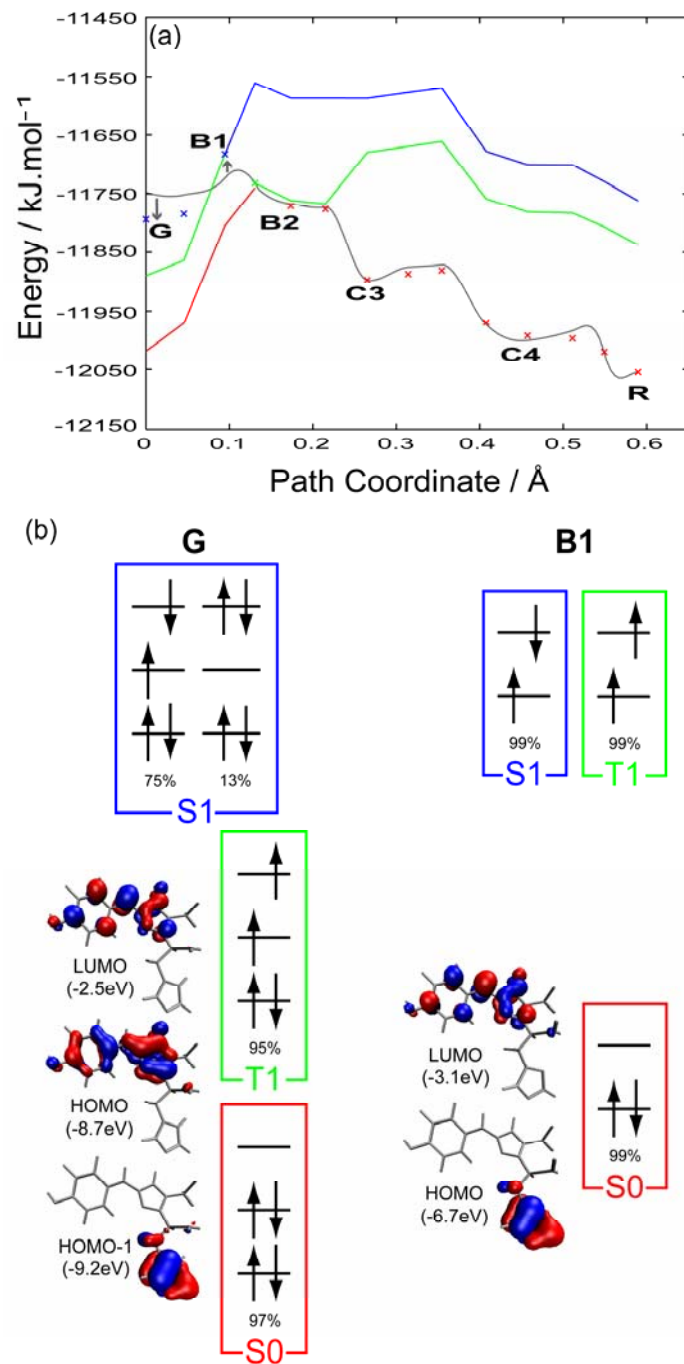


Figure 10.



SUPPORTING INFORMATION TO:

Photoconversion of the fluorescent protein EosFP: a hybrid potential simulation study reveals intersystem crossings

Mickaël Lelimosin, Virgile Adam, G. Ulrich Nienhaus, Dominique Bourgeois, and Martin J. Field

Table S1

Figures S1 to S9

Captions for Movies S1 to S4

Table S1. Comparison between measured and calculated wavelengths of absorption/emission for different states of EosFP. Deviations between calculated energies and experimental energies extracted from absorption/emission peak maxima are given in parentheses.

Measurement / Calculation	Method	Green states		Red states	
		anionic	neutral	anionic	neutral
λ_{\max} absorption, nm	UV-visible spectroscopy	506	390	571	460
λ S ₀ opt→S ₁ , nm		432	348	453	420
(deviation, kJ.mol ⁻¹) ₁	[10,9]-CISD(PDDG)/OPLS	(+ 40.7)	(+ 37.1)	(+ 54.8)	(+ 25.0)
λ S ₀ opt→S ₁ , nm		441	346	472	412
(deviation, kJ.mol ⁻¹) ₁	[10,9]-CISD(AM1)/OPLS	(+ 35.0)	(+ 39.1)	(+ 44.1)	(+ 30.3)
λ_{\max} emission, nm	UV-visible spectroscopy	516	-	581	-
λ S ₁ opt→S ₀ , nm		509	485	530	528
(deviation, kJ.mol ⁻¹) ₁	[10,9]-CISD(PDDG)/OPLS	(+ 3.0)	-	(+ 19.8)	-
λ S ₁ opt→S ₀ , nm		483	406	546	527
(deviation, kJ.mol ⁻¹) ₁	[10,9]-CISD(AM1)/OPLS	(+ 15.7)	-	(+ 13.1)	-

Figure S1. Energy profile of optimized NEB paths from A1 to A4 and from A4 to RA at the ground state S₀ calculated with a AM1/OPLS hybrid potential.

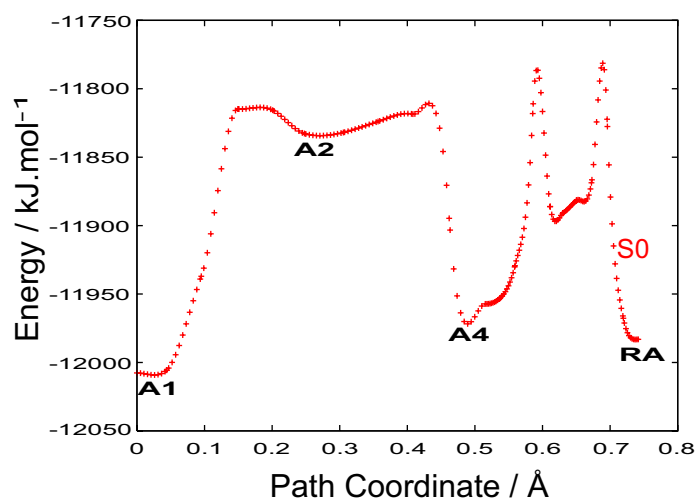


Figure S2. Energy profiles of optimized NEB paths calculated with a [10,9]-CISD(AM1)/OPLS hybrid potential: **(a)** from G to B4 at the excited state S1 (blue plus signs); **(b)** from B1 to B4 at the excited state T1 (green plus signs); and **(c)** from B1' to B4 and from B4 to R at the ground state S0 (red plus signs). Solid lines in each panel give the energies of the optimized NEB structures in the electronic states that were not used in the optimization.

It has to be notified that results are qualitatively similar as for PDDG calculations. Compared to the PDDG results, highest activation barriers are obtained along the NEB path at S1 **(a)**. This difference is due to the fact that energy-minimization of the AM1 path at S1 was unable to reach the desired convergence criteria. The same problem happened for the NEB path at T1 **(b)** and that may explain the additional step which is observed after ISC1: before backbone cleavage, B1 relaxes to an intermediate B1' which has a similar nuclear structure as B1 with a different electronic structure. B1' displays a lower energy on T1 than on S0. The barrier to break the C α -N bond on the T1 surface, though, remains relatively small at ~20 kJ/mol.

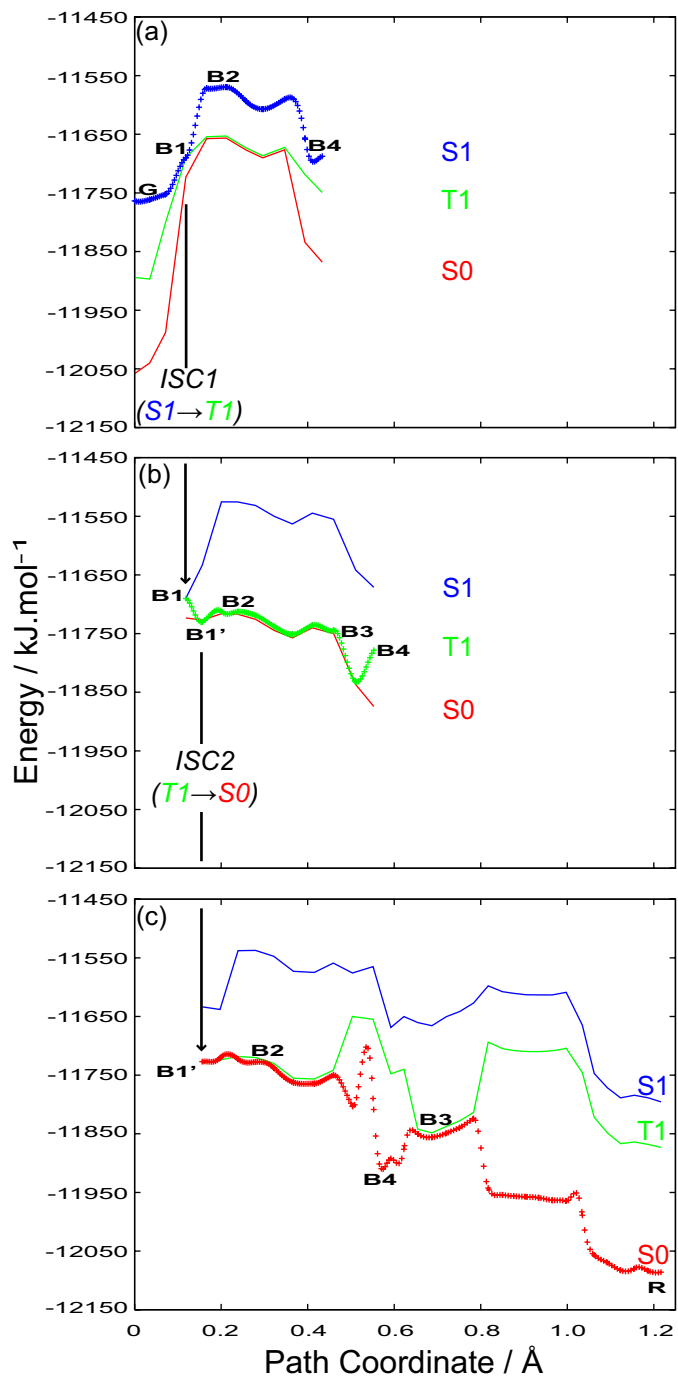


Figure S3. Energy profile of optimized NEB paths calculated with a [10,9]-CISD(AM1)/OPLS hybrid potential from G to B4: **(a)** at the excited state T1 (green plus signs); **(b)** at the ground state S0 (red plus signs); and **(c)** with a permanently anionic chromophore at the excited state S1 (blue plus signs). The solid lines give the energies of the optimized NEB structures but in the electronic states different from the one in which the optimization was performed.

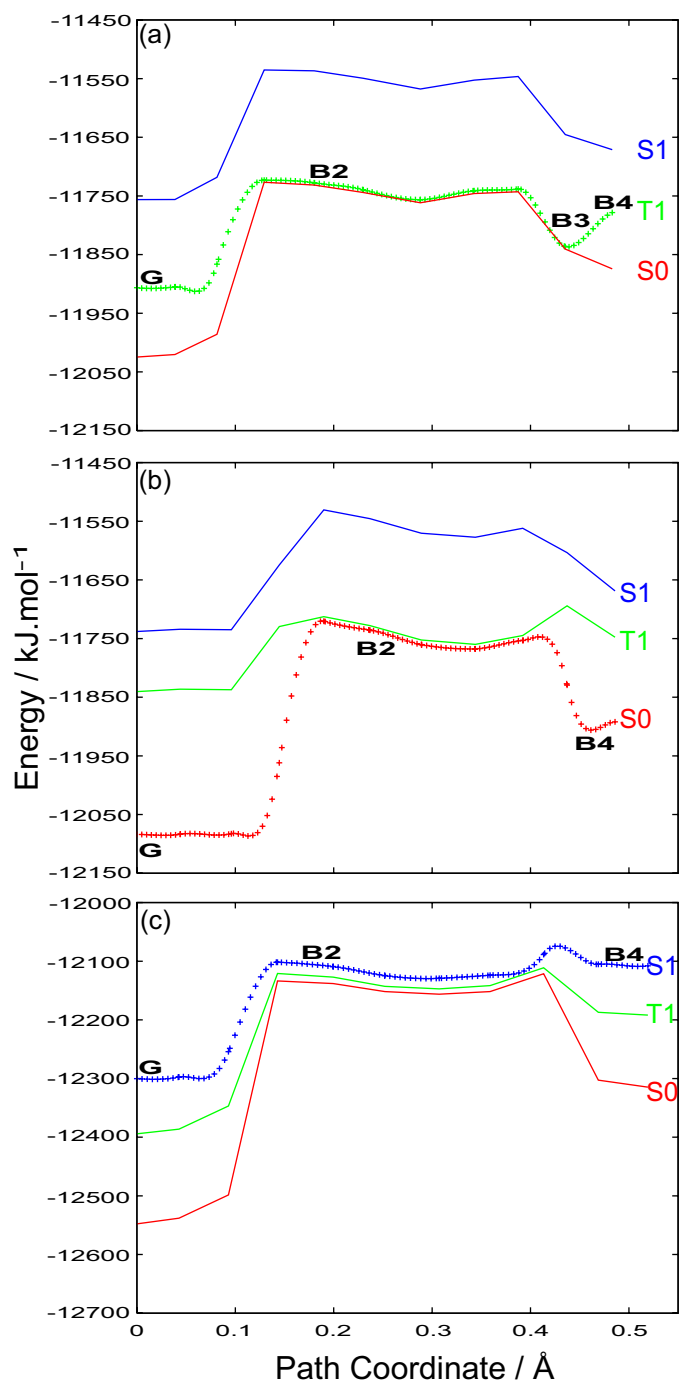


Figure S4. Energy profile of optimized NEB paths calculated with a [10,9]-CISD(AM1)/OPLS hybrid potential from G to B1 at the excited state S1 (blue plus signs), from B1 to B1' at the excited state T1 (green plus signs) and from B1' to R at the ground state S0 (red plus signs). The solid lines give the energies of the optimized NEB structures in electronic states that were not used during optimization.

As mentioned in the caption of **Figure S2**, the AM1 profile is qualitatively similar as the PDDG one, with an additional intermediate B1'. The step of backbone cleavage between B1' and B2 intermediates is calculated to have activation energy of only ~5 kJ/mol at the ground state S0. However, as intersystem crossing is a rare event, and it is possible that there would be time for breaking the C α -N bond on the T1 surface (see **Figure S2b**), before the ISC2 occurs.

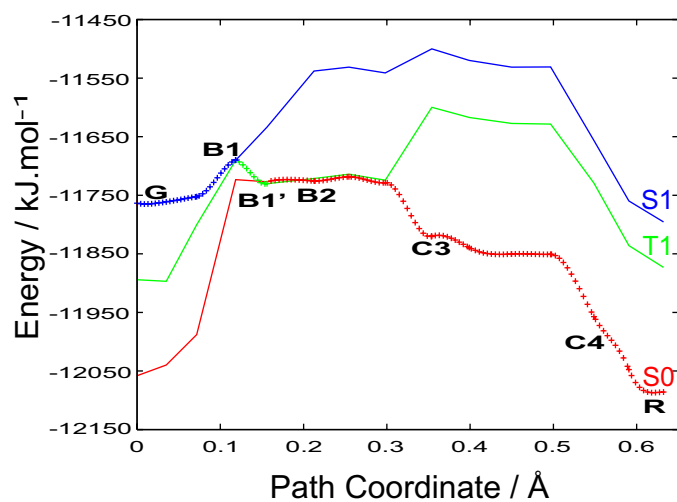


Figure S5. Energy profile of photoconversion calculated with a permanently anionic chromophore and a [10,9]-CISD(AM1)/OPLS hybrid potential. Energies of structures extracted from the real path (permanent neutral chromophore) are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states with a deprotonated chromophore. Crosses show the energies of the optimized path structures with the anionic chromophore, and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states.

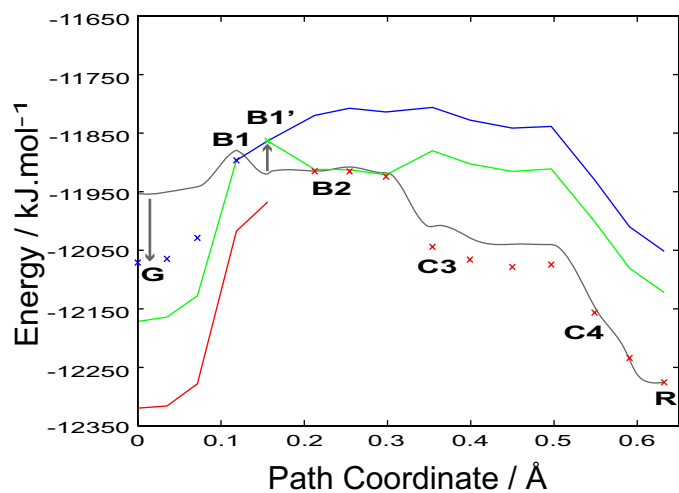


Figure S6. Energy profile of photoconversion calculated: (a) for the chromophore *in vacuo* with a [10,9]-CISD(AM1) potential; (b) *in protein* with null atomic charges (MM) on Glu212 and with a [10,9]-CISD(AM1)/OPLS hybrid potential. A new QC/MM partition is defined for the hybrid potential in which His194 and Glu212 are located in the MM region. Energies of structures extracted from the optimized NEB path are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states. Crosses show the energies of the optimized path structures with the modified potentials, and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states.

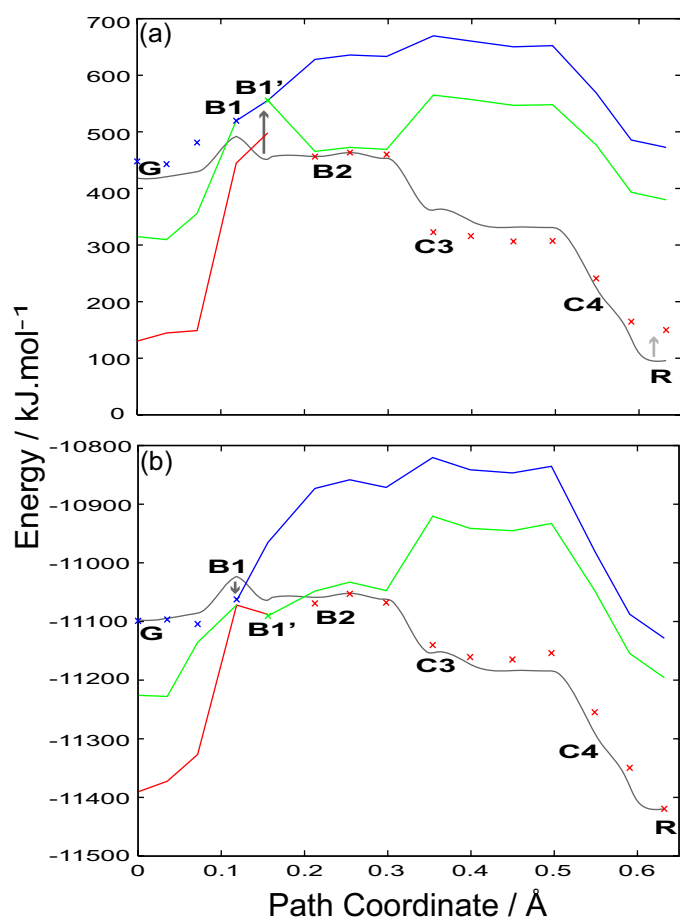


Figure S7. Energy profiles of photoconversion calculated with a modified QC/MM partitioning in which His194 and Glu212 are in the MM region: (a) a [10,9]-CISD(PDDG)/OPLS hybrid potential; and (b) a [10,9]-CISD(AM1)/OPLS hybrid potential. For each profile, energies of structures extracted from the real path (permanent neutral chromophore with initial QC/MM partitioning) are calculated for the S0 (red), T1 (green) and S1 (blue) electronic states with the modified hybrid potentials. Crosses show the energies of the optimized path structures with the modified hybrid potentials, and the real profile is shown in grey. The solid lines are the energies of the structures in the other electronic states.

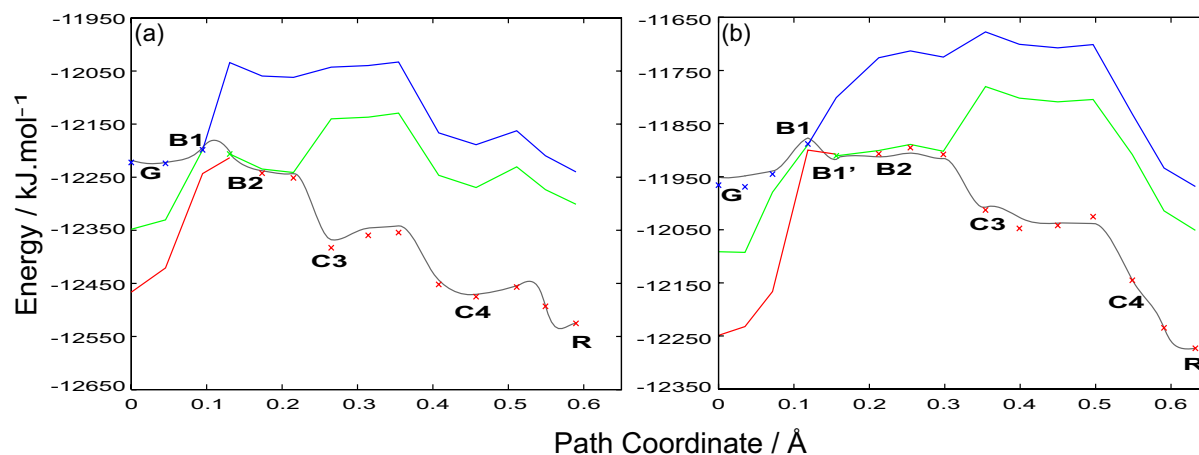


Figure S8. Energy profile of photoconversion calculated with null atomic charges (MM) on His194 and a [10,9]-CISD(AM1)/OPLS hybrid potential. The QC/MM partition is used in which His194 and Glu212 are located in the MM region. Energies of structures extracted from the real path (permanent neutral chromophore with initial QC/MM partitioning) are calculated in the S0 (red), T1 (green) and S1 (blue) electronic states. Crosses show the energies of the optimized path structures with the modified hybrid potential and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states.

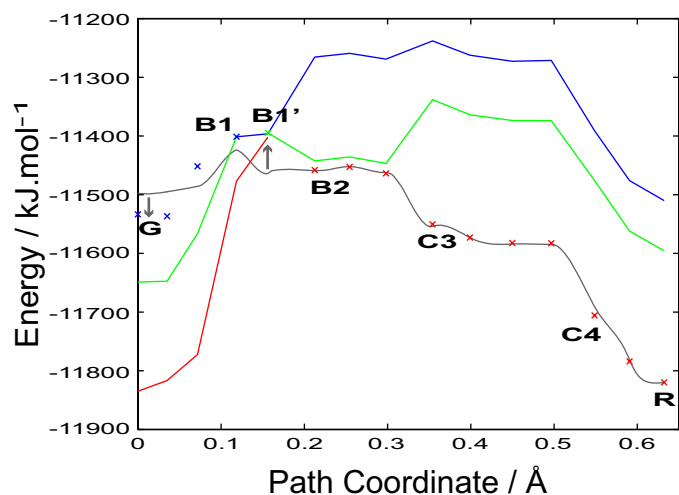
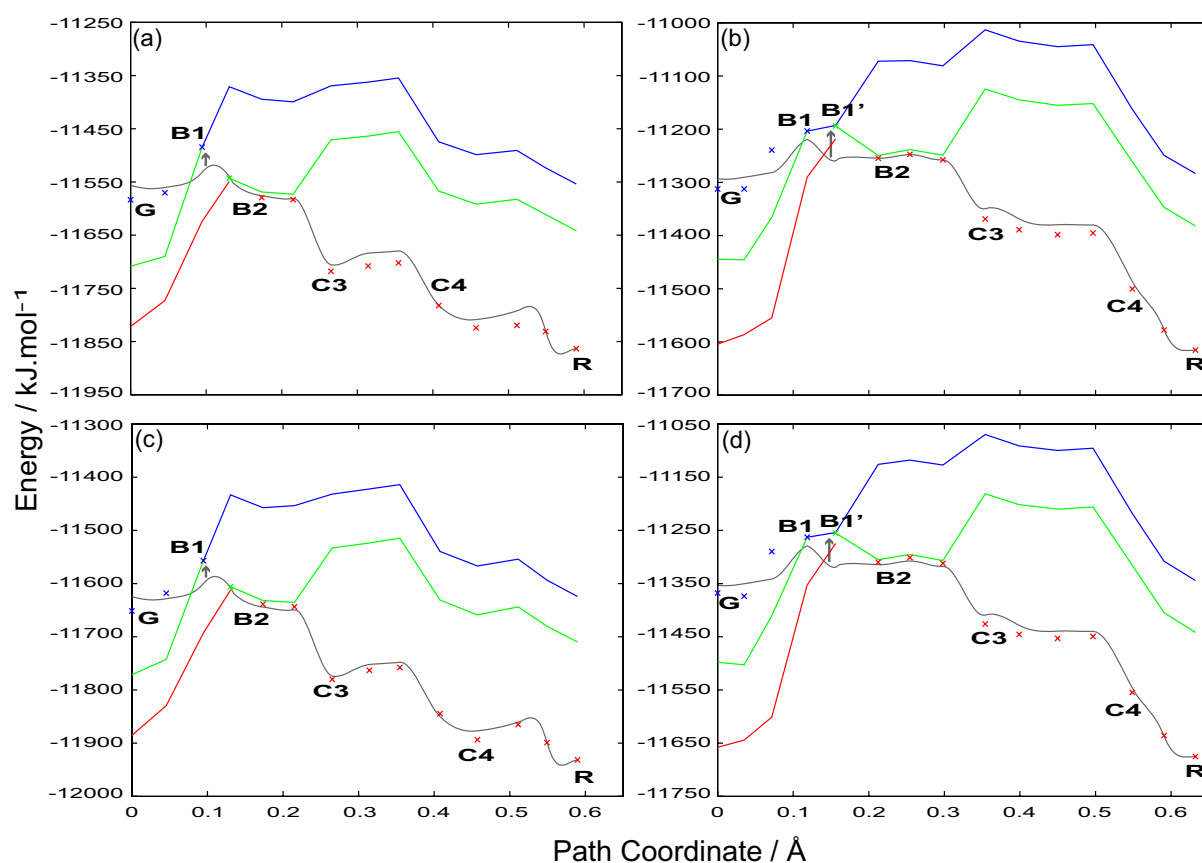


Figure S9. Energy profiles of photoconversion calculated with a modified QC/MM partitioning in which His194 and Glu212 are in the MM region. (a) and (c) are calculated with a [10,9]-CISD(PDDG)/OPLS hybrid potential and (b) and (d) with a [10,9]-CISD(AM1)/OPLS hybrid potential. (a) and (b) have null atomic charges (MM) for Arg66; (c) and (d) null atomic charges (MM) for Arg91. For each profile, energies of structures extracted from the real path (permanent neutral chromophore with initial QC/MM partitioning) are calculated for the S0 (red), T1 (green) and S1 (blue) electronic states with the modified hybrid potentials. Crosses show the energies of the optimized path structures with the modified hybrid potentials, and the main differences with the real profile (shown in grey) are indicated by arrows. The solid lines are the energies of the structures in the other electronic states.



Movie S1 Simulation of the photoconversion in EosFP determined by NEB calculations with the hybrid potential [10,9]-CISD(PDDG)/OPLS.

Movie S2 Simulation of the photoconversion in EosFP determined by NEB calculations with the hybrid potential [10,9]-CISD(AM1)/OPLS.

Movie S3 Hydrogen-bond partners of the water molecule W1 in EosFP during PDDG-simulation of the photoconversion.

Movie S4 Hydrogen-bond partners of the water molecule W1 in EosFP during AM1-simulation of the photoconversion.

Chapitre 6

Stabilité thermodynamique des différents états d'IrisFP

Une simple mutation dans EosFP a donné naissance à nouvelle protéine fluorescente nommé IrisFP. Ce mutant, développé par nos collaborateurs, présente des propriétés uniques de photoactivation. Il associe en effet la photoconversion irréversible de la forme sauvage à l'interruption réversible de la fluorescence des formes vertes et rouges. La combinaison de ces propriétés offre des possibilités d'application importantes aux protocoles émergents d'imagerie de super-résolution [123] ou encore aux techniques futures de stockage de masse sur des supports biologiques [124].

IrisFP constitue par ailleurs un archétype pour l'ingénierie des protéines fluorescentes. Nous avons tenté de comprendre comment une simple mutation peut engendrer de telles modifications des propriétés photophysiques. Notre démarche a été d'évaluer la stabilité thermodynamique des différents états d'IrisFP et de comparer celle-ci à des états existants ou virtuels chez EosFP. Une approche rigoureuse devrait être le calcul des différences d'enthalpie libre par des méthodes de perturbation ou d'intégration thermodynamiques [125, 126]. Dans l'article ci-dessous, nous montrons que de simples calculs d'enthalpies ont permis d'identifier des facteurs déterminants de la photophysique d'IrisFP.

Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations

Virgile Adam^a, Mickaël Lelimosin^b, Susan Boehme^c, Guillaume Desfonds^a, Karin Nienhaus^c, Martin J. Field^b, Joerg Wiedenmann^d, Sean McSweeney^a, G. Ulrich Nienhaus^{c,e,1}, and Dominique Bourgeois^{a,b,1}

^aEuropean Synchrotron Radiation Facility, 6 Rue Jules Horowitz, BP 220, 38043 Grenoble Cedex, France; ^bInstitut de Biologie Structurale Jean-Pierre Ebel, Commissariat à l'Énergie Atomique, Centre National de la Recherche Scientifique, Université Joseph Fourier, 41 Rue Jules Horowitz, 38027 Grenoble, France; ^cInstitute of Biophysics, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany; ^dNational Oceanography Centre, University of Southampton, Southampton SO14 3ZH, United Kingdom; and ^eDepartment of Physics, University of Illinois, Urbana, IL 61801

Edited by Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, New Mexico, and approved October 8, 2008 (received for review June 19, 2008)

Photoactivatable fluorescent proteins (FPs) are powerful fluorescent highlighters in live cell imaging and offer perspectives for optical nanoscopy and the development of biophotonic devices. Two types of photoactivation are currently being distinguished, reversible photoswitching between fluorescent and nonfluorescent forms and irreversible photoconversion. Here, we have combined crystallography and (*in crystallo*) spectroscopy to characterize the Phe-173-Ser mutant of the tetrameric variant of EosFP, named IrisFP, which incorporates both types of phototransformations. In its green fluorescent state, IrisFP displays reversible photoswitching, which involves *cis-trans* isomerization of the chromophore. Like its parent protein EosFP, IrisFP also photoconverts irreversibly to a red-emitting state under violet light because of an extension of the conjugated π -electron system of the chromophore, accompanied by a cleavage of the polypeptide backbone. The red form of IrisFP exhibits a second reversible photoswitching process, which may also involve *cis-trans* isomerization of the chromophore. Therefore, IrisFP displays altogether 3 distinct photoactivation processes. The possibility to engineer and precisely control multiple phototransformations in photoactivatable FPs offers exciting perspectives for the extension of the fluorescent protein toolkit.

fluorescent proteins | microspectrophotometry | photoactivation | photochromism | protein crystallography

In recent years, fluorescent proteins (FPs) from the green fluorescent protein (GFP) family have gained enormous popularity as genetically encoded fluorescence markers (1). They enable the visualization of a broad range of biological processes, including gene expression, protein translocation within cells, and cell movement during development. Photoactivatable FP variants have been recognized as particularly powerful imaging tools. Their fluorescence emission intensity or color can be controlled by irradiation with light of a specific wavelength, which allows selective highlighting of subsets of protein molecules within cells. This capability has been exploited in applications ranging from quantitative studies of protein movements in live cells (2, 3) to optical nanoscopy (4–6), and photoactivatable FPs may even be used in data-storage devices (7).

Two different groups of photoactivatable FPs are presently being distinguished. Members of the first group undergo reversible photoswitching between a fluorescent on state and a nonfluorescent off state (8–10), whereas members of the second group display irreversible photoconversion, either from a nonfluorescent to a fluorescent state (11–13) or between 2 fluorescent states with different emission wavelengths (14–19).

The structural basis of reversible photoswitching has been elucidated by crystallographic studies of several FPs including Dronpa (20), mTFP0.7 (9), and asFP595 (21). In the absence of light, the 4-(*p*-hydroxybenzylidene)-5-imidazolinone chro-

mophores of Dronpa and mTFP0.7 assume a *cis* conformation, in which the anionic, fluorescent state of the chromophore is predominant at physiological pH. Light-induced off-switching of the fluorescence was proposed to arise from a *cis-trans* photoisomerization of the chromophore accompanied by a change of its protonation state (22–24), a loss of chromophore planarity and, for mTFP0.7, a more disordered chromophore structure in the *trans* conformation (9). Alternatively, the loss of fluorescence in the off state was recently explained by light-induced protonation of the Dronpa chromophore, associated with increased chromophore flexibility, which enhances nonradiative deactivation pathways (25). In this scenario, *cis-trans* isomerization may occur as a side effect. In asFP595, the nonfluorescent *trans* state is thermodynamically more stable than the *cis* state, and light irradiation induces a *trans-cis* isomerization to the *cis* state (21), in which a minor, zwitterionic state has been proposed to promote fluorescence (26). When kept in the dark, reversibly photoswitching FPs relax within minutes to several hours to the thermodynamically more stable isomeric state.

For irreversible photoconversion, several mechanisms have been identified. For EosFP, a photoconverting FP that changes its emission color from green to red upon irradiation with violet light (18), the X-ray structures have been determined for the green and red forms (27, 28). They revealed a light-induced extension of the conjugated π -electron system of the chromophore, accompanied by a breakage of the protein backbone between the amide nitrogen and the α -carbon of His-62 (EosFP residue numbering is used throughout this article), with surprisingly little change of the overall protein structure (27). Kaede (29), KikGR (17), and Dendra (14) are further representatives of green-to-red photoconverters that all share the tripeptide His-62–Tyr-63–Gly-64 as the chromophore-forming unit. A β -elimination reaction has been invoked to explain backbone cleavage, which may involve excited state proton transfer (ESPT) from the neutral chromophore to His-62 (27, 29), possibly assisted by dislocation of a water molecule (30). For PA-GFP (11) and PS-CFP (13), irreversible photoactivation

Author contributions: V.A., K.N., G.U.N., and D.B. designed research; V.A., M.L., S.B., and G.D. performed research; M.J.F., J.W., and S.M. contributed new reagents/analytic tools; V.A., M.L., S.B., and G.D. analyzed data; and V.A., M.L., K.N., G.U.N., and D.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 2VVH, 2VVI, and 2VVJ).

¹To whom correspondence may be addressed. E-mail: dominique.bourgeois@ibs.fr or uli@uiuc.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805949105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

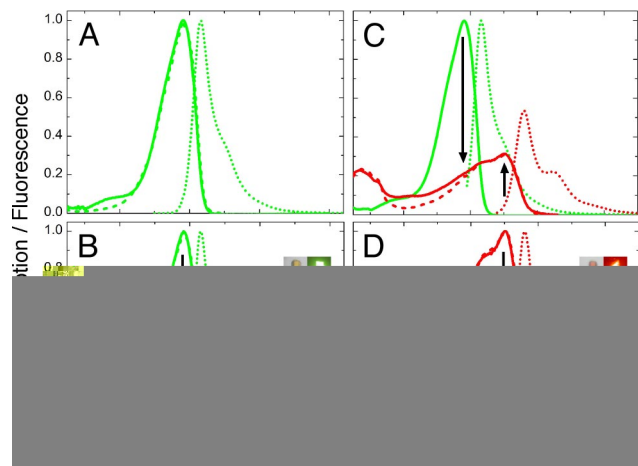


Fig. 1. Spectroscopic characterization of IrisFP. Absorption, excitation, and emission spectra scaled to equal maximum amplitudes are depicted by solid, dashed, and dotted lines, respectively. Emission spectra of green (red) IrisFP were obtained by exciting at 488 (532) nm. Excitation spectra of green (red) IrisFP were obtained by detecting at 550 (620) nm. (A) Green IrisFP. (B) Green IrisFP before (green lines) and after (gray lines) illumination with 488-nm light. (C) Red IrisFP (red lines) after photoconversion of green IrisFP (green lines) with 405-nm light. (D) Red IrisFP before (red lines) and after (gray lines) illumination with 532-nm light. Spectra were recorded on solution samples in potassium phosphate buffer, pH 9. *Insets* show pictures of green IrisFP crystals before (*B Upper*) and after (*B Lower*) 488-nm illumination and of red IrisFP crystals before (*D Upper*) and after (*D Lower*) 532-nm illumination, respectively. Crystals are shown in bright-field mode (*Insets, Left*) and in fluorescence mode (*Insets, Right*). Detailed experimental procedures are provided in *SI Text*.

arises from stabilization of the anionic chromophore upon irradiation with violet light, presumably due to decarboxylation of Glu-212, as was earlier observed for GFP (31). Irreversible photoactivation has also been noticed for PA-mRFP1 (12) and the asFP595 Ala-148-Gly variant, also known as KFP1 (32); the underlying structural changes have yet to be elucidated.

The presence of structurally diverse photoactivation mechanisms suggests the intriguing possibility that multiple photo-transformations can be implemented within a single photoactivatable FP. Such a highlighter would enable sophisticated photoactivation schemes that could help unravel complex protein interaction patterns. Here, we introduce IrisFP, a variant of the tetrameric form of EosFP, which can be switched reversibly between fluorescent and nonfluorescent forms in its green state, then photoconverted from green to red, and finally switched reversibly between fluorescent and nonfluorescent forms in its

red state. A structural characterization of IrisFP based on a combination of X-ray crystallography and UV-visible spectrophotometry on crystal and solution samples provides a detailed picture of the different states of IrisFP at near-atomic level.

Results

Spectroscopic Characterization. In the course of random mutagenesis experiments on EosFP, the mutant Phe-173-Ser was generated. We discovered that this mutant undergoes multiple photoactivation processes and, therefore, named it IrisFP (the Greek goddess Iris personifies the rainbow). Optical absorption and emission spectra of IrisFP solution samples and photographs of crystals recorded after exposure to a variety of illumination conditions are shown in Fig. 1. The corresponding spectra measured on single crystals are presented in [supporting information \(SI\) Fig. S1](#); spectroscopic parameters determined on solution samples are compiled in Table 1. In the green form of IrisFP (Fig. 1*A* and Fig. *S1A*), the main absorption band peaks at 487 nm; it is flanked by a vibronic shoulder at ≈ 460 nm. A minor absorption band resides at 390 nm. As for EosFP (18), the pH dependence of the absorption spectra (Fig. *S2*) reveals that the 487- and 390-nm bands are associated with the anionic and neutral states of the chromophore, respectively. Fluorescence emission, peaking at 516 nm, appears upon excitation of the anionic chromophore; the neutral form is only marginally fluorescent. The broader absorption/excitation bands of the anionic species [full width at half maximum (FWHM) 55 nm] as compared with EosFP (FWHM 25 nm) may reflect enhanced conformational freedom of the chromophore in IrisFP, which is also consistent with its decreased fluorescence quantum yield ($\Phi = 0.43$, Table 1).

Illumination of IrisFP solutions or crystals at room temperature by 488-nm light leads to pronounced changes in the optical spectra (Fig. 1*B* and Fig. *S1B*). The absorption band of the anionic species at 488 nm decays rapidly, concomitant with an increase of the band of the neutral species at 390 nm. A crisp isosbestic point at 426 nm (data not shown) suggests a light-driven conversion of the chromophore from anionic to neutral. The resulting species is essentially nonfluorescent. Such experiments are usually difficult to perform with crystals, owing to their high optical density (in IrisFP crystals, the penetration depth is only ≈ 40 μm at 488 nm). Nevertheless, an almost complete loss of fluorescence can be achieved *in crystallo* because progressive deactivation of layers of molecules in the crystal ensures penetration of the actinic light into the bulk. The nonfluorescent state of the chromophore is very stable; it persists for hours in the dark ($t_{1/2} \approx 5.5$ h in solution). The recovery (on switching) is markedly accelerated by illuminating with 405-nm light, occurring within seconds. The quantum yields for off and on switching are estimated as $\Phi_{\text{off}} = 0.014$ and $\Phi_{\text{on}} =$

Table 1. Spectroscopic properties of IrisFP in solution

Parameter	Green IrisFP	Red IrisFP	Green EosFP	Red EosFP
$\lambda_{\text{max, ex/em}}$, nm	488 / 516	551 / 580	506 / 516	571 / 581
ϵ , $\text{M}^{-1} \text{cm}^{-1}$	52,200	35,400	72,000	41,000
QY _{fluorescence}	0.43 ± 0.02	0.47 ± 0.02	0.70 ± 0.02	0.62 ± 0.03
QY _{off switching} *	0.014	0.0020	—	—
QY _{on switching} *	0.5	0.047	—	—
QY _{green-to-red conversion}	0.0018	—	0.0008	—
Eff _{off} †	≈ 0.85	≈ 0.75	—	—
$t_{1/2}$, thermal recovery, h	5.5	3.2	—	—

Detailed experimental procedures are described in *SI Text*. λ_{max} , peak wavelength; ex, excitation; em, emission; ϵ , extinction coefficient; QY, quantum yield; $t_{1/2}$, half-life.

*Standard deviations are estimated as 25%.

†Switching efficiencies based on the following illumination conditions: Green IrisFP: 473 nm, 40 mW, ≈ 10 min. Red IrisFP: 532 nm, 50 mW, ≈ 10 min.

0.5, respectively. Off/on switching can be repeated several times in solution and in the crystal with minimal loss of fluorescence due to photobleaching (Fig. S3). Overall, the reversible switching behavior of green IrisFP resembles previous observations made with Dronpa (23) or mTFP0.7 (9), which suggests that IrisFP may undergo light-activated chromophore isomerization.

When stimulated by 405-nm laser light, the green IrisFP chromophore turns red (Fig. 1C and Fig. S1C). This photoconversion process gives rise to changes in the absorption and excitation spectra analogous to those reported earlier for the parent protein EosFP (18). The various bands can be assigned to differently charged chromophore species based on their response to pH changes (Fig. S2) and by comparison with EosFP. The excitation spectrum shows that the anionic species absorbs at 551 nm (with a vibronic shoulder at ≈ 515 nm) and fluoresces at 580 nm, whereas the neutral species, with a broad absorption feature at ≈ 450 nm, does not contribute to the fluorescence at 580 nm. Higher-order electronic transitions of the red chromophore are visible in the spectral region <400 nm. Prolonged exposure to violet light results in a slow decay of red fluorescence, possibly due to photobleaching. However, as a result of intramolecular Förster resonance energy transfer (FRET) from green to red subunits within the tetramer, already a moderate 405-nm light exposure yields an almost complete change to red fluorescence. The quantum yield of photoconversion is enhanced as compared with that of the parent protein ($\Phi_{\text{green/red}} = 1.8 \times 10^{-3}$, Table 1).

Exposure of red-converted IrisFP to green light (532 nm) induces yet another photoswitching event, characterized by a decrease in the red fluorescence (Fig. 1D and Fig. S1D). Concomitantly, absorption at 551 nm decreases and absorption at 450 nm increases, which suggests that a neutral form of the chromophore becomes stabilized. Under our experimental conditions, a maximum switching yield of $\approx 75\%$ could be achieved (Table 1). In the dark, the new species slowly reverts to the red fluorescent state ($t_{1/2} \approx 3.2$ h). The return is strongly accelerated by exposure to 440-nm light. As for the green species, off/on switching can be repeated several times in solution and in the crystal with minimal loss of fluorescence due to photobleaching (Fig. S3). The quantum yields of photoactivated off and on switching of the red form are estimated to be $\Phi_{\text{off}} = 0.0020$ and $\Phi_{\text{on}} = 0.047$, respectively. Interestingly, at moderate levels of photoconversion to the red form, off-switching of the red fluorescence results in an increase in green fluorescence (Fig. S4), which arises from the loss of intramolecular energy transfer from green to red monomers within tetrameric molecules.

X-Ray Crystallographic Characterization. To reveal the structural basis of the spectroscopically observed photoactivation processes in IrisFP, we performed X-ray structure analyses on crystals that were exposed to light conditions chosen to enhance particular species. Data collection and refinement statistics are compiled in Table S1. The X-ray structure of the green form of IrisFP, crystallized in the dark, was determined at a resolution of 1.8 Å without prior light exposure. The asymmetric unit consists of a well-defined tetramer. The individual subunits of IrisFP show the classical β -can fold, with no major deviation from the EosFP structure (27), as indicated by the backbone rms deviation of 0.36 Å between the 2 models. The anionic chromophore is slightly nonplanar (Table S1), which may account for the lower fluorescence quantum yield as compared with EosFP (Table 1). The phenolate moiety is stabilized by a hydrogen bond to Ser-142 and 2 water molecules (W-2188 and W-2017, Fig. 2 and S5A). The remarkable photoactivation properties of IrisFP result from the replacement of Phe-173 by serine. In EosFP, Phe-173 forces the side chain of Met-159 to establish van der Waals contact with the hydroxybenzylidene moiety (Fig. 2). The smaller Ser-173 side chain in IrisFP reduces these steric con-

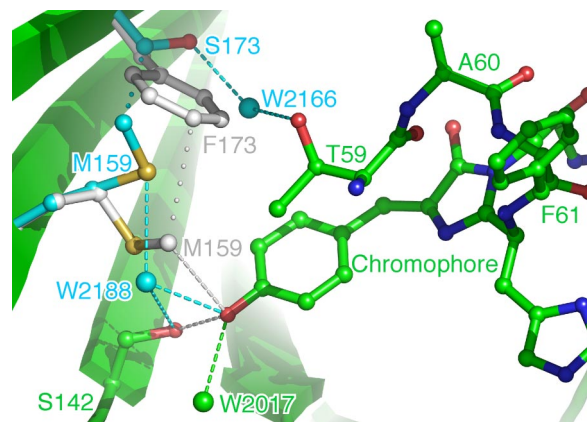


Fig. 2. Changes in the chromophore environment induced by the F173S mutation in EosFP. Common structural elements of EosFP and IrisFP are represented with green carbon atoms. Residues Phe-173 and Met-159 in EosFP are represented with gray carbon atoms, whereas the mutated residue Ser-173 and residue Met-159 in IrisFP are represented with cyan carbon atoms. The 2 water molecules (W-2188 and W-2166) are represented with cyan spheres, and W-2017 is represented with a green sphere. Hydrogen bonds are shown with dashed lines and van der Waals interactions with dotted lines.

straints, and the Met-159 side chain rotates away from the chromophore. Two cavities are created (total volume ≈ 13 Å³), each filled with a water molecule (W-2188 and W-2166, Fig. 2). The less-densely packed chromophore environment of IrisFP may lead to enhanced flexibility of the chromophore, which is in line with the spectroscopic observations that, compared with EosFP, the absorption and emission bands are broader and the fluorescence quantum yield is reduced.

To examine the structural basis of photoswitching of the green form of IrisFP, we illuminated a green crystal with 488-nm laser light and flash-cooled it immediately afterward. A diffraction dataset was collected to a resolution of 2.0 Å. An electron density omit map is shown in Fig. S5B. From the superposition of the refined models of the fluorescent and nonfluorescent forms in Fig. 3A, it is evident that off switching is accompanied by a *cis-trans* isomerization of the chromophore, as was earlier reported for Dronpa (20) and mTFP0.7 (9). The *trans* isomer of the chromophore is nonplanar (Table S1). Substantial rearrangements of residues Ser-142, His-194, Glu-212, Arg-66, and Ile-157 in the chromophore cavity are also apparent (Fig. 3A). The hydrogen bond between the phenolate oxygen of the *cis* chromophore and the Ser-142 hydroxyl side chain is removed and Ser-142 achieves a double conformation, either reorienting toward the solvent or H-bonding to His-194 and W-2188. The *trans* chromophore hydrogen-bonds to the presumably deprotonated carboxylate group of Glu-144 and to W-2032, most probably assuming a protonated phenolic form consistent with the observed loss of fluorescence and the increased absorption at 390 nm. The hydrogen-bonding network linking Glu-212 and Glu-144 via His-194 in the *cis* conformation involves Arg-66 in the *trans* conformation, implying large conformational rearrangements of the latter 2 residues similar to those observed in Dronpa (20) or mTFP0.7 (9). In addition, the side chain of Ile-157 undergoes a flip of $\approx 145^\circ$, possibly because of transient repulsion during chromophore isomerization. Based on these structures, the enthalpy difference between the 2 isomeric forms of green IrisFP was estimated from QM/MM molecular dynamics as 68 kJ/mol (SI Text and Table S2). This value is substantially smaller than the one estimated for EosFP (114 kJ/mol), assuming a similar *trans* structure. Calculations performed with null atomic charges for Ser-142 suggest that the stability of the *trans* isomeric state largely depends on the electrostatic interactions

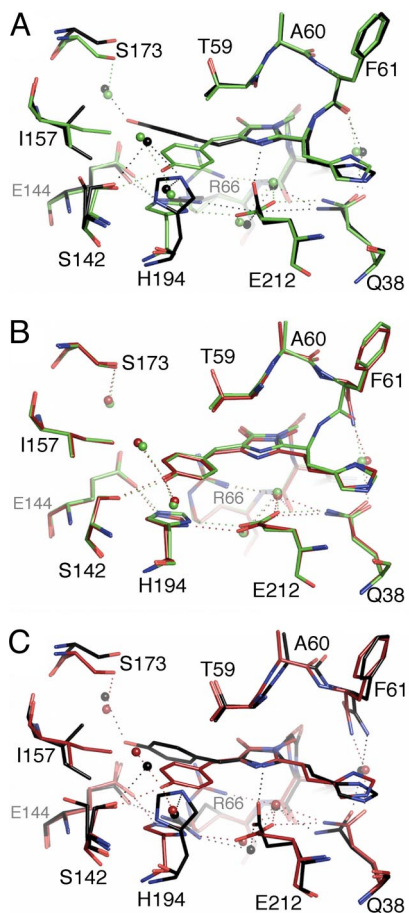


Fig. 3. Structural changes of the chromophore pocket upon phototransformation of IrisFP. (A) Superposition of the native green state (green) and the first reversibly switched state (black) obtained upon illumination at 488 nm. (B) Superposition of the native green state (green) and the irreversibly photoconverted red state (red) obtained after illumination at 405 nm. (C) Superposition of the red state (red) with the second reversibly switched state (black) obtained upon illumination at 405 nm, followed by illumination at 532 nm. In C, a putative model of the *trans* state of red IrisFP is shown (black). Hydrogen bonds are shown as dotted lines. Water molecules are shown as spheres.

between this residue and its environment. These interactions are much more favorable in IrisFP than in EosFP (Table S3). Diffraction data collected on a green IrisFP crystal after recovery in the dark from the off state confirmed the reversibility of the *cis*–*trans* isomerization, as the resulting structure was indistinguishable from the one of the initial fluorescent state (data not shown).

To determine the structure of red IrisFP, a green crystal was illuminated by 405-nm light and flash-cooled to collect a diffraction dataset at 2.0-Å resolution. Backbone cleavage between Phe-61 and His-62 was firmly established on the basis of omit electron density maps (Fig. S5C). Except for this modification, the superposition of the modeled chromophore environments of the green and red forms in Fig. 3B shows a nearly unperturbed chromophore pocket, as was observed earlier for red EosFP (27). A water molecule proximal to His-62 is dislocated in the red form, as was recently also reported for Kaede (30). These results support the notion of identical green-to-red photoconversion mechanisms for IrisFP and EosFP. Signs of radiation damage that could be caused by the long violet-light exposure of the sample, such as decarboxylation of Glu-212, were not noticeable.

To investigate the structural basis of reversible photoswitching of the red form of IrisFP, a green crystal was irreversibly

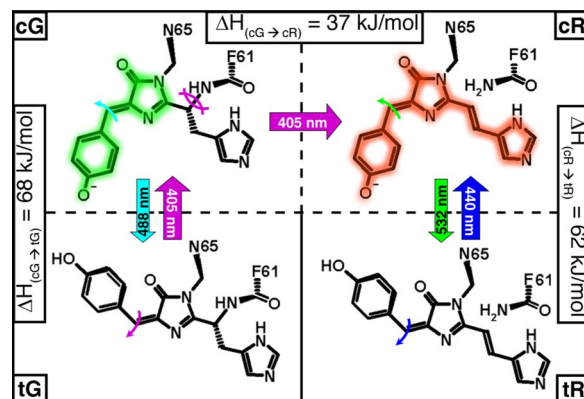


Fig. 4. Photoinduced transformations in IrisFP. Structural motions induced by light are represented by curved arrows of the same color as those used to represent light illumination at specific wavelengths. CG, *cis*-green Iris; tG, *trans*-green Iris; cR, *cis*-red Iris; tR, *trans*-red Iris.

photoconverted to the red form by illumination with 405-nm light and subsequently exposed to 532-nm light to induce reversible switching. After flash-cooling, diffraction data were collected to 2.2-Å resolution. In addition to the features characteristic of green-to-red photoconversion of IrisFP, the electron density map shows the *p*-hydroxybenzylidene ring in both the *cis* and *trans* conformations (Fig. S5D). This observation suggests a partial *cis*–*trans* isomerization of the chromophore, with features similar to those displayed by green off-switched IrisFP (Fig. 3A). However, the multiple conformations coexisting in the crystal, possibly as a result of limited penetration of green light within the crystal, made a reliable structural refinement difficult. Thus, assuming a *trans* conformation of the chromophore, only a tentative, energy-minimized model of the off-switched state could be produced (Fig. 3C and Fig. S5D). Based on this model, the enthalpy difference between the 2 isomeric states of red IrisFP was calculated to be 62 kJ/mol (Table S2). As for the green form, this difference is much higher for wild-type EosFP (122 kJ/mol) because of the less-favorable interactions of Ser-142 with its environment.

Discussion

From the spectroscopic and crystallographic data presented here, it is evident that 2 photoactivation mechanisms occur in IrisFP: reversible photoswitching between a fluorescent and a nonfluorescent state, based on *cis*–*trans* isomerization of the chromophore, and irreversible green-to-red photoconversion between 2 different fluorescent forms of the chromophore. Fig. 4 summarizes the different optical states of IrisFP, highlighting the richness of its photophysics.

Superposition of the structures of IrisFP, Dronpa, and mTFP0.7 in their fluorescent *cis* and nonfluorescent *trans* states (Fig. S6) suggests that reversible photoisomerization in these proteins proceeds by similar mechanisms. In all 3 proteins, rotation of the chromophore is coupled to concerted motions of His-194 and Arg-66. Glu-212 and Glu-144 act as nanotweezers, providing hydrogen bonds to His-194 in the *cis* configuration and to Arg-66 in the *trans* configuration and, thereby, stabilizing both conformations. However, this common configuration is clearly not sufficient to enable photoswitching, as it is also found in the *cis* state of nonphotoswitchable FPs such as EosFP or Kaede (29). Stabilization of the *trans* conformation of the chromophore requires additional interactions that, in particular, compensate for the cost of breaking the hydrogen bond linking the phenolate oxygen of the benzylidene group to Ser-142 in the *cis* state. Comparison between EosFP and IrisFP allows a precise identification of these additional interactions. Our enthalpy calcu-

lations (Table S3) suggest that an essential factor is the availability of a polar environment around Ser-142 in the *trans* state. In IrisFP, such an environment is provided by the repositioning of Met-159 and by the introduction of water molecule W-2188. As a consequence, Ser-142 may either hydrogen-bond to His-194 and W-2188 or reorient toward the solvent. Our molecular-dynamics simulations suggest that these 2 options are prohibited in EosFP because of the hydrophobic character of Met-159 residing in close vicinity of Ser-142 (Table S2 and Table S3). In Dronpa and mTFP0.7, Ser-142 is also observed to find hydrogen-bonding partners inside the barrel or to reorient toward the solvent in the *trans* configuration of the chromophore. Thus, there are several structural solutions to the problem of stabilizing Ser-142 (Fig. S6): in comparison with EosFP, Dronpa has a valine instead of Ile-157, mTFP0.7 a histidine instead of Met-159, and IrisFP a serine instead of Phe-173.

Another factor contributing to the stability of the *trans* state is the network of interactions available around the phenolic oxygen of the chromophore's benzylidene moiety in this state. In IrisFP, the phenolic oxygen is stabilized by Glu-144 and water molecule W-2032, which both provide hydrogen bonding. Therefore, in contrast to Dronpa and mTFP0.7, the chromophore environment in the *trans* conformation appears largely polar.

Whereas bistability is an essential property of photoswitchable FPs, efficient photoswitching requires a low transition barrier between the 2 isomeric states. Free space around the *p*-hydroxybenzylidene moiety of the chromophore can clearly influence the photoswitching kinetics. The superposition of the *cis* forms of EosFP, Dronpa, and mTFP0.7 (Fig. S6A) reveals that these proteins provide rather similar steric environments around the chromophore. In contrast, in IrisFP, the mutation Phe-173-Ser creates significantly more space around the chromophore, notably by displacing Met-159 (Fig. 2). This observation might relate to the exceptional switching quantum yields measured for this protein. Interestingly, the Dronpa mutant Met-159-Thr shows substantially faster photoswitching kinetics than Dronpa (33). Presumably, even more space is freed around the chromophore in this variant, further lowering the transition barrier at the expense of a much reduced fluorescence quantum yield.

It is interesting to note that the 3 proteins compared in Fig. S6 have a different first amino acid in the chromophore-forming triad, namely cysteine in Dronpa, histidine in IrisFP, and alanine in mTFP0.7. Therefore, on/off switching by photoisomerization apparently does not depend on the nature of this amino acid. We also noticed that the carboxylate group of Glu-212 in IrisFP rotates by 90° upon isomerization and forms a hydrogen bond to the presumably unprotonated nitrogen of the chromophore's imidazolinone moiety, contrary to what is observed in Dronpa or mTFP0.7. Therefore, Glu-212 in the neutral *trans* chromophore configuration appears to be protonated. Assuming that this residue is also protonated in the neutral *cis* chromophore configuration, this observation may implicate Glu-212 as a potential donor of a proton in the putative β -elimination reaction leading to green-to-red photoconversion in EosFP-type proteins (27, 29). In the *trans* configuration of green IrisFP, the Glu-212 carboxylate side chain appears too far removed from His-62 for efficient proton transfer (Fig. 3A). Therefore, green-to-red photoconversion is unlikely to occur from the neutral *trans* state. Rather, absorption of a violet photon may lead to deprotonation of the phenol oxygen (via excited-state proton transfer), favoring back-isomerization to the *cis* state, from which photoconversion to the red form may readily occur. The excited-state behavior of photoactivatable proteins is to a large extent determined by the protonation states of the chromophore and its environment (23, 26). Therefore, more detailed investigations of these protonation states and their interplay with

isomerization will be required to decipher the photoactivation mechanisms of IrisFP.

One could be concerned that the presence of multiple phototransformations in IrisFP would result in poor yields for each of them and in a high susceptibility to photobleaching. However, reversible photoswitching in green IrisFP displays a remarkable efficiency compared with other Dronpa-like proteins. The yield of green-to-red photoconversion is also slightly increased compared with its parent protein EosFP, although red IrisFP appears more susceptible to photobleaching than EosFP. However, upon illumination with violet light, no chromophore photodamage is apparent in our crystal structures.

Our structural data suggest that the mechanisms for reversible photoswitching are essentially the same for green and red IrisFP, although the inhomogeneous chromophore conformation in the X-ray structure of photoswitched red IrisFP still leaves some possibility that isomerization may not be strictly required for switching in this state, as suggested for Dronpa (25). From the backbone breakage next to the chromophore in red IrisFP, which has also been noticed for the photoisomerizable marker protein asFP595 (21), one may expect an enhanced conformational flexibility of the chromophore and thus a facilitated light-induced isomerization. We noticed, on the contrary, that red IrisFP can be photoswitched to the nonfluorescent *trans* form slightly less efficiently than green IrisFP.

Incomplete green-to-red photoconversion may pose some limitations in one or the other marker application. Nevertheless, these adverse effects are greatly alleviated in the tetrameric form of IrisFP if we deliberately choose to only partially photoconvert the marker proteins in the sample. The strong FRET coupling between the chromophores in the tetramer leads to red fluorescence emission upon excitation of the green chromophores even if only 1 of the 4 chromophores is converted into its red fluorescent *cis* form (34). Off switching of the red form then results in the reappearance of green emission because the *trans* configuration of the neutral red chromophore does not absorb green light (Fig. S4). Thus, the partially red-converted tetramer effectively shows an interconversion between 2 bright states, red and green (Fig. S7). Such a mechanism is reminiscent of a technique used in DsRed to induce a color change (35) and also relates to the recently designed protein Phamret (36). It could be exploited further in tandem dimer constructs.

To turn IrisFP into a widely applicable highlighter protein, we aim to develop a monomeric version by introducing similar modifications as in EosFP (18), hoping that it will exhibit the same multiple photoactivation processes that we have observed here for the tetramer.

Conclusions

Photoactivatable FPs hold great promises for applications in cell biology. However, reversibly photoswitchable FPs are not visible in their off state, and photoconvertible FPs are limited by the nonreversible nature of photoconversion. These drawbacks may be alleviated with new highlighters such as IrisFP, which combine the 2 modes of photoactivation. Other applications taking advantage of the multiple phototransformations displayed by IrisFP will undoubtedly emerge, including two-color nanoscopy and sequential photoactivation schemes that can be beneficial for unraveling complex protein–protein interactions. IrisFP also hints at the possibility to combine read-only and rewritable capabilities in future mass storage media.

Materials and Methods

The experimental procedures are briefly sketched in the following; a detailed description is provided as *SI Text*. IrisFP was overexpressed and purified as EosFP (18). Rod-shaped crystals ($0.2 \times 0.2 \times 0.8 \text{ mm}^3$) were grown at 20 °C in 2.4 M ammonium sulfate, 0.1 M bicine (pH 8.4).

X-ray data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF). Crystal structures of IrisFP were solved by molecular replacement using EosFP as a starting model. To induce phototransformations, crystals were illuminated in their crystallization drops before flash-cooling. Illumination conditions are described in *SI Text*.

Spectra on solutions were taken at room temperature, by using protein concentrations of $\approx 10 \mu\text{M}$ (absorption) and $1 \mu\text{M}$ (fluorescence) in 100 mM potassium phosphate buffer (pH 9.0). *In crystallo* spectroscopy was carried out at 100 K by using the microspectrophotometer of the Cryobench laboratory (37).

Enthalpy differences between pairs of phototransformed states in IrisFP and wild-type EosFP were calculated with the *fDynamo* library (38), by using a QM/MM potential. The chromophore and the His-194 and Glu-212 side

chains were in the QM region, and the remaining atoms were in the MM region. Enthalpy differences were estimated in 2 ways, by energy minimization of single structures and by Langevin molecular dynamics at 300 K for 100 ps with a time step of 1 fs.

ACKNOWLEDGMENTS. The European Synchrotron Radiation Facility is acknowledged for support of the Cryobench laboratory and for providing in-house beam time on beamlines ID14-3, ID23-1, and ID29. This work was supported by Agence Nationale de la Recherche Grant ANR-07-BLAN-0107-01 (to D.B.), Deutsche Forschungsgemeinschaft Grants NI 291/9 and SFB 497 (to G.U.N.), the Fonds der Chemischen Industrie (G.U.N.), and the Landesstiftung Baden-Württemberg (Elite Postdoc Program) (J.W.).

- Shaner NC, Patterson GH, Davidson MW (2007) Advances in fluorescent protein technology. *J Cell Sci* 120:4247–4260.
- Lukyanov KA, Chudakov DM, Lukyanov S, Verkhusha VV (2005) Innovation: Photoactivatable fluorescent proteins. *Nat Rev Mol Cell Biol* 6:885–891.
- Wiedenmann J, Nienhaus GU (2006) Live-cell imaging with EosFP and other photoactivatable marker proteins of the GFP family. *Exp Rev Proteomics* 3:361–374.
- Hess ST, Girirajan TP, Mason MD (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J* 91:4258–4272.
- Betzig E, et al. (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313:1642–1645.
- Hofmann M, Eggeling C, Jakobs S, Hell SW (2005) Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc Natl Acad Sci USA* 102:17565–17569.
- Sauer M (2005) Reversible molecular photoswitches: A key technology for nanoscience and fluorescence imaging. *Proc Natl Acad Sci USA* 102:9433–9434.
- Ando R, Mizuno H, Miyawaki A (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* 306:1370–1373.
- Henderson JN, Ai HW, Campbell RE, Remington SJ (2007) Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc Natl Acad Sci USA* 104:6672–6677.
- Lukyanov KA, et al. (2000) Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J Biol Chem* 275:25879–25882.
- Patterson GH, Lippincott-Schwartz J (2002) A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* 297:1873–1877.
- Verkhusha VV, Sorkin A (2005) Conversion of the monomeric red fluorescent protein into a photoactivatable probe. *Chem Biol* 12:279–285.
- Chudakov DM, et al. (2004) Photoswitchable cyan fluorescent protein for protein tracking. *Nat Biotechnol* 22:1435–1439.
- Gurskaya NG, et al. (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat Biotechnol* 24:461–465.
- Mizuno H, et al. (2003) Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. *Mol Cell* 12:1051–1058.
- Oswald F, et al. (2007) Contributions of host and symbiont pigments to the coloration of reef corals. *FEBS J* 274:1102–1109.
- Tsutsui H, Karasawa S, Shimizu H, Nukina N, Miyawaki A (2005) Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO Rep* 6:233–238.
- Wiedenmann J, et al. (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc Natl Acad Sci USA* 101:15905–15910.
- Ivanchenko S, et al. (2007) Two-photon excitation and photoconversion of EosFP in dual-color 4Pi confocal microscopy. *Biophys J* 92:4451–4457.
- Andresen M, et al. (2007) Structural basis for reversible photoswitching in Dronpa. *Proc Natl Acad Sci USA* 104:13005–13009.
- Andresen M, et al. (2005) Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc Natl Acad Sci USA* 102:13070–13074.
- Habuchi S, et al. (2005) Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc Natl Acad Sci USA* 102:9511–9516.
- Habuchi S, et al. (2006) Photo-induced protonation/deprotonation in the GFP-like fluorescent protein Dronpa: Mechanism responsible for the reversible photoswitching. *Photochem Photobiol Sci* 5:567–576.
- Fron E, et al. (2007) Ultrafast excited-state dynamics of the photoswitchable protein Dronpa. *J Am Chem Soc* 129:4870–4871.
- Mizuno H, et al. (2008) Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proc Natl Acad Sci USA* 105:9227–9232.
- Schäfer LV, Groenhof G, Boggio-Pasqua M, Robb MA, Grubmüller H (2008) Chromophore protonation state controls photoswitching of the fluoroprotein asFP595. *PLoS Comput Biol* 4:e1000034.
- Nienhaus K, Nienhaus GU, Wiedenmann J, Nar H (2005) Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proc Natl Acad Sci USA* 102:9156–9159.
- Nienhaus GU, et al. (2006) Photoconvertible fluorescent protein EosFP: biophysical properties and cell biology applications. *Photochem Photobiol* 82:351–358.
- Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci USA* 99:12651–12656.
- Hayashi I, et al. (2007) Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. *J Mol Biol* 372:918–926.
- van Thor JJ, Gensch T, Hellingwerf KJ, Johnson LN (2002) Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222. *Nat Struct Biol* 9:37–41.
- Chudakov DM, et al. (2003) Kindling fluorescent proteins for precise in vivo photolabeling. *Nat Biotechnol* 21:191–194.
- Stiel AC, et al. (2007) 1.8 Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochem J* 402:35–42.
- Cotlet M, et al. (2001) Identification of different emitting species in the red fluorescent protein DsRed by means of ensemble and single-molecule spectroscopy. *Proc Natl Acad Sci USA* 98:14398–14403.
- Marchant JS, Stutzmann GE, Leissing MA, LaFerla FM, Parker I (2001) Multiphoton-evoked color change of DsRed as an optical highlighter for cellular and subcellular labeling. *Nat Biotechnol* 19:645–649.
- Matsuda T, Miyawaki A, Nagai T (2008) Direct measurement of protein dynamics inside cells using a rationally designed photoconvertible protein. *Nat Methods* 5:339–345.
- Bourgeois D, Vernède X, Adam V, Fioravanti E, Ursby T (2002) A microspectrophotometer for UV-visible and fluorescence studies of protein crystals. *J Appl Crystallogr* 35:319–326.
- Field MJ, Albe M, Bret C, Proust-de-Martin F, Thomas A (2000) The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J Comput Chem* 21:1088–1100.

Supporting Information

Adam et al. 10.1073/pnas.0805949105

SI Text

Crystallization. IrisFP was expressed in *Escherichia coli* (strain M15pREP4) and purified as described for EosFP (1). Crystals were grown at 20 °C in 2.4 M ammonium sulfate, 0.1 M bicine (pH 8.4), using the hanging-drop vapor-diffusion method. Rod-shaped crystals appeared within 24 h and reached their final size of (0.2 × 0.2 × 0.8) mm³ within a few days. For low-temperature measurements, crystals were rapidly transferred to a cryoprotectant solution [10% glycerol/2.4 M ammonium sulfate/0.1 M bicine (pH 8.4)] before being flash-frozen in liquid or gaseous nitrogen at 100 K.

X-Ray Structure Analysis. X-ray diffraction experiments were performed at the European Synchrotron Radiation Facility (ESRF). IrisFP crystals were kept at 100 K during data collection by using a nitrogen stream. Data collection of IrisFP in its green fluorescent (on) and nonfluorescent (off) states was performed on beamline ID14-3 (X-ray wavelength $\lambda = 0.931$ Å). Data of the red state were collected on beamline ID23-1 ($\lambda = 0.816$ Å). Data of the red off state were collected on beamline ID29 ($\lambda = 0.979$ Å). Datasets were integrated and scaled with XDS (2). Data collection and model refinement statistics are compiled in Table S1.

Crystal structures were solved by molecular replacement with the program Molrep (3), using the coordinates of the wild-type EosFP (PDB ID code 1ZUX) as a search model. Crystallographic refinement was performed with Refmac (4) and Coot (5). Chromophore restraints were generated in Libcheck (6) from idealized coordinates found in the Hic-Up database (7). Attempts to force chromophore planarity resulted in residual difference electron density in $F_{\text{obs}} - F_{\text{calc}}$ maps. Overall rms deviations were calculated with the program Superpose (8). Cavity volumes were determined by using the program CASTp (9). $2F_{\text{obs}} - F_{\text{calc}}$ composite omit maps were calculated with OMIT (10), using phases from the green structure of IrisFP. The putative model of the IrisFP red off state was energy-minimized with a QM/MM potential using the *fDynamo* library (11). Hydrogen atoms were added according to pK_a calculations at pH 7.0 (12). Each monomer was optimized separately. The chromophore and the side chains of Arg-66, Ser-142, Glu-144, His-194, and Glu-212 were placed in the QM region and treated with the PDDG semiempirical method (13), whereas the remaining atoms were in the MM region and treated with the OPLS force field (14). Figs. 2 and 3 and Figs. S5 and S6 are based on the structures of IrisFP monomer B. Minor deviations from these structures are observed in monomers A, C, and D that do not affect the conclusions drawn.

Solution Spectroscopy. A few microliters of a concentrated stock solution of IrisFP were added to 100 mM potassium phosphate buffer to obtain a final concentration of ≈ 10 μM (absorption) and 1 μM (fluorescence), respectively. Absorption spectra were recorded at room temperature on a Cary 1 spectrophotometer (Varian) with a resolution of 1 nm. Fluorescence excitation and emission spectra were measured at 25 °C on a SPEX Fluorolog II spectrofluorometer (Spex Industries) with the excitation line width set to 0.85 nm. Emission spectra were recorded with 2.2-nm resolution and corrected for the wavelength dependence of the detector efficiency.

Extinction coefficients of green and red IrisFP chromophores were obtained in the following way. An absorption spectrum of green IrisFP at pH 7 was measured, from which the extinction

coefficient of IrisFP at 280 nm was estimated by considering the number of aromatic amino acids contributing to the band (20). The extinction coefficient of the green anionic chromophore was calculated with reference to the absorption at 280 nm, taking into account that, at pH 7, only a fraction of the chromophores is anionic. Subsequently, the sample was partially photoconverted, and the spectrum was modeled with the contributions assigned to green and red IrisFP. From the loss of the green species and the concomitant gain of the red species due to photoconversion, we calculated the extinction coefficient of the red species, taking into account that the extinction coefficient at 551 nm solely originates from anionic red chromophores. This latter method avoids problems arising from partial deterioration of the sample due to photoconversion. In Fig. 1C, full photoconversion to red IrisFP was accompanied by such partial deterioration; hence, an estimation of the extinction coefficient of the red form based on the absorption ratio between 280 nm and 551 nm leads to an artificially low value.

Reversible photoswitching of green IrisFP was achieved by illuminating the sample solution with light of 488 nm at 85 $\mu\text{W}/\text{cm}^2$ (on/off switching) and 405 nm at 85 $\mu\text{W}/\text{cm}^2$ (off/on switching), respectively. For reversible photoswitching of red IrisFP, 532-nm (on/off switching, 2 mW/cm²) and 458-nm (off/on switching, 11 mW/cm²) light was used. Green-to-red photoconversion was induced by illumination with 405-nm light (30 mW/cm²).

To estimate the quantum yields of the different phototransformations, 20 μl of dilute protein solution (concentration ≈ 1 μM) were filled into a 1 × 10 × 30 mm quartz cuvette and illuminated with light of the respective wavelength (see above). The light source was carefully adjusted to evenly illuminate the entire sample volume. The fluorescence intensity was recorded as a function of time at the appropriate wavelength ($\lambda_{\text{green-Em}} = 520$ nm, $\lambda_{\text{red-Em}} = 620$ nm) by using a SPEX Fluorolog II spectrofluorometer (Spex Industries). The kinetics were fitted with a stretched exponential,

$$y = A_1 \cdot \exp(-(\tau/t)^\beta) + A_0.$$

Average phototransition times $\langle\tau\rangle$; were determined according to

$$\langle\tau\rangle = \frac{\tau}{\beta} \Gamma\left(\frac{1}{\beta}\right),$$

where Γ represents the gamma function. For photoswitching processes, absolute quantum yields Φ were calculated by using

$$\Phi = \frac{\frac{1}{\langle\tau\rangle} * hc * A_{\text{ill}} * N_A}{P * \lambda * \epsilon * \ln 10},$$

with Planck constant, h , speed of light, c , sample cross-section, A_{ill} , Avogadro constant, N_A , power density, P , wavelength, λ , and extinction coefficient, ϵ . To ensure that our experimental protocol was reliable, photobleaching experiments were performed on Coumarin 307, and we obtained results that were identical to published data (15). For the photoconversion process with 405-nm illumination, Coumarin 307 was taken as a reference and the quantum yield was obtained as

$$\Phi = \Phi_{\text{coumarin}} \frac{\tau_{\text{coumarin}} \epsilon_{\text{coumarin}}}{\tau \epsilon}.$$

In crystallo Spectroscopy. Absorption and fluorescence spectra in crystals were recorded at 100 K by using the microspectrophotometer of the Cryobench laboratory (16). Crystals in standard cryoloops were quickly dipped into the cryoprotectant solution [10% glycerol/2.4 M ammonium sulfate/0.1 M bicine (pH 8.4)] and placed under a nitrogen flow stream. Spectra were recorded by using a CCD-based spectrometer (HR2000+; Ocean Optics). Absorption spectra were recorded with a white light source (DH2000-BAL; Ocean Optics). Green and red fluorescence were excited by using lasers emitting at 488 nm (532-MAP-A01; Melles-Griot) and 532 nm (NG-11010-110; JDS Uniphase), respectively. Fluorescence emission spectra were collected by using short excitation light pulses (≈ 10 ms) to prevent photobleaching.

To induce the various phototransformations in the crystals, IrisFP crystals were illuminated in their crystallization drops before freezing, using optical fibers. The reversible switching of the green form was induced by exposing the crystals for ≈ 2 min to 488-nm laser light at ≈ 1 W/cm². The nonreversible photoconversion from green to red was induced by exposing the crystals for ≈ 30 min to 405-nm laser light at ≈ 2.4 W/cm². Reversible switching of the red form was induced by illuminating the crystals for ≈ 10 min with laser light at 532 nm (≈ 370 mW/cm²).

Enthalpy Calculations. Enthalpy differences between pairs of phototransformed states in IrisFP and wt-EosFP were calculated with the *fDynamo* library (11). The states of the proteins were modeled starting from the crystallographic structures; their protonation states at pH 7 were estimated by using pK_a calculations (12). The resulting structures were then solvated in a cubic box of water molecules that contained sufficient sodium ions to ensure a total system charge of zero.

The potential energy of the system was calculated by using a QM/MM potential. The chromophore and the His-194 and

Glu-212 side chains (monomer B) were in the QM region and described by either the AM1 (17) or PDDG (13) semiempirical methods. The remaining atoms were in the MM region and treated with the OPLS force field (14) and the TIP3P water model (18). Nonbonding interactions within the MM region and between the QM and MM atoms were calculated by using an atom-based force-switching truncation function with inner and outer cutoffs of 8.0 Å and 12.0 Å, respectively. Periodic boundary conditions were applied.

To focus on local modifications around the chromophore, atoms at a distance of more than ≈ 10 Å from the C β of Tyr-63 were not allowed to move in each of the simulations. The fixed atoms were identical for calculations on different states of the same protein so as to ensure a valid comparison of the potential energies. Likewise, the condition that the number of atoms in each state is the same required us to keep Glu-212 deprotonated in the structures with the *trans* conformation of the chromophore. We considered this a reasonable approximation, given that protonation of glutamate is unnecessary in the photoswitching reaction of Dronpa-like proteins.

Enthalpy differences between different states of the same protein were estimated in 2 ways (Table S2): (i) by energy minimization of single structures for each state, which is equivalent to a calculation at 0 K, and (ii) by Langevin molecular dynamics simulations at 300 K with time steps of 1 fs. Each simulation covered ≈ 100 ps. The enthalpy of each state was estimated by averaging the potential energies of 1,000 structures after the system had equilibrated. To identify differences in the intramolecular interactions between IrisFP and wild-type EosFP, energy calculations were done on structures resulting from the above simulations, using null atomic charges on specific residues. All unfixed residues of the MM region were scanned in this way, and it was found that Ser-142 has a predominant effect (Table S3).

1. Wiedenmann J, et al. (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc Natl Acad Sci USA* 101:15905–15910.
2. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* 26:795–800.
3. Vagin A, Teplyakov A (1997) MOLREP: An automated program for molecular replacement. *J Appl Crystallogr* 30:1022–1025.
4. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
5. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
6. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
7. Kleywegt GJ, Jones TA (1998) Databases in protein crystallography. *Acta Crystallogr D* 54:1119–1131.
8. Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: A simple server for sophisticated structural superposition. *Nucleic Acids Res* 32:W590–W594.
9. Dundas J, et al. (2006) CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34:W116–W118.
10. Vellieux FMD, Dijkstra BW (1997) Computation of Bhat's OMIT maps with different coefficients. *J Appl Crystallogr* 35:319–326.
11. Field MJ, et al. (2000) The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J Comput Chem* 21:1088–1100.
12. Antosiewicz J, McCammon JA, Gilson MK (1994) Prediction of pH-dependent properties of proteins. *J Mol Biol* 238:415–436.
13. Repasky MP, Chandrasekhar J, Jorgensen WL (2002) Improved semiempirical heats of formation through the use of bond and group equivalents. *J Comput Chem* 23:498–510.
14. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110:1657–1666.
15. Eggeling C, Brand L, Seidel CAM (1997) Laser-induced fluorescence of coumarin derivatives in aqueous solution: Photochemical aspects for single molecule detection. *Biolmaging* 5:105–115.
16. Bourgeois D, et al. (2002) A microspectrophotometer for UV-visible and fluorescence studies of protein crystals. *J Appl Crystallogr* 35:319–326.
17. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909.
18. Jorgensen WL, et al. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
19. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck—A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
20. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411–2423.

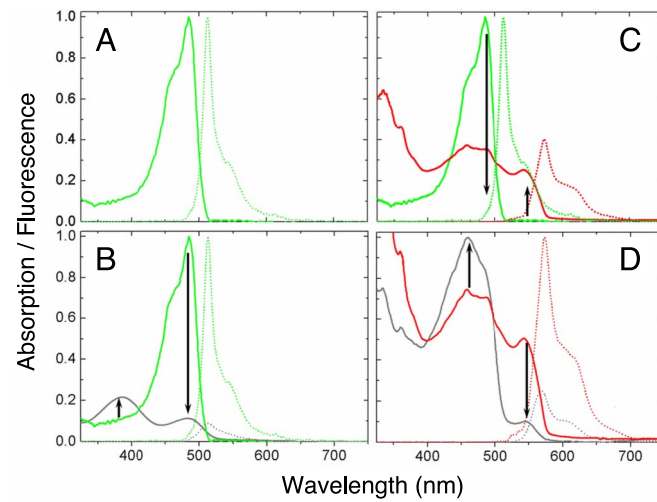


Fig. S1. Absorption and fluorescence spectra of IrisFP crystals at 100 K, depicted by solid and dotted lines, respectively. The emission spectra of the green and the red forms of IrisFP were obtained by exciting at 488 nm and 532 nm, respectively. (A) Green IrisFP. (B) Green IrisFP before (green lines) and after (gray lines) illumination with 488-nm light. (C) Red IrisFP (red lines) after photoconversion of green IrisFP (green lines) at 405 nm. (D) Red IrisFP before (red lines) and after (gray lines) illumination with 532-nm light.

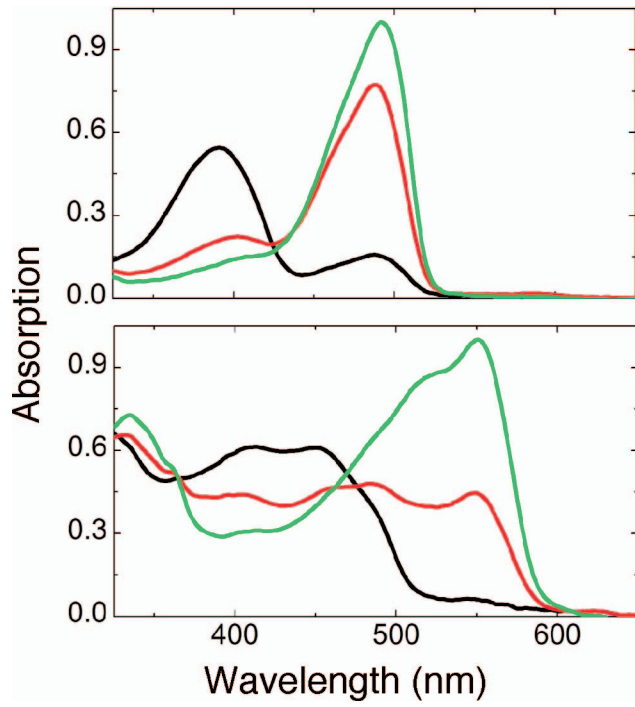


Fig. S2. Solution absorption spectra of green (*Upper*) and red (*Lower*) IrisFP at pH 5 (black), pH 7 (red), and pH 9 (green).

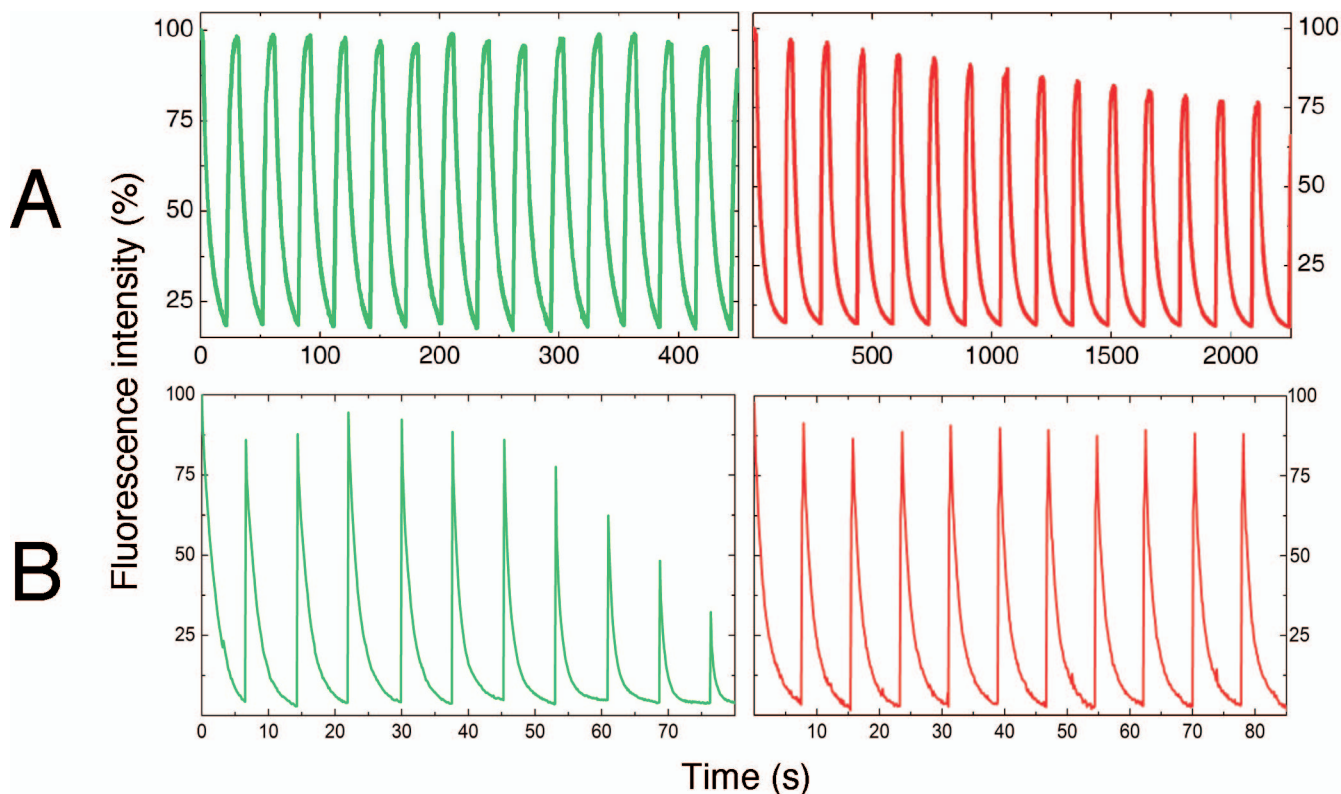


Fig. S3. Reversible photoswitching of IrisFP. (A) Solution state. (Left) Green IrisFP (0.1 M bicine, pH 9). The sample was exposed to continuous illumination with 488-nm light (50 mW/cm^2), while 405-nm light (37 mW/cm^2) was switched on every 30 s for 10 s. (Right) Red IrisFP (0.1 M potassium-phosphate, pH 7). After photoconversion, the sample was exposed to continuous illumination with 561-nm light (49 mW/cm^2), while 440-nm light (31 mW/cm^2) was switched on every 120 s for 30 s. (B) Crystalline state. (Left) Green IrisFP. A capillary mounted crystal was exposed to continuous illumination with 488-nm light (2.5 W/cm^2), while 405-nm light (1 W/cm^2) was switched on every 7.5 s for 25 ms. Crystal drying was observed at the end of the series, explaining the faster decay as compared with the solution state. (Right) Red IrisFP. A capillary mounted crystal was photoconverted and exposed to continuous illumination with 532-nm light (7.5 W/cm^2), while 440-nm light (1 W/cm^2) was switched on every 7.5 s for 200 ms.

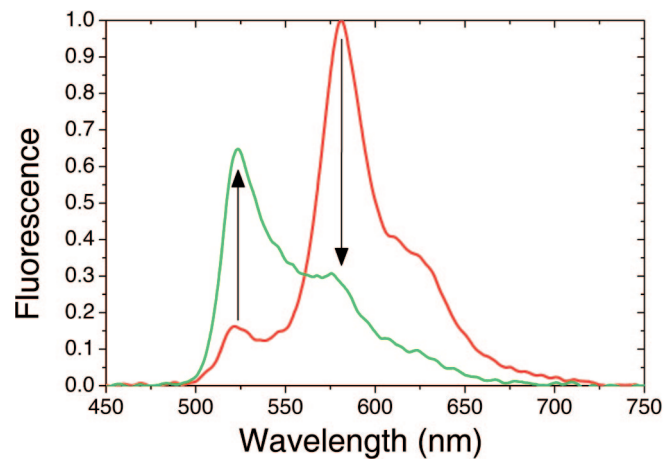


Fig. S4. Red–green photoswitching due to the intramolecular FRET effect. Partial green-to-red photoconversion of the IrisFP tetramer leads to predominantly red fluorescence, even when excited at 488 nm, because of intramolecular FRET from green to red monomers. Actinic illumination of these partially red tetramers with green light ($\lambda = 532$ nm) reversibly photoswitches the red monomers to their nonfluorescent form. As a consequence, excitation at 488 nm does not result in FRET anymore and hence generates an increase in green fluorescence.

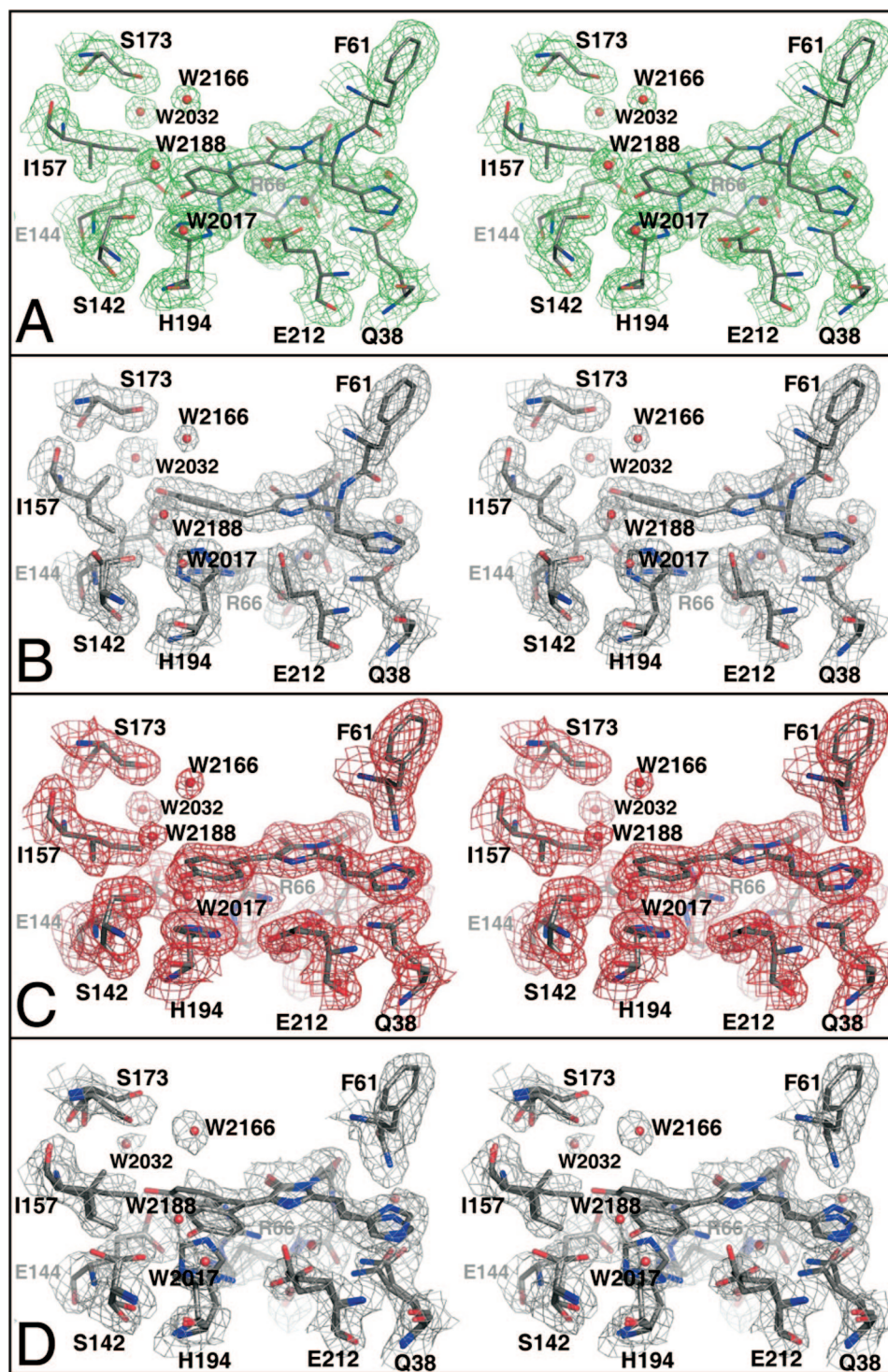
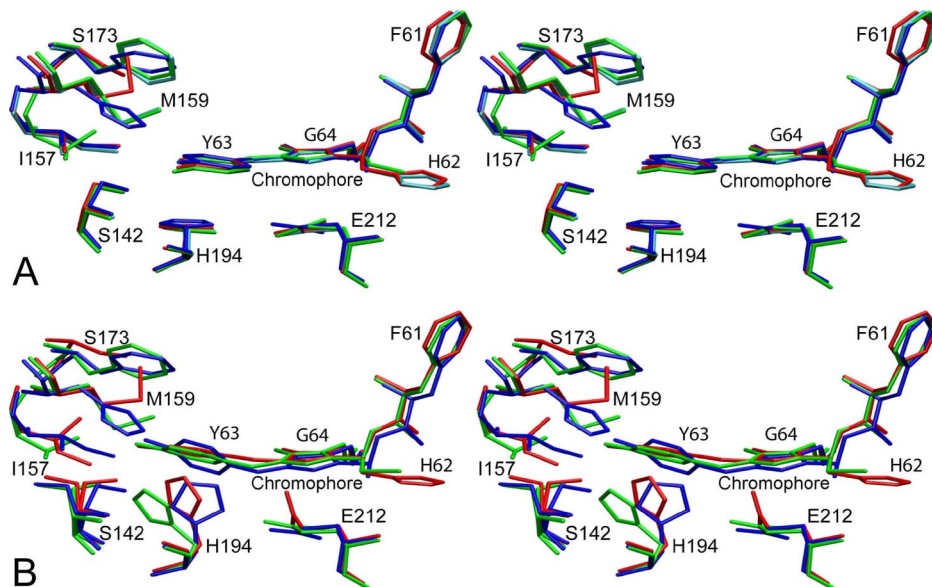


Fig. S5. Omit electron density maps at $+1.0 \sigma$ of the various states of IrisFP. Stereo views are shown. (A) Green state. (B) First reversibly switched state (green "off"). (C) Red state. (D) Second reversibly switched state (red "off"). All structures unambiguously show the conformation of the chromophore triad. After the first reversible switching (B), isomerization is essentially complete in the crystal. No increased disorder is noticed in the *trans* state. (C and D) For red IrisFP, the maps clearly indicate backbone cleavage between Phe-61 and His-62. In D, the density map suggests a mixed occupation by the *cis* and *trans* states of the chromophore.



Corresponding residues of IrisFP, EosFP, Dronpa and mTFP0.7 are listed in the table below.

IrisFP	F61	H62	Y63	G64	S142	I157	M159	S173	H194	E212
EosFP	F61	H62	Y63	G64	S142	I157	M159	F173	H194	E212
Dronpa	F61	C62	Y63	G64	S142	V157	M159	F173	H193	E211
mTFP0.7	F65	A66	Y66	G68	S146	I161	H163	F177	H197	E215

Fig. S6. Superposition of the chromophores and close environments of IrisFP, EosFP, Dronpa, and mTFP0.7 in stereoview representation. (A) Chromophores and surrounding side chains of monomer B of green IrisFP, green EosFP (PDB ID code: 1ZUX), Dronpa (PDB ID code: 2IOV) and mTFP0.7 (PDB ID code: 2OTB) are shown in the *cis* conformation in red, cyan, green, and blue colors, respectively. (B) *Trans* conformations of IrisFP, Dronpa, and mTFP0.7. Residues are labeled according to IrisFP.

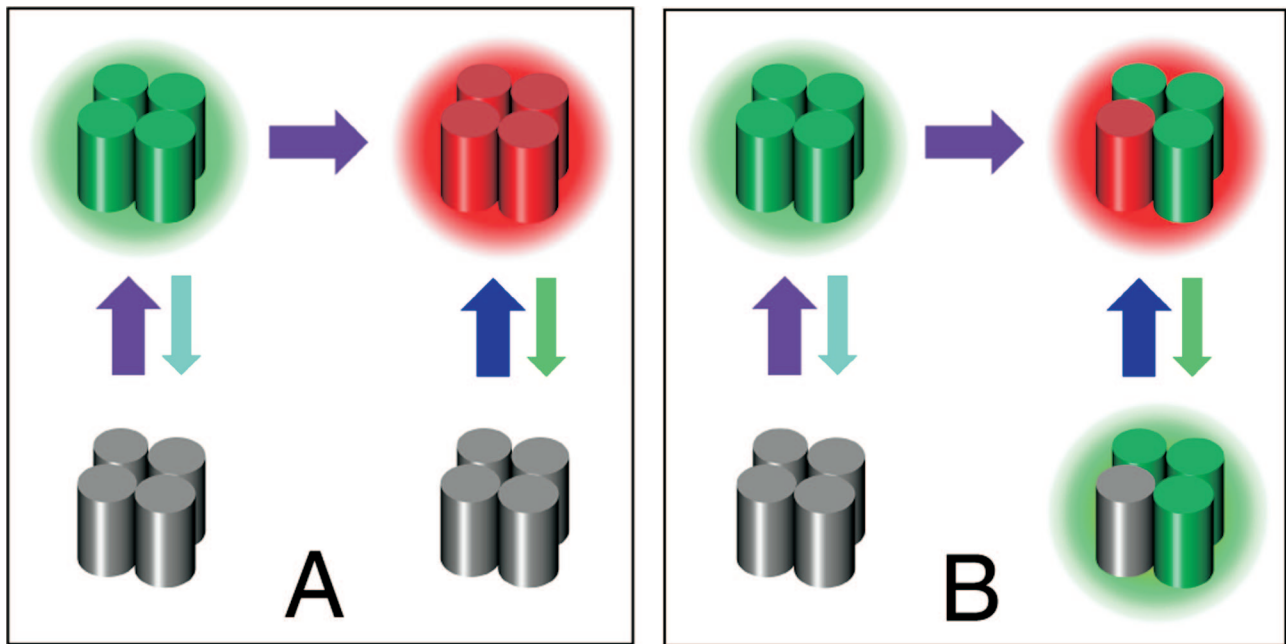


Fig. S7. Reversible red–green photoswitching in tetrameric IrisFP. (A) Complete photoconversion of the tetramer from green to red IrisFP is assumed, resulting in the “normal” photoswitching behavior between a red bright state and a dark state. (B) Because of intramolecular FRET, partial photoconversion of the tetramer can be used to generate reversible photoswitching between 2 bright states. Colored halos represent the emission colors.

Table S1. Data collection, refinement, and structure quality statistics of the different forms of IrisFP reported in this work

	Dataset			
	Green <i>cis</i>	Green <i>trans</i>	Red <i>cis</i>	Red <i>trans</i>
PDB ID code	2VVH	2VVI	2VVJ	—
Data collection				
Beamline	ESRF / ID14-3	ESRF / ID14-3	ESRF / ID23-1	ESRF / ID29
Wavelength, Å	0.931	0.931	0.816	0.979
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions				
<i>a</i> , Å	86.34	81.31	85.67	81.29
<i>b</i> , Å	96.55	96.38	96.62	96.60
<i>c</i> , Å	139.83	140.61	140.17	141.10
Resolution, Å	45.6–1.8 (1.9–1.8)	46.6–2.0 (2.2–2.0)	45.7–2.0 (2.1–2.0)	42.3–2.2 (2.5–2.2)
<i>R</i> _{sym} , %*	5.9 (42.8)	10.3 (46.8)	6.7 (43.6)	8.2 (54.7)
Mean <i>I</i> /σ(<i>I</i>)	19.06 (3.46)	10.1 (2.8)	15.37 (3.44)	12.30 (2.82)
Completeness, %	99.6 (99.6)	97.3 (98.6)	99.1 (99.0)	98.4 (99.4)
Redundancy	4.7 (4.6)	3.7 (3.7)	4.0 (3.9)	4.0 (4.0)
No. of unique reflections	108,316 (15,990)	73,301 (18,250)	78,543 (10,568)	56,167 (17,806)
Wilson B factor, Å ²	29.6	33.9	35.8	42.1
Refinement				
Observations/parameters	3.22	2.22	2.35	—
<i>R</i> _{work} / <i>R</i> _{free} [†]	0.185 / 0.218	0.193 / 0.251	0.186 / 0.232	—
No. of atoms				
Total	8,444	8,189	8,217	—
Protein	7166	7086	7159	—
Other	91	96	91	—
Water	1,187	1,007	967	—
Mean B factor, Å ²	23.1	27.4	30.4	—
Rmsd				
Bond lengths, Å	0.009	0.011	0.008	—
Bond angles, °	1.209	1.322	1.163	—
Chromophore planarity [‡]				
Dihedral angle 1, °	5.7 ± 1.9	−176.4 ± 0.8	14.8 ± 2.3	—
Dihedral angle 2, °	12.3 ± 0.7	40.3 ± 8.4	21.6 ± 2.7	—
Ramachandran plot, [§] %				
Most favored	93.7	93.3	93.5	—
Additionally allowed	6.3	6.7	6.5	—
Generously allowed	0.0	0.0	0.0	—
Disallowed	0.0	0.0	0.0	—

Values in parentheses refer to the highest-resolution shell.

$$*R_{\text{sym}} = \frac{\sum_j \sum_h |I_{h,j} - \langle I \rangle|}{\sum_j \sum_h I_{h,j}}$$

[†]*R*_{work} = $\sum h |F_{\text{obs}} - F_{\text{calc}}| / \sum h F_{\text{obs}}$. *R*_{free} is calculated with a small fraction (5%) of reflections chosen to be part of a test group.

[‡]Chromophore planarity was assessed by the measurement of the dihedral angles linking atoms N15-C9-C8-C7 (dihedral angle 1) and C9-C8-C7-C5 (dihedral angle 2) of the chromophore (averaged over the 4 molecules of the asymmetric unit).

[§]Determined by PROCHECK (19).

Table S2. Calculated enthalpy differences between the isomerized states of the chromophore in IrisFP and EosFP

Simulation method	QM/MM potential	ΔH (cG \rightarrow tG), kJ·mol ⁻¹		ΔH (cR \rightarrow tR), kJ·mol ⁻¹	
		IrisFP	EosFP	IrisFP	EosFP
Optimization	AM1/OPLS	62	107	69	110
Optimization	PDDG/OPLS	55	129	66	116
Dynamics	AM1/OPLS	68	114	62	122

cG, *cis*-green; tG, *trans*-green; cR, *cis*-red; tR, *trans*-red.

Table S3. Enthalpy differences between the isomerized states of the chromophore in IrisFP and EosFP, calculated with null atomic charges (MM) for Ser-142

Simulation method	QM/MM potential	ΔH (cG \rightarrow tG), kJ·mol ⁻¹		ΔH (cR \rightarrow tR), kJ·mol ⁻¹	
		IrisFP	EosFP	IrisFP	EosFP
Optimization	AM1/OPLS	64	67	70	70
Optimization	PDDG/OPLS	58	97	60	83
Dynamics	AM1/OPLS	51	53	45	58

cG: *cis*-green; tG, *trans*-green; cR, *cis*-red; tR, *trans*-red.

Conclusion et perspectives

Nous avons appliqué des techniques de simulation numérique à différentes protéines fluorescentes d'intérêt. L'étude de dynamique moléculaire a permis de détecter les fines interactions de van der Waals qui conditionnent l'amélioration des propriétés de fluorescence dans la Cerulean par rapport à l'ECFP (*Enhanced Cyan Fluorescent Protein*). Nous avons également élucidé le mécanisme de photoconversion de la protéine fluorescente EosFP par l'utilisation de potentiels MQ/MM appropriés au traitement des états excités. L'analyse des interactions électrostatiques révèle le pouvoir catalytique déterminant de certains résidus polaires dans l'environnement du chromophore. Finalement, la stabilité thermodynamique des différents états structuraux de la protéine IrisFP a été évaluée par des simulations MQ/MM. Là encore, les interactions électrostatiques jouent un rôle manifeste dans les propriétés photophysiques complexes de ce nouveau marqueur biologique.

Il serait désormais intéressant d'exploiter les résultats de ces études par des projets de mutagenèse dirigée. Certaines mutations devraient permettre d'améliorer le rendement quantique de fluorescence des protéines cyans. Egalement, l'insertion de nouveaux résidus polaires dans l'environnement du chromophore des protéines photoconvertibles devrait faciliter la rupture photo-induite de leur chaîne peptidique. Enfin, une étude des mécanismes réactionnels de photo-isomérisations d'IrisFP permettrait de mieux comprendre la coordination des grands changements structuraux qui accompagnent ces transformations.

Conclusion générale

La simulation numérique est un moyen approprié à l'étude des systèmes biologiques à l'échelle atomique. Nous avons présenté certaines applications de simulation numérique aux protéines fluorescentes. Ces études fournissent des informations qui sont difficiles à obtenir par des techniques expérimentales. En particulier, elles révèlent des interactions intramoléculaires déterminantes pour la fonction biologique.

Nous avons examiné de faibles interactions de van der Waals qui influencent les propriétés de certaines protéines. Cependant, les interactions électrostatiques demeurent les plus fortes et conditionnent souvent les propriétés d'un système moléculaire. Ces interactions peuvent avoir un effet catalytique déterminant sur une réaction chimique. Elles peuvent par ailleurs stabiliser un état structural donné. Il est donc nécessaire de modéliser convenablement les interactions électrostatiques dans les systèmes moléculaires.

Des travaux comme ceux que nous avons présentés pour le traitement des interactions électrostatiques devraient contribuer à améliorer l'efficacité et la précision des méthodes de simulation. Ces efforts devraient permettre à la modélisation moléculaire d'acquérir une place de plus en plus importante dans le domaine des sciences du vivant.

Annexe A

Informations supplémentaires sur les tests des méthodes de partition

TAB. A.1: Charge de l'oxygène de l'eau pour différentes méthodes à partir d'une fonction d'onde HF calculée avec différents jeux de fonctions de base.

Base	ESP	MLK	LNR			
STO-3G	-0,64	-0,38	-0,44			
3-21G	-0,89	-0,74	-0,70			
6-31G*	-0,82	-0,90	-0,83			
6-311G**	-0,80	-0,50	-0,49			
6-311++G**	-0,83	-0,51	-0,50			
cc-pVTZ	-0,75	-0,48	-0,49			
cc-pVQZ	-0,75	-0,52	-0,52			
cc-pV5Z	-0,74	-0,57	-0,57			
cc-pV6Z	-0,74	-0,40	-0,46			
Base	DHB1	DHB2	DHB3	DHB4	DHB5	DHB6
STO-3G	-0,47	-0,48	-0,49	-0,49	-0,48	-0,48
3-21G	-0,68	-0,66	-0,64	-0,64	-0,64	-0,64
6-31G*	-0,81	-0,79	-0,78	-0,81	-0,83	-0,84
6-311G**	-0,49	-0,49	-0,48	-0,46	-0,43	-0,40
6-311++G**	-0,50	-0,50	-0,49	-0,47	-0,44	-0,41
cc-pVTZ	-0,50	-0,50	-0,50	-0,48	-0,46	-0,46
cc-pVQZ	-0,52	-0,53	-0,52	-0,50	-0,47	-0,46
cc-pV5Z	-0,57	-0,57	-0,57	-0,57	-0,55	-0,54
cc-pV6Z	-0,48	-0,49	-0,50	-0,50	-0,48	-0,44
Base	DXB1	DXB2	DXB3	DXB4	DXB5	DXB6
STO-3G	-0,62	-0,65	-0,67	-0,68	-0,69	-0,70
3-21G	-0,84	-0,84	-0,81	-0,76	-0,70	-0,66
6-31G*	-0,97	-0,97	-0,95	-0,91	-0,86	-0,84
6-311G**	-0,68	-0,72	-0,75	-0,79	-0,83	-0,85
6-311++G**	-0,70	-0,73	-0,76	-0,80	-0,84	-0,87
cc-pVTZ	-0,71	-0,74	-0,75	-0,76	-0,77	-0,79
cc-pVQZ	-0,75	-0,79	-0,81	-0,84	-0,88	-0,93
cc-pV5Z	-0,78	-0,82	-0,85	-0,87	-0,90	-0,93
cc-pV6Z	-0,63	-0,67	-0,69	-0,71	-0,74	-0,77

Base	MHB1	MHB2	MHB3	MHB4	MHB5	MHB6
STO-3G	-0,46	-0,46	-0,46	-0,45	-0,44	-0,43
3-21G	-0,67	-0,64	-0,61	-0,60	-0,60	-0,60
6-31G*	-0,81	-0,79	-0,78	-0,79	-0,81	-0,82
6-311G**	-0,50	-0,49	-0,46	-0,44	-0,40	-0,38
6-311++G**	-0,52	-0,50	-0,48	-0,45	-0,42	-0,39
cc-pVTZ	-0,50	-0,49	-0,47	-0,45	-0,43	-0,42
cc-pVQZ	-0,53	-0,51	-0,49	-0,46	-0,43	-0,42
cc-pV5Z	-0,59	-0,58	-0,57	-0,56	-0,54	-0,53
cc-pV6Z	-0,53	-0,53	-0,53	-0,53	-0,50	-0,47
Base	MXB1	MXB2	MXB3	MXB4	MXB5	MXB6
STO-3G	-0,58	-0,60	-0,61	-0,62	-0,64	-0,65
3-21G	-0,81	-0,79	-0,76	-0,71	-0,66	-0,61
6-31G*	-0,96	-0,95	-0,93	-0,89	-0,85	-0,83
6-311G**	-0,68	-0,69	-0,72	-0,76	-0,80	-0,83
6-311++G**	-0,70	-0,71	-0,74	-0,78	-0,82	-0,85
cc-pVTZ	-0,69	-0,70	-0,71	-0,72	-0,73	-0,75
cc-pVQZ	-0,73	-0,74	-0,76	-0,79	-0,83	-0,88
cc-pV5Z	-0,79	-0,81	-0,83	-0,85	-0,88	-0,91
cc-pV6Z	-0,69	-0,71	-0,73	-0,75	-0,78	-0,81

TAB. A.2: Charge du carbone du formaldéhyde pour différentes méthodes à partir d'une fonction d'onde HF calculée avec différents jeux de fonctions de base.

Base	ESP	MLK	LNR			
STO-3G	0,36	0,09	0,21			
3-21G	0,49	0,14	0,28			
6-31G*	0,45	0,14	0,26			
6-311G**	0,44	0,21	0,37			
6-311++G**	0,48	0,13	0,34			
cc-pVTZ	0,42	0,21	0,38			
cc-pVQZ	0,42	0,33	0,46			
cc-pV5Z	0,42	0,41	0,54			
cc-pV6Z	0,42	0,23	0,39			
Base	DHB1	DHB2	DHB3	DHB4	DHB5	DHB6
STO-3G	0,24	0,27	0,27	0,24	0,19	0,14
3-21G	0,31	0,33	0,33	0,33	0,33	0,32
6-31G*	0,29	0,31	0,34	0,38	0,43	0,46
6-311G**	0,41	0,45	0,47	0,45	0,42	0,41
6-311++G**	0,40	0,44	0,46	0,45	0,42	0,40
cc-pVTZ	0,43	0,47	0,49	0,47	0,42	0,37
cc-pVQZ	0,50	0,53	0,54	0,52	0,48	0,43
cc-pV5Z	0,58	0,60	0,62	0,62	0,60	0,58
cc-pV6Z	0,44	0,48	0,50	0,51	0,51	0,50
Base	DXB1	DXB2	DXB3	DXB4	DXB5	DXB6
STO-3G	-0,10	-0,12	-0,11	-0,06	0,01	0,04
3-21G	0,02	-0,00	-0,01	0,00	0,02	0,02
6-31G*	-0,05	-0,09	-0,10	-0,08	-0,06	-0,04
6-311G**	0,04	0,02	0,04	0,10	0,17	0,23
6-311++G**	0,04	0,03	0,05	0,11	0,17	0,23
cc-pVTZ	0,07	0,05	0,07	0,12	0,18	0,20
cc-pVQZ	0,13	0,10	0,09	0,12	0,16	0,19
cc-pV5Z	0,19	0,16	0,13	0,13	0,14	0,15
cc-pV6Z	0,04	0,01	-0,01	-0,02	-0,01	0,01

Base	MHB1	MHB2	MHB3	MHB4	MHB5	MHB6
STO-3G	0,20	0,29	0,33	0,32	0,27	0,21
3-21G	0,39	0,46	0,51	0,52	0,52	0,51
6-31G*	0,35	0,45	0,53	0,59	0,65	0,69
6-311G**	0,44	0,55	0,61	0,61	0,58	0,55
6-311++G**	0,49	0,58	0,64	0,63	0,59	0,57
cc-pVTZ	0,54	0,65	0,72	0,72	0,68	0,62
cc-pVQZ	0,65	0,76	0,82	0,82	0,78	0,73
cc-pV5Z	0,74	0,85	0,92	0,93	0,91	0,90
cc-pV6Z	0,45	0,57	0,64	0,67	0,66	0,65
Base	MXB1	MXB2	MXB3	MXB4	MXB5	MXB6
STO-3G	-0,08	-0,04	-0,01	0,05	0,11	0,14
3-21G	0,18	0,21	0,23	0,25	0,27	0,29
6-31G*	0,10	0,13	0,15	0,18	0,22	0,24
6-311G**	0,17	0,22	0,27	0,33	0,40	0,46
6-311++G**	0,25	0,30	0,34	0,40	0,46	0,52
cc-pVTZ	0,31	0,37	0,41	0,47	0,54	0,58
cc-pVQZ	0,44	0,49	0,51	0,56	0,62	0,67
cc-pV5Z	0,52	0,57	0,58	0,61	0,63	0,65
cc-pV6Z	0,15	0,20	0,21	0,22	0,23	0,25

TAB. A.3: Liste des 200 molécules utilisées pour le calcul des moments dipolaires dont les valeurs expérimentales sont indiquées en Debye [58].

Molécule	μ_{exp} (Debye)
1,1,1,2,2,3,3-Heptafluoropropane	1.620
1,1,1-Trichloroethane	1.780
1,1,1-Trifluoroethane	2.347
1,1-Dichloroethene	1.340
1,1-Difluoroallene	2.070
1,1-Difluoroethane	2.270
1,1-Difluoroethene	1.384
1,2,5,6-Tetrahydropyridine	1.007
1,2-Dichlorobenzene	2.500
1,2-Dichloroethene	1.900
1,2-Difluoroethene	2.420
1,2-Ethanediol	2.280
1,2-Propanediol	2.320
1,3-Dichlorobenzene	1.720
1,3-Difluorobenzene	1.580
1,3-Dioxane	2.060
1-Bromopropane	2.180
1-Chloropropene	1.970
1-Cyanopropene	4.080
1-Fluoropropene	1.460
2,5H-Furanone	4.905
2-Aminopropane	1.190
2-Bromopropane	2.210
2-Butanone	2.780
2-Butenal	3.670
2-Chloropropene	1.647
2-Cyanopyridine	5.780
2-Fluoro-2-Methylpropane	1.959
2-Fluoropropane	1.960
2-Fluoropropene	1.610
2-Methoxyethanoic Acid	4.720
2-methylpyrimidine	1.676
2-Propanol	1.224
3,3,3-Trifluoropropene	2.433
3,3,3-Trifluoropropyne	2.360
3,4-Dihydro-2,4-pyran	1.283
3-Bromopropyne	1.540
3-Chloropropene	1.940
3-Chloropropyne	1.680
3-Cyanopyridine	3.660
3-Cyclopentenone	2.790

Molécule	μ_{exp} (Debye)
3-Methylene-Oxetane	1.630
3-Oxetanone	0.887
4,5-Dihydroisoxazole	1.770
4-Cyanopyridine	1.960
4-Cyclopentene-1,3-dione	1.680
Acetaldehyde	2.750
Acetaldimine	2.560
Acetamide	3.760
AceticAcid	1.700
Acetone	2.880
Acetonitrile	3.925
AcetylChloride	2.720
Acetylcyanide	3.450
AcetylFluoride	2.960
AcrylicAcid	2.020
Acrylonitrile	3.560
Aminoacetonitrile	2.640
Ammonia	1.470
Aniline	1.530
Anisole	1.380
Aziridine	1.890
Benzonitrile	4.180
Benzylsilane	0.845
Beta-Propiolactone	4.180
BisTrifluoromethylEther	0.540
Bromobenzene	1.700
Bromoethene	1.420
Bromoethyne	0.230
CH ₂ CN _m	1.330
CH ₃ CNH _p	1.390
CH ₃ COCH _{2m}	2.910
CH ₃ COHCH _{3p}	1.550
CH ₃ COO _m	3.260
CH ₃ NH ₂ CH _{3p}	1.500
CH ₃ NH _{3p}	2.190
CH ₃ NH _m	2.320
CH ₃ OH _{2p}	1.900
CH ₃ OHCH _{3p}	1.200
CH ₃ O _m	1.630
CH ₃ SH _{2p}	1.300
Chlorobenzene	1.690
Chloroethene	1.452
Chloroethyne	0.440
CN _m	0.770

Molécule	μ_{exp} (Debye)
Cyanoacetylene	3.724
Cyanoallene	4.280
Cyanoethane	4.010
CyanogenChloride	2.800
CyanogenFluoride	2.170
Cyclobutane-1,2-dione	3.831
CyclobutaneCarbonitrile	4.110
Cyclobutanol	1.620
Cyclobutanone	2.890
Cyclopentadienecarbonitrile	4.250
Cyclopentadienone	3.132
Cyclopentanone	3.300
CyclopropaneCarbonitrile	4.131
Cyclopropanol	1.460
Cyclopropanone	2.670
CyclopropylAmine	1.190
Dibromomethane	1.430
Dichloromethane	1.600
DicyanogenSulfide	3.020
Dicyanomethane	3.730
Difluoromethane	1.978
Dimethylamine	1.010
DimethylEther	1.300
DimethylPhosphine	1.230
DimethylSilane	0.713
Dimethylthioether	1.500
Ethanethiol	1.580
Ethanol	1.441
Ethylamine	1.220
Ethylbromide	2.030
EthylChloride	2.050
EthylFormate	1.980
EthylSilane	0.810
Fluoroallene	1.970
Fluorobenzene	1.600
Fluoroethene	1.427
Formaldehyde	2.332
Formamide	3.730
FormicAcid	1.410
FormylformicAcid	1.860
Furan	0.661
Gamma-Butyrolactone	4.270
HCNHp	0.190
HOOm	2.650
HydrogenCyanide	2.985

Molécule	μ_{exp} (Debye)
HydrogenSulfide	0.970
Indole	2.100
Isobutyronitrile	4.290
Isoxazole	2.950
Ketene	1.420
Me3NHp	0.870
MePOOMe2	2.029
meta-Fluorotoluene	1.860
Methacrylonitrile	3.690
Methanethiol	1.520
Methanol	1.700
MethylAcetate	1.720
Methylamine	1.310
Methylbromide	1.821
MethylChloride	1.892
MethylFluoride	1.858
MethylFormate	1.770
Methylketene	1.790
MethylPhosphine	1.100
MethylSilane	0.735
NH2m	1.580
N-Methylacetaldimine	1.498
N-MethylAcetamide	3.730
N-Methylformaldimine	1.530
N-Methylformamide	3.860
NO2m	0.160
OCNm	1.490
OH3p	1.530
OHm	1.350
OPF3	1.770
OPH2OH	4.663
OPH3	3.710
OPH	2.345
OPMeOH2	1.061
ortho-Chlorotoluene	1.560
ortho-Fluorotoluene	1.370
Oxetane	1.940
para-Chlorotoluene	2.210
para-Fluorotoluene	2.000
Pentachloroethane	0.920
Pentafluoroethane	1.540
Phenol	1.450
Phosphine	0.574
Piperidine	1.189
Propanal	2.520

Molécule	μ_{exp} (Debye)
PropanoicAcid	1.460
Propanol	1.580
Propenal	3.120
Pyridine	2.150
Pyrrole	1.740
SH3p	1.580
SHm	0.650
t-ButylCyanide	3.950
Tetrafluoropropyne	1.710
Tetrahydrofuran	1.630
Tetrahydropyran	1.530
Thietane	1.850
Thioacetaldehyde	2.330
Thioformaldehyde	1.647
Thiophene	0.550
Tribromomethane	0.990
Trichloromethane	1.040
Trifluoroacetonitrile	1.262
Trifluoroethene	1.320
Trifluoromethane	1.652
Trimethylamine	0.612
TrimethylSilane	0.520
VinylmethylEther	0.960
VinylSilane	0.660
Water	1.854

TAB. A.4: Erreur absolue moyenne dans le calcul des 200 dipôles (en Debye) par rapport aux valeurs expérimentales pour les différentes méthodes à partir de différentes fonctions d'onde.

Méthode	DEN	ESP	LDW	MLK	LNR	
MNDO	0,44	-	0,76	0,68	0,63	
PM3	0,40	-	0,74	0,53	0,56	
AM1	0,39	-	0,62	0,52	0,59	
PDDG	0,36	-	0,65	0,56	0,61	
RM1	0,35	-	0,64	0,60	0,62	
HF/3-21G	0,43	0,44	-	1,38	1,07	
HF/6-31G*	0,35	0,37	-	0,98	0,84	
HF/6-311++G**	0,44	0,46	-	1,87	0,73	
DFT/3-21G	0,29	0,30	-	0,99	0,78	
DFT/6-31G*	0,22	0,23	-	0,60	0,52	
DFT/6-311++G**	0,25	0,27	-	1,53	0,60	
Méthode	DHB1	DHB2	DHB3	DHB4	DHB5	DHB6
MNDO	0,62	0,63	0,64	0,65	0,67	0,71
PM3	0,61	0,66	0,71	0,73	0,73	0,72
AM1	0,63	0,66	0,69	0,71	0,73	0,75
PDDG	0,71	0,80	0,87	0,91	0,92	0,92
RM1	0,65	0,70	0,73	0,75	0,76	0,77
HF/3-21G	0,92	0,79	0,72	0,71	0,74	0,78
HF/6-31G*	0,80	0,77	0,75	0,76	0,78	0,79
HF/6-311++G**	0,57	0,51	0,52	0,63	0,78	0,93
DFT/3-21G	0,68	0,58	0,53	0,51	0,52	0,53
DFT/6-31G*	0,50	0,49	0,48	0,48	0,49	0,49
DFT/6-311++G**	0,49	0,44	0,45	0,54	0,67	0,77
Méthode	DXB1	DXB2	DXB3	DXB4	DXB5	DXB6
MNDO	0,75	0,82	0,87	0,88	0,85	0,82
PM3	0,62	0,71	0,77	0,81	0,84	0,87
AM1	0,73	0,82	0,89	0,92	0,93	0,93
PDDG	0,86	1,00	1,10	1,16	1,20	1,24
RM1	0,91	1,01	1,08	1,14	1,18	1,20
HF/3-21G	1,08	1,00	0,97	1,02	1,10	1,16
HF/6-31G*	1,12	1,14	1,18	1,26	1,35	1,40
HF/6-311++G**	0,79	0,86	0,97	1,08	1,21	1,26
DFT/3-21G	0,77	0,70	0,70	0,75	0,82	0,87
DFT/6-31G*	0,76	0,81	0,86	0,96	1,05	1,11
DFT/6-311++G**	0,65	0,70	0,76	0,86	0,96	0,99

Méthode	MHB1	MHB2	MHB3	MHB4	MHB5	MHB6
MNDO	0,66	0,64	0,64	0,65	0,67	0,71
PM3	0,61	0,65	0,68	0,70	0,70	0,69
AM1	0,68	0,68	0,69	0,71	0,73	0,76
PDDG	0,72	0,77	0,83	0,87	0,89	0,89
RM1	0,67	0,66	0,70	0,72	0,72	0,73
HF/3-21G	0,95	0,96	0,98	1,04	1,11	1,17
HF/6-31G*	0,71	0,76	0,83	0,88	0,92	0,93
HF/6-311++G**	2,87	2,72	2,84	2,98	3,11	3,28
DFT/3-21G	0,67	0,70	0,73	0,77	0,83	0,87
DFT/6-31G*	0,45	0,48	0,53	0,56	0,60	0,64
DFT/6-311++G**	2,46	2,40	2,46	2,54	2,63	2,74
Méthode	MXB1	MXB2	MXB3	MXB4	MXB5	MXB6
MNDO	0,85	0,87	0,90	0,91	0,88	0,84
PM3	0,60	0,66	0,71	0,75	0,78	0,81
AM1	0,84	0,89	0,94	0,98	1,00	1,01
PDDG	0,93	1,02	1,10	1,16	1,21	1,24
RM1	0,98	1,03	1,08	1,14	1,17	1,18
HF/3-21G	1,27	1,26	1,33	1,48	1,62	1,70
HF/6-31G*	1,31	1,34	1,41	1,57	1,75	1,87
HF/6-311++G**	2,62	2,76	3,01	3,18	3,32	3,48
DFT/3-21G	0,85	0,85	0,92	1,05	1,19	1,26
DFT/6-31G*	0,87	0,92	1,01	1,18	1,37	1,50
DFT/6-311++G**	2,31	2,41	2,58	2,69	2,79	2,89

TAB. A.5: Liste des 62 dimères pour le test des potentiels hybrides MQ/MM. Les structures initiales ont été optimisées au niveau HF/6-31G*. Les énergies de formation de ces dimères calculées à ce niveau HF/6-31G* sont reportées ici en kJ/mol. Elles ont été utilisées comme énergies de référence pour nos tests. L'ensemble de ces données a été obtenue à l'adresse http://www.ccl.net/cca/data/ff_evaluation_suite.

Dimère	$E_{formation}$ (kJ/mol)
HOH...OH2	-23.532
HOH...HOH, cyclique	-18.262
HOH...OHCH3	-23.246
CH3OH...OH2	-23.387
CH3OH...OHCH3	-23.170
C6H5OH...OH2	-30.818
HOH...OHC6H5	-19.709
T-NMA...OH2	-22.707
HOH...T-NMA	-30.526
FORMAMIDE, cyclique	-56.232
FORMAMIDE	-30.886
H2NH...HNH2	-13.350
H2NH...NH3	-12.857
HOH...NH3	-27.444
HOH...NH2CH3	-27.305
H3CCOOH...HOOCCH3, cyclique	-65.057
H2CCOOH...OHH	-44.588
IMIDAZOLE...OH2	-26.633
HOH...IMIDAZOLE	-29.552
INDOLE...OH2	-24.073
PYRROLE...OH2	-22.416
H3CNH3+... OH2	-80.744
OHH...(-)O2CCH3, bidentate	-91.412
HOH...PYRIDINE	-25.231
HOH...FURANE	-15.230
HOH...THIOPHENE	-10.126
HOH...PYRIDINE N-OXIDE	-40.663
HOH...METHYLE FORMATE (=O)	-24.299
HOH...METHYLE FORMATE (-O-)	-14.532
HOH...FORMALDEHYDE	-22.109
HOH...ACETONE	-26.147
HOH...FCH3	-18.512
FORMAMIDINE...WATER, cyclique	-41.963
HOH...FORMALDEHYDE IMINE	-25.651
GUANIDINE...OHH	-30.264
GUANIDINIUM...OHH	-77.318
ANILINE...OHH	-17.593
HOH...ANILINE	-21.494

Dimère	$E_{formation}$ (kJ/mol)
HOH...O(CH3)2	-22.222
HOH...S(CH3)2	-13.539
OHH...CH3SSCH3, cyclique	-14.387
HOH...SHC6H5	-8.3595
T-NMA, parallele	-29.282
T-NMA, antiparallele	-20.568
HOH...N-METHYLEFORMAMIDE	-30.043
N-METHYLEFORMAMIDE...OH2	-23.023
CH3COOH...NH3, bidentate	-45.397
FORMAMIDINIUM...OH2	-71.027
FORMAMIDINIUM...OH2 , central	-81.960
VINYLAMINE...OH2	-15.889
T-VINYLE ALCOOL...OH2	-31.104
HOH...T-VINYLE ALCOOL	-19.628
OXYDE DE METHYLETHYLAMINE...OHH, cyclique	-58.892
METHYLETHYLHYDROXYLAMINE...OH2 cyclique	-30.910
T,T-ACIDE OXYLIQUE...OHH cyclique	-42.714
CIS,CIS-ACIDE OXYLIQUE...OHH cyclique	-18.641
OHH...T,T-ACIDE OXYLIQUE cyclique	-14.516
HOH...METHYLE VINYLE ETHER	-19.796
HOH...CHLOROETHANE	-11.735
H2NH...O(CH3)2	-11.869
T-N-OH,N-METHYLACETAMIDE...OH2	-31.230
HOH...T-N-OH,N-METHYLACETAMIDE	-15.582

TAB. A.6: Erreur absolue moyenne dans le calcul des énergies de formation de dimères (en kJ/mol) par rapport aux valeurs HF/6-31G* pour les différents potentiels (107 énergies en MQ/MM; 62 énergies en MQ). L'erreur absolue moyenne dans le calcul des 45 énergies de formation de dimères était de 4,16 kJ/mol pour OPLS.

Potentiel	DEN	LDW	MLK	LNR	Potentiel	DEN
MNDO/OPLS	9,14	12,12	9,68	8,10	MNDO	24,06
PM3/OPLS	8,14	11,66	10,28	7,41	PM3	14,63
AM1/OPLS	7,75	9,63	8,53	7,18	AM1	9,35
PDDG/OPLS	6,82	10,94	10,41	8,55	PDDG	13,88
RM1/OPLS	6,25	9,98	8,57	7,26	RM1	15,68
Potentiel	DHB1	DHB2	DHB3	DHB4	DHB5	DHB6
MNDO/OPLS	7,73	7,44	7,25	7,21	7,26	7,49
PM3/OPLS	7,01	6,89	7,05	7,08	7,01	6,89
AM1/OPLS	7,26	7,26	7,06	7,01	6,98	7,17
PDDG/OPLS	8,24	8,37	8,84	8,90	9,01	9,04
RM1/OPLS	7,16	7,18	7,53	7,15	6,93	6,40
Potentiel	DXB1	DXB2	DXB3	DXB4	DXB5	DXB6
MNDO/OPLS	6,72	6,95	7,16	7,39	7,25	7,10
PM3/OPLS	7,34	8,08	8,59	8,94	9,11	9,00
AM1/OPLS	7,93	8,99	9,71	10,07	9,97	9,81
PDDG/OPLS	9,26	10,78	11,69	12,17	12,07	12,46
RM1/OPLS	8,63	9,83	10,70	11,19	11,25	11,08
Potentiel	MHB1	MHB2	MHB3	MHB4	MHB5	MHB6
MNDO/OPLS	7,80	7,59	7,55	7,56	7,75	7,95
PM3/OPLS	7,59	7,26	7,31	7,34	7,22	7,08
AM1/OPLS	7,02	7,12	7,20	7,29	7,33	7,48
PDDG/OPLS	8,93	8,52	8,56	8,77	8,89	8,55
RM1/OPLS	7,36	7,45	7,49	7,46	7,13	6,73
Potentiel	MXB1	MXB2	MXB3	MXB4	MXB5	MXB6
MNDO/OPLS	7,60	7,38	7,71	7,76	7,67	7,57
PM3/OPLS	8,16	8,46	8,68	8,86	9,01	9,01
AM1/OPLS	9,19	8,78	9,04	9,26	9,28	9,19
PDDG/OPLS	10,05	9,86	10,54	11,17	11,32	10,96
RM1/OPLS	9,59	9,25	9,48	9,77	9,75	9,50

Annexe B

Base structurale du photoblanchiment d'une protéine fluorescente verte photactivable

Dans le travail reporté ci-dessous, nos collaborateurs ont exploré des voies de photoblanchiment réversible et irréversible de la protéine fluorescente IrisFP. Ces résultats apportent des données structurales nouvelles et des hypothèses mécanistiques importantes pour la compréhension de ces phénomènes. Ce travail pourrait être un point de départ avisé pour une prochaine étude de modélisation moléculaire.

Structural Basis of Photobleaching in a Photoactivatable Green Fluorescent Protein

Virgile Adam, Philippe Carpentier, Sebastien Violot, Mickaël Lelimosin, Claudine Darnault, G. Ulrich Nienhaus, and Dominique Bourgeois*

Fluorescent proteins (FPs) are invaluable fluorescent markers in cell biology.^[1] They enclose a highly fluorescent chromophore that forms autocatalytically within a rigid β -can structure. Naturally occurring FPs have been optimized to emit fluorescence over a wide spectral range, and photoactivatable FPs have recently been developed that enable protein tracking^[2] and play a crucial role in super-resolution microscopy.^[3]

Upon illumination, all fluorescent proteins eventually undergo non-reversible bleaching. Typically, the 4-(*p*-hydroxybenzylidene)-5-imidazolinone chromophore emits $10^4 - 10^5$ photons before permanent photodestruction. Although photobleaching can be exploited in certain cases,^[4, 5] it is generally considered as a major nuisance, notably in single-molecule or time-resolved studies, and it limits the achievable resolution in various nanoscopy applications.^[3]

Photobleaching results from the reactive nature of electronically

excited states of the chromophore. Residues interacting with the chromophore and solvent accessibility to the chromophore pocket are known to affect the photostability of FPs.^[6] However, contrary to organic dyes, the details of the underlying molecular mechanisms have remained largely unexplored. Thus, the development of FPs with superior photostability has mainly followed empirical approaches to date.^[6, 7]

Although high-level singlet excited states can lead to photobleaching,^[8] it is generally recognized that the triplet state T_1 , due to its long lifetime, is the most likely starting point for reactions leading to chromophore destruction.^[9] In organic dyes, radical formation from T_1 has been pointed out as an important pathway for photobleaching.^[10, 11] Elegant methods have been proposed to scavenge such radical states so as to minimize fluorescence blinking or bleaching.^[12-14] Radical chemistry can also occur in fluorescent proteins. For example, in some FPs, decarboxylation of the conserved Glu222 (GFP numbering) may occur *via* a bi-radical intermediate (Kolbe mechanism), leading to fluorescence activation^[15-17] or bleaching.^[18] The structural distortions of the chromophore induced by radical formation, however, have not yet been elucidated.

Here we have used X-rays as a tool to generate a transient radical state in the fluorescent protein IrisFP,^[19] a variant of the photoactivatable fluorescent protein EosFP from *Lobophyllia hemprichii*.^[20] By combining crystallography with *in crystallo* UV-vis absorption, fluorescence and Raman spectroscopy,^[21] we have characterized the associated structural modifications of the chromophore and its environment. The results are discussed in terms of the photobleaching mechanisms induced by visible light.

In the course of diffraction experiments carried out on crystals of IrisFP, we noticed that the fluorescence emission decays very rapidly under a synchrotron X-ray beam, even at 100 K (Fig. 1a). This decay occurs much faster than the time needed to collect a complete diffraction data set. The loss of fluorescence is mostly reversible; only a small, irreversible component is observed (Fig. 1a, inset). The corresponding absorbance decay is slower, composed of a rapid phase followed by a slower phase, and is much less pronounced at comparable doses (Fig. 1a). The absorbance loss is also partly reversible (Fig. S1).

To obtain a structural view of the observed transient spectroscopic changes, we collected composite X-ray data sets,^[22] *i.e.*, subsets of diffraction data using different parts of an IrisFP crystal, so as to reconstruct two complete data sets acquired at low and moderate X-ray doses, respectively (Fig. 1, Supporting Methods and Table S1). The first (median dose: ~ 0.1 MGy) and second (~ 0.5 MGy) data sets correspond to the beginning and end of the fast absorbance decay phase, respectively (Fig. 1a). At 0.5 MGy, the loss of fluorescence is almost entirely reversible (Fig. 1a, inset). The difference electron density map between the two data sets reveals structural changes induced by X-irradiation, localized on the chromophore and its immediate environment (Fig. 1b). We estimate that < 20 % of the molecules in the crystal caused these changes.

[*] Dr. P. Carpentier [+,§], M. Lelimosin [‡], C. Darnault [§], Dr. D. Bourgeois [§]
Laboratoire de Cristallogénèse et Cristallographie des Protéines [§] and Laboratoire de Dynamique Moléculaire [‡] IBS, Institut de Biologie Structurale Jean-Pierre Ebel, CEA; CNRS; Université Joseph Fourier
41 rue Jules Horowitz, F-38027 Grenoble (France)
Fax: (+33) 4 38 78 51 22
E-mail: dominique.bourgeois@ibs.fr

V. Adam [+,], Dr. D. Bourgeois [§]
ESRF
6 rue Jules Horowitz, BP 220, 38043 Grenoble Cedex (France)

Dr. S. Violot
Laboratoire de Physiologie Cellulaire Végétale
Institut de Recherches en Technologie et Sciences pour le Vivant, CEA ; CNRS ; INRA ; Université Joseph Fourier
17 rue des Martyrs, F-38054 Grenoble (France)

Prof. G.U. Nienhaus
Institute of Biophysics
University of Ulm
Albert-Einstein-Allee 11, 89081 Ulm (Germany)
and
Department of Physics
University of Illinois at Urbana-Champaign
1110 West Green Street, Urbana, Illinois 61801 (USA)

[+] These authors contributed equally to this work

[**] We thank Martin Weik, Antoine Royant and Chantal Houée-Lévin for insightful discussions. D.B. acknowledges support by the Agence Nationale de la Recherche (ANR-07-BLAN-0107-01). G.U.N. acknowledges support by the Deutsche Forschungsgemeinschaft (Grants NI 291/9 and SFB 497).



Supporting information for this article is available on the WWW under <http://www.angewandte.org>

Pairs of negative and positive difference density features show a downward motion of the chromophore's bridging methylene moiety and an upward motion of the benzylidene phenolate oxygen, resulting in a loss of planarity of the chromophore. The motion of the methylene bridge induces a displacement of Glu212 (corresponding to Glu222 in GFP), which seems to explore two conformations, with the carboxylate group H-bonding either to His194 or to Gln38. Glu212, however, does not decarboxylate noticeably. The motion of the chromophore phenolate oxygen is accompanied by a shift of the Met159 side chain, clearly visible due to the heavy sulphur atom. Other structural changes typically observed upon X-ray exposure of protein crystals are not visible at the doses employed here (Fig. S2).

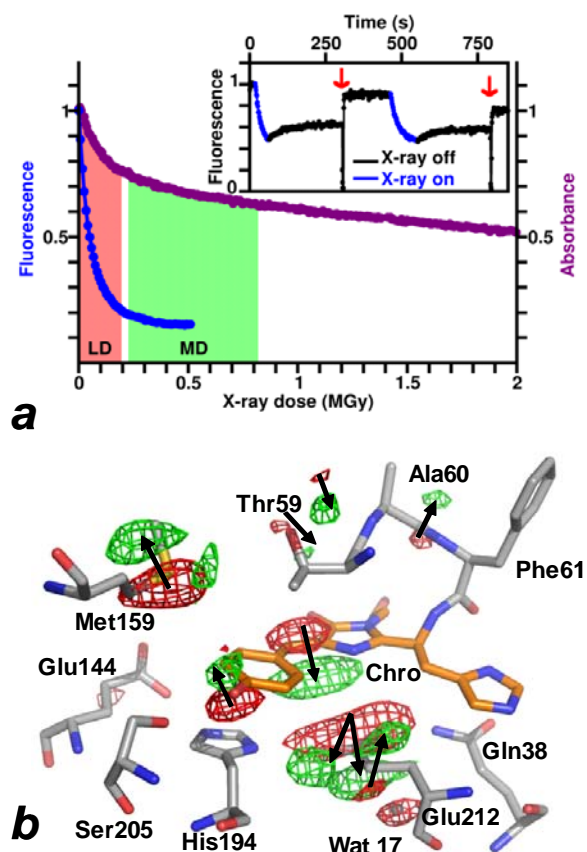


Figure 1. Reversible bleaching of IrisFP *in crystallo* induced by X-rays at 100 K. (a) Fluorescence (blue) and absorbance (purple) loss upon X-irradiation. The doses used for low- and moderate-dose (LD and MD) data collection are indicated (red and green shading). (inset) Reversibility of the fluorescence decay. Partial recovery (in black) occurs at 100 K, whereas almost complete recovery is rapidly induced by short excursions to room temperature (red arrows at 300 s and 780 s). The process can be repeated several times. (b) Experimental electron difference density map (red, -9.5σ ; green: $+9.5 \sigma$) between the LD and MD data sets, overlaid on a model of IrisFP (PDB code 2VVH). The difference map averaged over the four IrisFP monomers in the crystal asymmetric unit is shown (Supporting Methods). The chromophore is displayed in orange, while neighbouring residues are shown in grey. Arg66, in contact with the chromophore, is omitted for clarity. Arrows indicate structural motions.

In crystallo Raman spectra of IrisFP were collected with 785-nm excitation before and immediately after exposure to a moderate X-ray dose (0.9 MGy). Although the spectral changes include a component from the slow absorbance decay phase, they reveal band

modifications that are consistent with the observed structural changes (Supporting Methods, Fig. 2). Notably, the 1545 cm^{-1} mode, which has been assigned predominantly to a stretching mode of the $C_{\alpha}-C_5$ exocyclic double bond of the anionic chromophore^[23], is significantly reduced. The absence of other significant changes in the Raman spectra suggests that the phenol moiety remains chemically intact and that the chromophore does neither isomerise nor change its protonation state.^[24] Altogether, the diffraction, UV-vis absorbance and Raman data strongly suggest that the $C_{\alpha}-C_5$ π -bond is removed upon moderate X-ray exposure, implying that the conjugated π -electron system is reversibly disrupted.

Electron difference density maps obtained from data sets collected at moderate and much higher (~ 20 MGy) X-ray doses, respectively, that is, mostly representative of structural changes associated with the slow absorbance decay phase, show a strongly tilted chromophore and a clear decarboxylation of Glu212 (Fig. S3). These structural data are indicative of a permanently bleached state.

To relate these observations with the mechanism of photobleaching induced by visible light, an IrisFP crystal was illuminated at 100 K with 488-nm laser light (Supporting Methods). An irreversible loss of fluorescence and absorbance was achieved, although at short illumination times a transient rise of the 390 nm absorbance band was also noticed. We attribute this rise to a photon-induced protonation of the chromophore possibly involved in the photoswitching mechanism of IrisFP, as previously suggested for Dronpa^[25] (Figs. S4 and S5 and Supporting Information). Due to this side-reaction, and because the high optical density of IrisFP crystals at 488 nm prevents an even distribution of light throughout the crystal volume, structural changes associated with early photobleaching by visible light are difficult to capture in X-ray diffraction experiments. However, Raman spectra collected on a crystal after illumination revealed a decrease of the band at 1545 cm^{-1} , as in the X-ray case (Fig. 2). Other significant modifications in the Raman spectrum also appear (e. g., at 1255 cm^{-1} , 1415 cm^{-1} and 1645 cm^{-1}), which might be associated with protonation of the hydroxybenzylidene group. These results suggest an irreversible rupture of the $C_{\alpha}-C_5$ π -bond induced by visible light, accompanied by further modifications of the chromophore and/or its immediate environment.

To interpret our results, we first address the apparent discrepancy between the absorption and fluorescence decays in Fig. 1a. Most likely, the fluorescence decay kinetics is accelerated by quenching effects attributable to radical species generated by X-rays. Chemically intact chromophores may thus lose their ability to fluoresce, and the absorbance decay is judged more relevant to quantitatively evaluate the results (Supporting Information). This decay suggests that, upon X-ray exposure, an intermediate state rapidly builds up, reaching a steady state level of $\sim 20\%$. The associated structural changes are those seen in Fig. 1b. This intermediate (or the starting green emitting species) infrequently converts to a permanently bleached state, which corresponds to the slow phase in the absorbance decay and yields the structural changes of Fig. S3. Importantly, the fast build-up of the intermediate, its steady state level at $\sim 20\%$, together with its slow recovery at 100 K in the absence of X-rays, imply that the intermediate is not only induced, but also efficiently back-converted ('repaired') to the initial fluorescent state by X-irradiation. Such repair mechanism may constitute a general feature of X-ray induced radiation damage chemistry in biological molecules.

We propose that X-ray bleaching of IrisFP at 100 K results predominantly from electrons released by water radiolysis. These electrons (and electron holes) migrate into the chromophore pocket,

where they react with the chromophore to form a transient non-fluorescent radical, eventually leading to permanent photodestruction.

Absorption of a visible photon in FPs may induce radical states arising from intramolecular electron transfer, likely in the triplet state. Therefore, radical formation by X-rays and visible light proceeds by different mechanisms. Upon absorption of a visible photon, charge transfer from Glu212 to the chromophore may occur in IrisFP (as in GFP^[15]), leaving an unpaired electron on this residue. This bi-radical state then relaxes back to the ground state, or occasionally leads to decarboxylation of Glu212. A different radical forms when a low dose of X-rays is employed at 100 K, probably because Glu212 remains chemically intact at such dose. At higher doses, however, the susceptibility of Glu212 to decarboxylation is evident (Fig. S3).

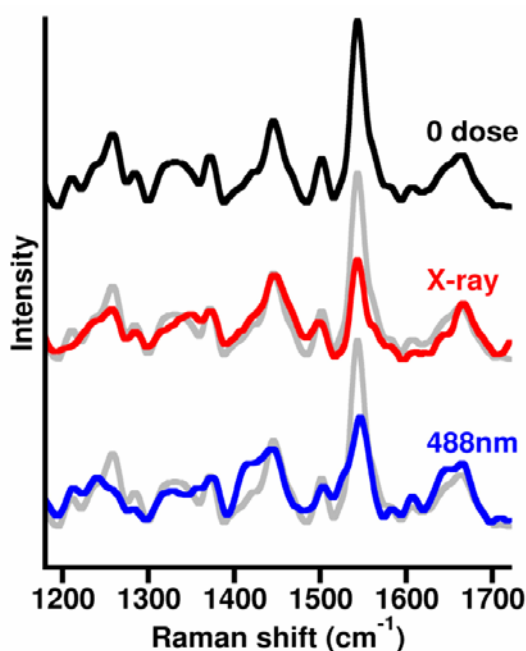


Figure 2. *In crystallo* Raman spectra of IrisFP. Spectra collected at 100 K are shown before X-irradiation (top and grey line), after X-irradiation (0.9 MGy, middle), and after 488-nm illumination (26 min at 1.2 kW/cm², bottom).

Despite the above mentioned differences, we argue that the observed X-ray induced structural changes (Fig. 1b) are relevant to the mechanism of bleaching by visible light, because our Raman data indicate that the C_α-C₅ double bond, key to the chromophore's optical properties, is primarily affected along the bleaching pathways induced by both X-rays and visible light. Thus, the transient structure observed in Fig. 1b probably represents an early intermediate along a photobleaching pathway, trapped before permanent photodestruction. This intermediate may also relate to some of the reversible dark states observed in numerous single-molecule studies.^[26] We suggest that the strongly reducing electrons and oxidizing holes generated by X-ray-induced water radiolysis provide an effective shunt pathway to populate a radical state directly from the ground state (Fig. 3). Because a radical state would form with low probability from the triplet state upon illumination by visible light, substantial accumulation is not expected at the

ensemble level due to competition with other reactions such as the photo-induced protonation of the chromophore that was observed here.

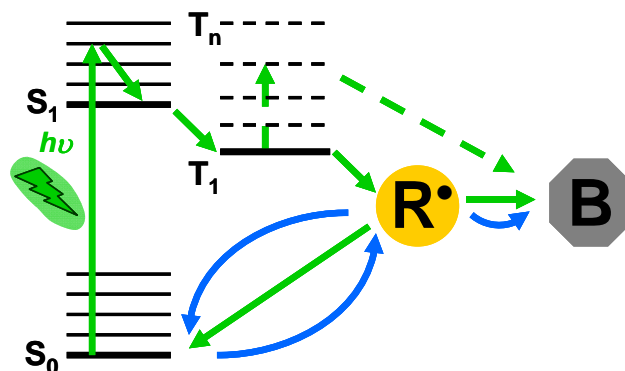


Figure 3. Proposed mechanism for X-ray induced radical generation in IrisFP and possible photobleaching pathways. Green (blue) lines show photobleaching pathways induced by visible (X-ray) light, respectively. Green dashed lines show other possible pathways not discussed in this article. R•: radical; B: bleached.

In organic dyes, a well-known mechanism of photobleaching is the destruction of the chromophore by reactive oxygen species (ROS), which form when molecular oxygen reacts with the excited chromophore.^[27, 28] However, some evidence exists that ROS may not be the dominant cause of photobleaching in FPs,^[29] as their β -can structure protects the chromophore from the solvent and dissolved O₂. Yet, fluorescent proteins are typically one order of magnitude less resistant to photobleaching as compared to popular organic dyes such as cyanines or rhodamines. This observation suggests that photobleaching *via* radical formation might play a comparatively important role in FPs. In fact, in the same way as the protein matrix tends to protect the chromophore against O₂, it may also confine radical states within the chromophore pocket, preventing relaxation via redox reactions with the surrounding medium.

Finally, our results bear consequences for the structure determination of fluorescent proteins by X-ray diffraction. It is known that chromophores in FPs may not be strictly planar due to constraints imposed by the protein matrix.^[30] However, the additional distortion induced by even very mild X-irradiation observed here for IrisFP suggests that the chromophore geometry in a number of FPs whose structure have been solved using synchrotron radiation may be slightly over-bent.

In conclusion, the reversible loss of π -conjugation directly observed in IrisFP may help unravelling the mechanisms of blinking and/or photobleaching in (photoactivatable) fluorescent proteins, which is of importance to rationally design more photostable variants.

Experimental Section

The experimental procedures are briefly sketched in the following; a detailed description is provided as Supporting Information. IrisFP crystals were obtained as described.^[19] X-ray data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF), using beamline ID14-2 (X-ray wavelength $\lambda = 0.931$ Å; X-ray flux $\approx 4.2 \times 10^{10}$ ph/sec/0.1 \times 0.1 mm²). The technique of composite data sets^[22] was used to reconstruct the “low dose” (LD) and

“moderate dose” (MD) data sets. The LD and MD data sets (1.9 Å and 2.0 Å resolution, respectively) were each obtained from 6 subsets of 20 images, collected at different locations of the crystal, translating and rotating the latter between subsequent subsets so as to scan the entire reciprocal space. The absorbed dose during the LD data set was very small (median dose: ~ 0.1 MGy, *i.e.*, 0.05 % of the Henderson limit^[31]), and it was still small during the MD data set (~ 0.5 MGy, *i.e.*, 2% of the Henderson limit).

Absorption, fluorescence emission and Raman spectra from IrisFP crystals were recorded at 100 K using the microspectrophotometer of the Cryobench laboratory, either “offline” for 488nm-photobleaching^[21, 32] or “online” for X-ray-photobleaching (beamline ID14-2 was used for absorption and fluorescence measurements, and beamline ID23-1 with an attenuated flux (4.4×10^{10} ph/s) was used for Raman measurements).^[21, 33] For X-ray photobleaching, a portion of a needle-shaped crystal was irradiated with an X-ray beam of larger dimensions than the probing spectroscopic beam. Bleaching with visible light was carried out at 488 nm (1.2 kW/cm^2) with an argon ion laser.

Supporting Information Available: Crystallization; X-ray data collection; *in crystallo* spectroscopy; Kinetic model for X-ray induced absorbance decay; Phototransformations in IrisFP; Data collection statistics and illustrations.

Received: ((will be filled in by the editorial staff))

Published online on ((will be filled in by the editorial staff))

Keywords: Fluorescent probes • Photobleaching • Radicals • X-ray diffraction • Microspectrophotometry.

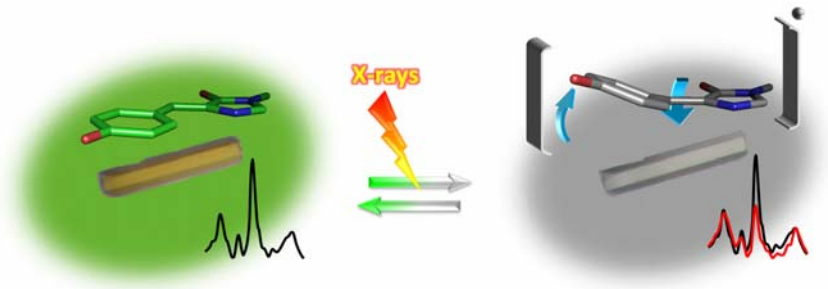
- [1] N. C. Shaner, G. H. Patterson, M. W. Davidson, *J Cell Sci* **2007**, *120*, 4247.
- [2] J. Wiedenmann, G. U. Nienhaus, *Expert Rev Proteomics* **2006**, *3*, 361.
- [3] S. W. Hell, *Nat Methods* **2009**, *6*, 24.
- [4] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, H. F. Hess, *Science* **2006**, *313*, 1642.
- [5] J. White, E. Stelzer, *Trends Cell Biol* **1999**, *9*, 61.
- [6] N. C. Shaner, M. Z. Lin, M. R. McKeown, P. A. Steinbach, K. L. Hazelwood, M. W. Davidson, R. Y. Tsien, *Nat Methods* **2008**, *5*, 545.
- [7] H. W. Ai, J. N. Henderson, S. J. Remington, R. E. Campbell, *Biochem J* **2006**, *400*, 531.
- [8] X. Kong, E. Nir, K. Hamadani, S. Weiss, *J Am Chem Soc* **2007**, *129*, 4643.
- [9] G. Donnert, C. Eggeling, S. W. Hell, *Nat Methods* **2007**, *4*, 81.
- [10] J. P. Hoogenboom, E. M. van Dijk, J. Hernando, N. F. van Hulst, M. F. Garcia-Parajo, *Phys Rev Lett* **2005**, *95*, 097401.
- [11] R. Zondervan, F. Kulzer, M. A. Kol'chenko, M. Orrit, *J. Phys. Chem. B* **2004**, *108*, 1657.
- [12] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, P. Tinnefeld, *Angew Chem Int Ed Engl* **2008**, *47*, 5465.
- [13] C. Steinhauer, C. Forthmann, J. Vogelsang, P. Tinnefeld, *J Am Chem Soc* **2008**, *130*, 16840.
- [14] J. Widengren, A. Chmyrov, C. Eggeling, P. A. Löfdahl, C. A. M. Seidel, *J. Phys. Chem. A* **2007**, *111*, 429.
- [15] J. J. van Thor, T. Gensch, K. J. Hellingwerf, L. N. Johnson, *Nat Struct Biol* **2002**, *9*, 37.
- [16] A. F. Bell, D. Stoner-Ma, R. M. Wachter, P. J. Tonge, *J Am Chem Soc* **2003**, *125*, 6919.
- [17] K. A. Lukyanov, D. M. Chudakov, S. Lukyanov, V. V. Verkhusha, *Nat Rev Mol Cell Biol* **2005**, *6*, 885.
- [18] T. B. McAnaney, W. Zeng, C. F. Doe, N. Bhanji, S. Wakelin, D. S. Pearson, P. Abbyad, X. Shi, S. G. Boxer, C. R. Bagshaw, *Biochemistry* **2005**, *44*, 5510.
- [19] V. Adam, M. Lelimosin, S. Boehme, G. Desfonds, K. Nienhaus, M. J. Field, J. Wiedenmann, S. McSweeney, G. U. Nienhaus, D. Bourgeois, *Proc Natl Acad Sci U S A* **2008**, *105*, 18343.
- [20] J. Wiedenmann, S. Ivanchenko, F. Oswald, F. Schmitt, C. Rocker, A. Salih, K. D. Spindler, G. U. Nienhaus, *Proc Natl Acad Sci U S A* **2004**, *101*, 15905.
- [21] P. Carpentier, A. Royant, J. Ohana, D. Bourgeois, *J Appl Cryst* **2007**, *40*, 1113.
- [22] G. I. Berglund, G. H. Carlsson, A. T. Smith, H. Szoke, A. Henriksen, J. Hajdu, *Nature* **2002**, *417*, 463.
- [23] X. He, A. F. Bell, P. J. Tonge, *J Phys Chem B* **2002**, *106*, 6056.
- [24] S. Luin, V. Voliani, G. Lanza, R. Bizzarri, P. Amat, V. Tozzini, M. Serresi, F. Beltram, *J Am Chem Soc* **2009**, *131*, 96.
- [25] H. Mizuno, T. K. Mal, M. Walchli, A. Kikuchi, T. Fukano, R. Ando, J. Jeyakanthan, J. Taka, Y. Shiro, M. Ikura, A. Miyawaki, *Proc Natl Acad Sci U S A* **2008**, *105*, 9227.
- [26] W. E. Moerner, *J Chem Phys* **2002**, *117*, 10925.
- [27] K. Jacobson, Z. Rajfur, E. Vitriol, K. Hahn, *Trends Cell Biol* **2008**, *18*, 443.
- [28] A. Jimenez-Banzo, S. Nonell, J. Hofkens, C. Flors, *Biophys J* **2008**, *94*, 168.
- [29] R. Swaminathan, C. P. Hoang, A. S. Verkman, *Biophys J* **1997**, *72*, 1900.
- [30] S. L. Maddalo, M. Zimmer, *Photochem Photobiol* **2006**, *82*, 367.
- [31] R. Henderson, *Proc. R. Soc. Lond. B Biol. Sci.* **1990**, *241*, 6
- [32] D. Bourgeois, X. Vernede, V. Adam, E. Fioravanti, T. Ursby, *J. Appl. Cryst* **2002**, *35*, 319.
- [33] J. McGeehan, R. B. Ravelli, J. Murray, R. L. Owen, F. Cipriani, S. McSweeney, M. Weik, E. Garman, *J Synchrotron Radiat* **2009**, *16*, 163.

Structure of a GFP dark state

Virgile Adam, Philippe Carpentier,
Sebastien Violot, Mickaël Lelimosin,
Claudine Darnault, G. Ulrich Nienhaus,
and Dominique Bourgeois*

Page – Page

Structural Basis of Photobleaching in a
Photoactivatable Green Fluorescent
Protein



The photo-activatable fluorescent protein IrisFP was observed in a transient radical state *en route* to photobleaching, with near-atomic resolution. X-rays were used to populate the radical, which under illumination with visible light presumably forms with low probability from the triplet state. Combined X-ray diffraction and *in crystallo* spectroscopic data reveal that radical formation involves pronounced but reversible distortion of the chromophore, suggesting a transient loss of π -conjugation.

Structural Basis of Photobleaching in a Photoactivatable Green Fluorescent Protein

Adam, Carpentier *et al.*

Supporting Information

Crystallization.

IrisFP was expressed in *Escherichia coli* (strain M15pREP4) and purified as described.^[1] Crystals were grown at 20°C in 2.4 M ammonium sulphate, 0.1 M bicine, pH 8.4, using the hanging drop vapour diffusion method. Rod-shaped crystals appeared within 24 h and reached their final size of (0.2 × 0.2 × 0.8) mm³ within a few days. For X-ray diffraction and UV-visible spectroscopy, crystals were rapidly transferred to a cryoprotectant solution (10% glycerol/2.4 M ammonium sulphate/0.1 M bicine, pH 8.4) before being flash-frozen in liquid or gaseous nitrogen at 100 K. Glycerol was not used for Raman spectroscopy to avoid spectral contamination by this molecule.

X-Ray data collection.

Composite X-ray diffraction data sets were collected at the European Synchrotron Radiation Facility (ESRF) on beamline ID14-2 (X-ray wavelength $\lambda = 0.931 \text{ \AA}$). They were integrated and scaled with the XDS program package.^[2] Data collection statistics are compiled in Table S1. The technique of composite data sets^[3] was used to reconstruct the “low dose” (LD) and “moderate dose” (MD) data sets. The LD data set corresponded to a total exposure of 32 sec under a non-attenuated X-ray beam ($4.2 \times 10^{10} \text{ ph/s}$, 0.1×0.1

mm²), whereas the MD data set corresponded to a total exposure of 140 sec under the same flux. The LD and MD data sets were each obtained from 6 subsets of 20 images (1° oscillation per frame), collected at different locations of the crystal, translating the latter by 50 μm and rotating it by 15° between subsequent subsets (overlap of 5°) so as to scan the entire reciprocal space. Only the 16 first images of each subset were used to reconstruct the LD data set, whereas all 20 images were utilized to reconstruct the MD data set. The absorbed dose during the LD data set was very small, running from 0 to 0.18 MGy (*i. e.*, the median dose corresponded to 0.05 % of the Henderson limit^[4]), and it was still small during the MD data set running from 0.23 to 0.80 MGy (*i.e.* 2 % of the Henderson limit). Hence, no sensible degradation of the diffractions patterns could be noticed. Based on an estimation of the crystal content, the approximate number of photons absorbed per unit cell was evaluated with the program Raddose.^[5] Assuming that a single absorbed X-ray photon generates a cascade of up to ~500 e⁻, ~30 e⁻ and ~160 e⁻ (on average) per unit cell may eventually be guided to the chromophores in the LD and MD cases, respectively. Therefore, with ~16 chromophores per unit cell, these findings are consistent with very small and moderate photobleaching of the chromophores in the LD and MD data sets, respectively.

Experimental difference electron density maps ($F_{\text{obs, MD}} - F_{\text{obs, LD}}$) were calculated with CNS,^[6] using phases from the green structure of IrisFP (PDB accession ID 2VVH) and Bayesian q-weighting of the difference structure factor amplitudes.^[7]

Initial model refinement of IrisFP against LD and MD data did not give significant modifications as compared to the deposited model of IrisFP, which was collected under standard synchrotron conditions.^[1] Thus, model refinement was not further attempted. Experimental electron difference density maps are extremely sensitive to conformational changes by only a small fraction of the molecules in the crystal. The fraction of molecules

that changed their state between the LD, MD and “standard” data sets are estimated to be $\leq 20\%$, which does not allow for a reliable refinement at the resolution of our data.

Experimental difference electron density maps ($F_{\text{obs, HD}} - F_{\text{obs, Standard}}$) were calculated with the same procedure as described above for the ($F_{\text{obs, MD}} - F_{\text{obs, LD}}$) map. The structure factor amplitudes $F_{\text{obs, HD}}$ and $F_{\text{obs, Standard}}$ were obtained from a separate experiment, in which a series of standard diffraction data sets were collected one after another on an IrisFP crystal. We estimate that the doses employed for $F_{\text{obs, Standard}}$ and $F_{\text{obs, HD}}$ are \sim twice and ~ 8 times that used for the MD data set, respectively. Due to dose accumulation on the crystal and associated expansion of the crystal unit-cell, the quality of the ($F_{\text{obs, HD}} - F_{\text{obs, Standard}}$) map is lower than that of the ($F_{\text{obs, MD}} - F_{\text{obs, LD}}$) map.

The NCS averaged ($F_{\text{obs, MD}} - F_{\text{obs, LD}}$) and ($F_{\text{obs, HD}} - F_{\text{obs, Standard}}$) maps presented in Figs. 1b and S3, respectively, were calculated with Coot.^[8] Figs. 1b and S3 were prepared with Pymol.^[9]

***In crystallo* spectroscopy.**

Absorption, fluorescence emission and Raman spectra from IrisFP crystals were recorded at 100 K using the microspectrophotometer of the Cryobench laboratory, either “offline” for 488nm-photobleaching^[10, 11] or “online” for X-ray-photobleaching (ESRF beamline ID14-2).^[11, 12] Crystals were mounted with standard cryoloops. Absorption and fluorescence emission spectra were collected using a CCD based spectrometer (HR2000+, Ocean Optics, Dunedin, FL, USA). For online experiments, we verified that the entire volume of the crystal probed by optical beams (cross sections: 25 μm and 50 μm diameter spots in absorbance/fluorescence and Raman mode, respectively) was bathed into the X-ray beam (cross section 0.1 \times 0.1 mm^2). Raman spectra were collected with a Renishaw

InVia spectrophotometer (Renishaw, Gloucestershire, United Kingdom) with 50 mW excitation at 785 nm from a diode laser. Spectra were collected in back-scattering geometry, with a 50× objective (focal spot: $\sim 20 \mu\text{m}^2$) offline, and a 20× objective (focal spot: $\sim 50 \mu\text{m}^2$) online.

For X-ray photobleaching, a portion of a needle shaped crystal was irradiated with a synchrotron X-ray beam (average flux: $\sim 6.2 \times 10^{10}$ ph/sec/ $0.1 \times 0.1 \text{mm}^2$, wavelength: 0.931 Å). Bleaching with visible light was carried out at 488 nm (1.2kW/cm^2) with an argon ion laser (532-MAP-A01, Melles-Griot, Carlsbad, CA, USA). In the latter case, the fluorescence and absorbance decays are complicated by the heterogeneous sample excitation resulting from the small laser penetration depth within the optically dense crystals, as well as by possible heating effects.

Fluorescence mode

During X-ray photobleaching, short laser pulses (~ 20 ms) at 440 nm ($\sim 1 \text{kW/cm}^2$) were used to excite fluorescence emission at 1 Hz. We verified that, in the absence of X-rays, no noticeable fluorescence decay occurred due to the laser excitation beam (not shown). For visible light photobleaching, the actinic 488-nm light was also used to excite fluorescence.

Absorption mode

Spectra were recorded with a white light source (DH2000-BAL, Ocean Optics, Dunedin, FL, USA). For visible light photobleaching, 488-nm photobleaching (5 s) and absorption spectra recording (10 ms) were interleaved.

Raman mode

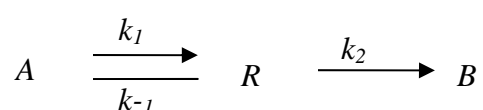
Raman spectra were recorded at 100 K from IrisFP crystals mounted in standard cryoloops. For X-ray photobleaching, Raman spectra were collected online (ESRF beamline ID23-1), before and after 180 s exposure to an attenuated X-ray beam (4.4×10^{10} ph/s, 1% transmission), equivalent to a 0.9 MGy absorbed dose. For visible light photobleaching, Raman spectra were collected offline, before and after 1560 s exposure to the 488-nm laser (1.7 kW/cm^2 at sample). For recording full spectra ($200 - 2000 \text{ cm}^{-1}$), a total exposure time of 15 min was used. Data-smoothing, baseline subtraction and cosmic ray removal were carried out with the program Wire (<http://www.renishaw.com>).

Solution studies

In absorption and fluorescence mode, photobleaching experiments were also performed with thin films of frozen IrisFP solution containing 30% glycerol. Under X-ray illumination, reversible and non-reversible bleaching were observed similarly to the crystalline phase (not shown). Under illumination by 488 nm light, non-reversible bleaching was also observed, although a pH-dependant fraction of the molecules apparently underwent a reversible photon-induced protonation reaction as a side pathway, as in crystals (not shown).

Kinetic model for X-ray induced absorbance decay

The absorbance decay of IrisFP at 485 nm (Fig. S1, inset) was fitted to a kinetic model representative of the X-ray induced bleaching pathway depicted in Fig. 3:



Model fitting was performed with a home-made routine based on the IDL software (ITT, Boulder, USA). Under illumination by X-rays (first part of the decay), the following rate constants were obtained: $k_1 = 0.0134 \text{ s}^{-1}$ ($\tau_{1/2} = 51 \text{ s}$); $k_{-1} = 0.028 \text{ s}^{-1}$ ($\tau_{1/2} = 24 \text{ s}$); $k_2 = 0.0032 \text{ s}^{-1}$ ($\tau_{1/2} = 216 \text{ s}$). Under the used X-ray flux density, these values convert to: $k_1 = 2.16 \text{ MGy}^{-1}$ ($\tau_{1/2} = 0.32 \text{ MGy}$); $k_{-1} = 4.51 \text{ MGy}^{-1}$ ($\tau_{1/2} = 0.15 \text{ MGy}$); $k_2 = 0.52 \text{ MGy}^{-1}$ ($\tau_{1/2} = 1.35 \text{ MGy}$). The recovery of absorbance at 100 K is characterized by a rate constant $k_{-1, \text{off}} = 0.0048 \text{ s}^{-1}$ ($\tau_{1/2} = 146 \text{ s}$), as deduced by exponential fitting of the second part of the absorbance trace (Fig. S1, inset). Thus, the rate constant k_{-1} is essentially photon driven.

It should be noted that a kinetic model in which species B is attained directly from A fits the data equally well, with little changes in the values of k_1 , k_{-1} and k_2 . Thus our data do not tell whether or not the bleached state of Fig. S3 is produced *via* the radical state R.

Photo-transformations in IrisFP.

IrisFP is a photo-activatable fluorescence protein that has been shown to undergo a variety of photo-transformations, notably on/off reversible photoswitching and green/red irreversible photoconversion.^[1] Because they involve dark, thermally activated steps, the reaction pathways that normally lead to these photo-transformations cannot come to completion at 100 K, so that all molecules eventually end up in a irreversibly photobleached state at this temperature. In contrast, irreversible ensemble bleaching of IrisFP with 488-nm light cannot be easily achieved at room-temperature because photoswitching rather than photobleaching occurs.

SI Tables

Table S1. Data collection statistics for the composite LD and MD data sets. Values in parentheses refer to the highest resolution shell.

	LD	MD
Median absorbed dose, MGy	0.09	0.5
Beamline	ESRF / ID14-2	ESRF / ID14-2
Wavelength, Å	0.933	0.933
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , Å	85.54	85.51
<i>b</i> , Å	96.47	96.47
<i>c</i> , Å	139.81	139.80
Resolution, Å	50.0-2.00 (2.11-2.00)	50.0-1.86 (1.97-1.86)
<i>R</i> _{sym} , %*	10.1 (42.9)	7.9 (40.4)
Mean <i>I</i> / σ (<i>I</i>)	10.3 (3.1)	13.9 (3.2)
Completeness, %	95.6 (94.4)	93.3 (70.7)
Redundancy	3.8 (3.8)	4.8 (3.9)
No. of unique reflections	75305 (10884)	91100(10786)
Wilson B factor, Å ²	19.9	19.0

$$*R_{\text{sym}} = \frac{\sum_j \sum_h |I_{h,j} - \langle I_h \rangle|}{\sum_j \sum_h I_{h,j}}$$

References.

- [1] V. Adam, M. Lelimosin, S. Boehme, G. Desfonds, K. Nienhaus, M. J. Field, J. Wiedenmann, S. McSweeney, G. U. Nienhaus, D. Bourgeois, *Proc Natl Acad Sci U S A* **2008**, *105*, 18343.
- [2] W. Kabsch, *J App Cryst* **1993**, *26*, 795.
- [3] G. I. Berglund, G. H. Carlsson, A. T. Smith, H. Szoke, A. Henriksen, J. Hajdu, *Nature* **2002**, *417*, 463.
- [4] R. Henderson, *Proc. R. Soc. Lond. B Biol. Sci.* **1990**, *241*, 6
- [5] J. W. Murray, E. Rudino-Pinera, R. L. Owen, M. Grininger, R. B. Ravelli, E. F. Garman, *J Synchrotron Radiat* **2005**, *12*, 268.
- [6] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr D Biol Crystallogr* **1998**, *54* (Pt 5), 905.

- [7] T. Ursby, D. Bourgeois, *Acta Crystallogr. A* **1997**, *53*, 564.
- [8] P. Emsley, K. Cowtan, *Acta Cryst* **2004**, *D60*, 2126.
- [9] (Ed.: D. S. LLC), San Carlos, CA, **2003**.
- [10] D. Bourgeois, X. Vernede, V. Adam, E. Fioravanti, T. Ursby, *J. Appl. Cryst* **2002**, *35*, 319.
- [11] P. Carpentier, A. Royant, J. Ohana, D. Bourgeois, *J Appl Cryst* **2007**, *40*, 1113.
- [12] J. McGeehan, R. B. Ravelli, J. Murray, R. L. Owen, F. Cipriani, S. McSweeney, M. Weik, E. Garman, *J Synchrotron Radiat* **2009**, *16*, 163.

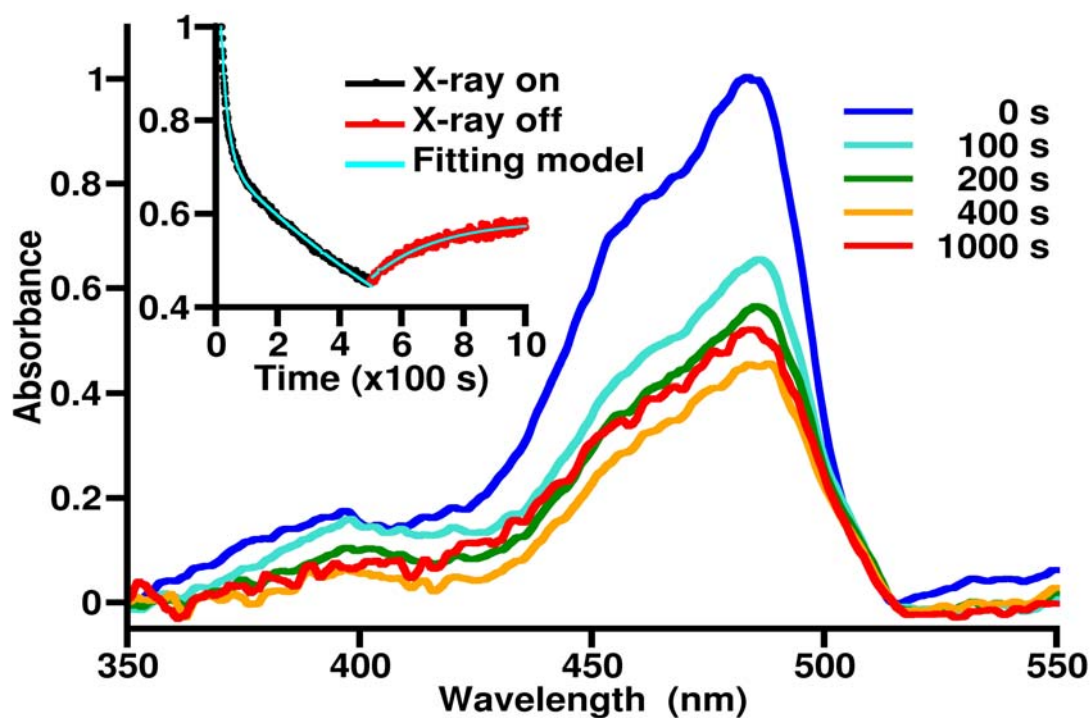


Fig. S1

Absorbance decay of an IrisFP crystal under an X-ray beam (6.2×10^{10} ph/s) at 100 K. A selected set of baseline corrected spectra are shown. The absorbance decay kinetics at 485 nm (black color) is displayed in the inset, showing a clear biphasic behavior. Recovery in the absence of X-rays is shown in red. Fitting to the kinetic model described in the SI is shown in blue. Attempts to record absorbance recovery after excursion to room temperature failed, due to strong distortions of the baseline induced by the annealing procedure.

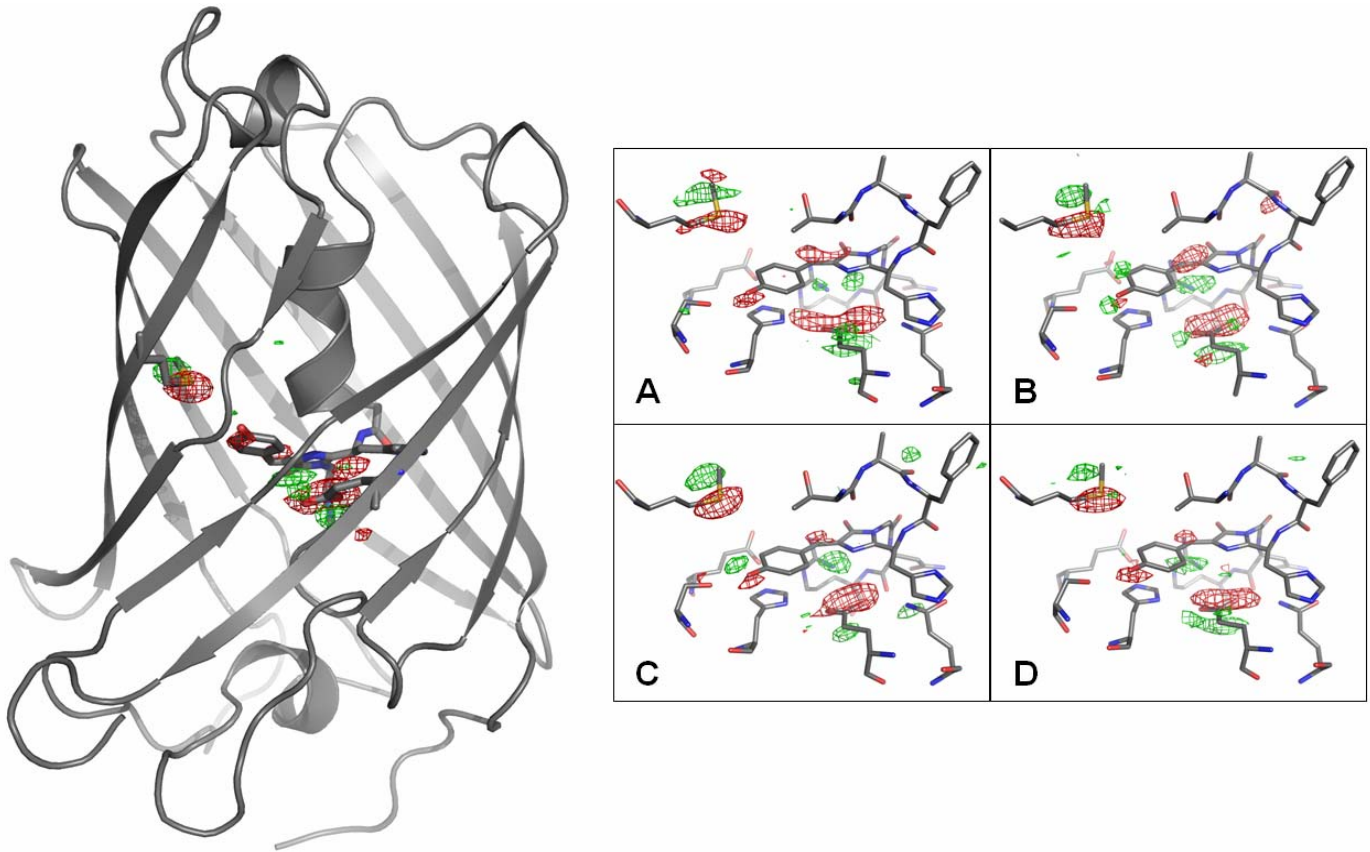


Fig. S2

Left panel: Overall structural view of a monomer of IrisFP (monomer A). The chromophore, Met159 and Glu212 are shown as sticks. The NCS averaged electron difference density map (red: negative; green: positive) between the LD and MD data sets is superimposed. Only structural changes at the chromophore and its immediate environment are seen. At the low X-ray doses employed, no other signs of noise or radiation damage (such as decarboxylation of Asp and Glu residues) can be seen elsewhere in the protein.

Right panel: Chromophore pocket in all 4 monomers A, B, C and D of the crystal asymmetric unit. The non-NCS averaged electron difference density map (red: negative; green: positive) between the LD and MD data sets is shown. No substantial variation from monomer to monomer is observed.

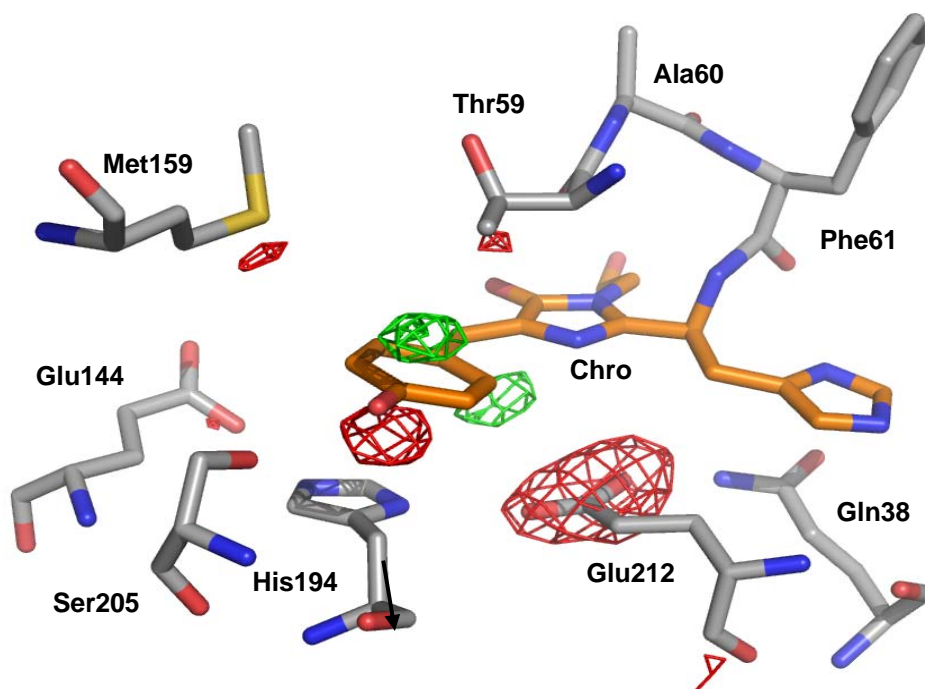


Fig. S3

Experimental electron difference density map (red: -8.0σ ; green: 8.0σ) between the HD and standard data sets, overlaid on the model of IrisFP. The NCS difference map is shown, averaged over the 4 IrisFP monomers in the crystal asymmetric unit. The chromophore is shown in orange, while neighboring residues are shown in grey. Arg66, in contact with the chromophore, is omitted for clarity.

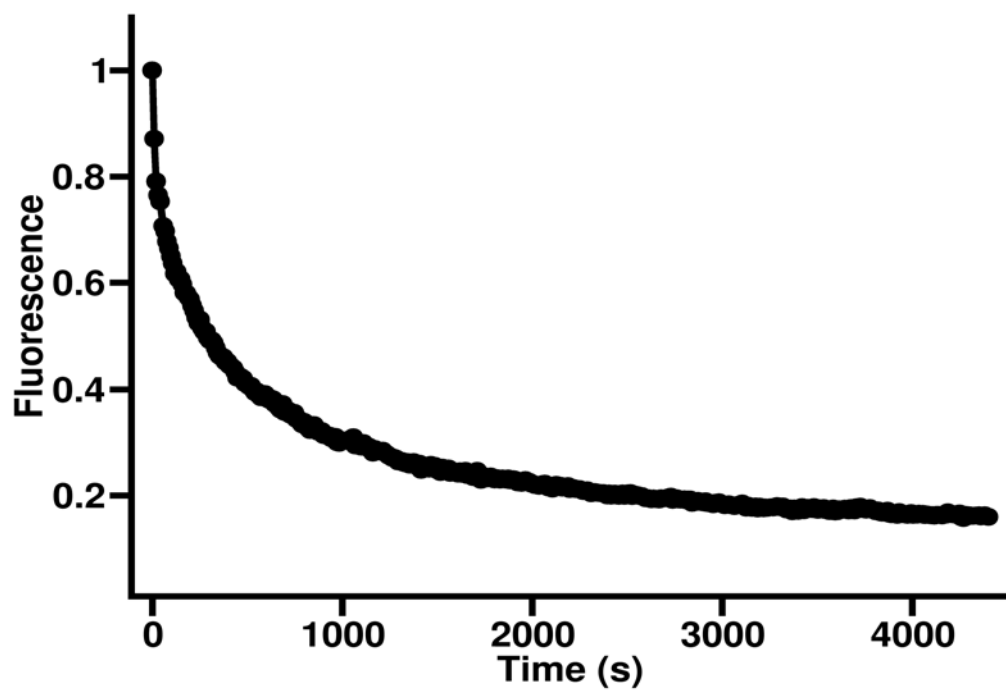


Fig. S4

Fluorescence emission decay of an IrisFP crystal under 488 nm illumination (1.2 kW/cm²) at 100 K.

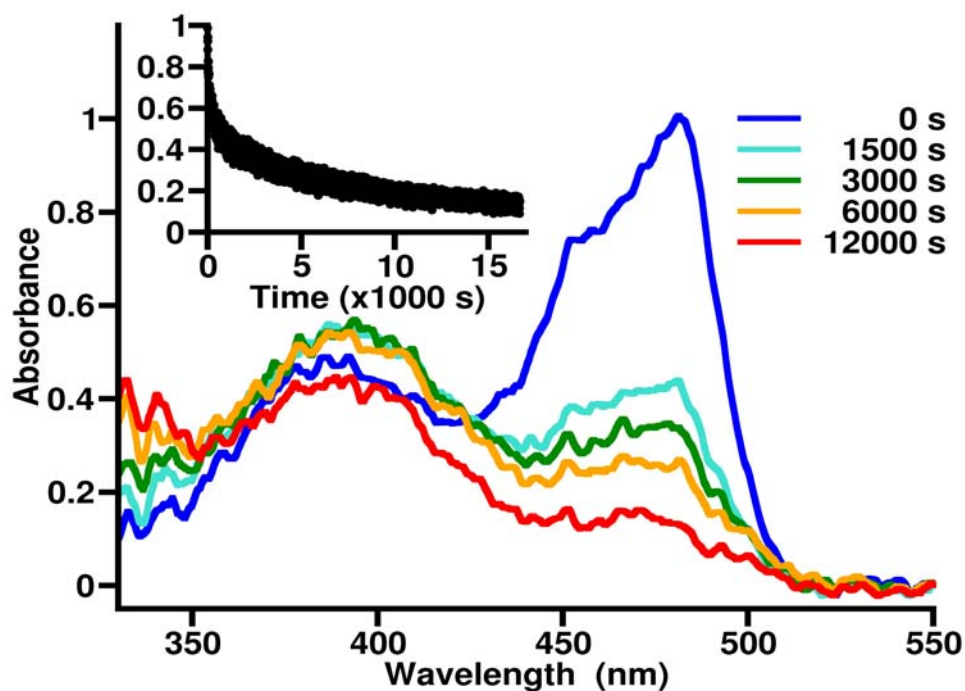


Fig. S5

Absorbance decay of an IrisFP crystal under 488 nm illumination (1.2 kW/cm^2) at 100 K. A selected set of baseline corrected spectra are shown. The absorbance decay kinetics at 485 nm is displayed in the inset. The fast decay phase is assigned to a reversible photo-induced protonation of the chromophore, whereas the slow phase is assigned to irreversible photo-bleaching.

Bibliographie

- [1] A. Warshel, S. T. Russell, and A. K. Churg. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 81(15) :4785–4789, 1984.
- [2] A. Warshel. *Journal of Biological Chemistry*, 273(42) :27035–27038, 1998.
- [3] M. Planck. *Annalen der Physik*, 4 :553–563, 1901.
- [4] A. Einstein. *Annalen der Physik*, XVII :132–148, 1905.
- [5] L.V. de Broglie. *Thèse de doctorat*. 1924.
- [6] C.J. Davisson and L.H. Germer. *Nature*, 119 :558–560, 1927.
- [7] W. Heisenberg. *Z. für Phys.*, 43 :172–198, 1927.
- [8] E. Schrödinger. *Annalen der Physik*, 79(4) :361–376, 1926.
- [9] N. Bohr. *Philos. Mag.*, 26(6) :1–25, 1913.
- [10] M. Born and R. Oppenheimer. Zur quantumtheorie der molekeln. *Ann. Physik*, 84 :457–484, 1927.
- [11] R.S. Mulliken. *Phys.Rev.*, 32 :186, 1927.
- [12] J.C. Slater. *Phys. Rev.*, 36 :57, 1930.
- [13] W. Pauli. *Naturwissenschaften*, 12 :741, 1924.
- [14] G. Uhlenbeck and S. Goudsmit. *Naturwissen*, 13 :953, 1925.
- [15] P. A. M. Dirac. *Proceedings of the Royal Society (London)*, 117 :610–612, 1928.
- [16] W. Pauli. *Zeitschrift für Physik*, 31 :765, 1925.
- [17] D.R. Hartree. *Roc. Camb. Phil. Soc.*, 24 :111, 1928.
- [18] V. Fock. *Z. Phys.*, 61 :126, 1930.
- [19] C.C.J. Roothaan. *Reviews of Modern Physics*, 23(69), 1951.
- [20] G.G. Hall. *Proceedings of the Royal Society (London) A*, 205(541), 1951.

- [21] P. Hohenberg and W. Kohn. *Physical Review B*, 136(3B) :864, 1964.
- [22] W. Kohn and L. J. Sham. *Physical Review*, 140(4A) :1133, 1965.
- [23] D. M. Ceperley and B. J. Alder. *Physical Review Letters*, 45(7) :566–569, 1980.
- [24] A. D. Becke. *Physical Review A*, 38 :3098–3100, 1988.
- [25] C. T. Lee, W. T. Yang, and R. G. Parr. *Physical Review B*, 37(2) :785–789, 1988.
- [26] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. *Journal of Physical Chemistry*, 98(45) :11623–11627, 1994.
- [27] M. J. S. Dewar and W. Thiel. *Journal of the American Chemical Society*, 99(15) :4899–4907, 1977.
- [28] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. *Journal of the American Chemical Society*, 107(13) :3902–3909, 1985.
- [29] G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart. *Journal of Computational Chemistry*, 27(10) :1101–1111, 2006.
- [30] J.J.P. Stewart. *J. Comp. Chem.*, 10 :209–220, 1989.
- [31] M. P. Repasky, J. Chandrasekhar, and W. L. Jorgensen. *Journal of Computational Chemistry*, 23(16) :1601–1622, 2002.
- [32] S.M. Bachrach. *Rev. Comp. Chem.*, 5 :171, 1994.
- [33] K. B. Wiberg and P. R. Rablen. *Journal of Computational Chemistry*, 14(12) :1504–1518, 1993.
- [34] R. S. Mulliken. *Journal of Chemical Physics*, 36(12) :3428, 1962.
- [35] I. Mayer. *Chemical Physics Letters*, 97(3) :270–274, 1983.
- [36] P.-O. Löwdin. *Adv. Quant. Chem.*, 5 :185, 1970.
- [37] A. E. Reed, L. A. Curtiss, and F. Weinhold. *Chemical Reviews*, 88(6) :899–926, 1988.
- [38] R.F.W. Bader. *Chem. Rev.*, 91 :893, 1991.
- [39] F. L. Hirshfeld. *Theoretica Chimica Acta*, 44(2) :129–138, 1977.
- [40] F. De Proft, C. Van Alsenoy, A. Peeters, W. Langenaeker, and P. Geerlings. *Journal of Computational Chemistry*, 23(12) :1198–1209, 2002.
- [41] C. F. Guerra, J. W. Handgraaf, E. J. Baerends, and F. M. Bickelhaupt. *Journal of Computational Chemistry*, 25(2) :189–210, 2004.

- [42] B. Rousseau, A. Peeters, and C. Van Alsenoy. *Journal of Molecular Structure-Theochem*, 538 :235–238, 2001.
- [43] P. Politzer and R. R. Harris. *Journal of the American Chemical Society*, 92(22) :6451, 1970.
- [44] P.M.W. Gill. *J. Phys. Chem.*, 100(15) :421, 1996.
- [45] J. Cioslowski. *Journal of the American Chemical Society*, 111(22) :8333–8336, 1989.
- [46] U. C. Singh and P. A. Kollman. *Journal of Computational Chemistry*, 5(2) :129–145, 1984.
- [47] B. H. Besler, K. M. Merz, and P. A. Kollman. *Journal of Computational Chemistry*, 11(4) :431–439, 1990.
- [48] M. M. Francl and L. E. Chirlian. *Reviews in Computational Chemistry, Vol 14*, 14 :1–31, 2000.
- [49] D.E. Williams. *Rev. Comp. Chem.*, 2 :219, 1991.
- [50] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. *Journal of Physical Chemistry*, 97(40) :10269–10280, 1993.
- [51] C. A. Reynolds, J. W. Essex, and W. G. Richards. *Journal of the American Chemical Society*, 114(23) :9075–9079, 1992.
- [52] R. T. Sanderson. *Science*, 114(2973) :670–672, 1951.
- [53] W. J. Mortier, K. Vangenechten, and J. Gasteiger. *Journal of the American Chemical Society*, 107(4) :829–835, 1985.
- [54] A. K. Rappe and W. A. Goddard. *Journal of Physical Chemistry*, 95(8) :3358–3363, 1991.
- [55] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J. P. Tollenaere. *Journal of Physical Chemistry A*, 106(34) :7887–7894, 2002.
- [56] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, C. Van Alsenoy, and J. P. Tollenaere. *Journal of Physical Chemistry A*, 106(34) :7895–7901, 2002.
- [57] J. W. Storer, D. J. Giesen, C. J. Cramer, and D. G. Truhlar. *Journal of Computer-Aided Molecular Design*, 9(1) :87–110, 1995.

- [58] J. B. Li, T. H. Zhu, C. J. Cramer, and D. G. Truhlar. *Journal of Physical Chemistry A*, 102(10) :1820–1831, 1998.
- [59] J. D. Thompson, J. D. Xidos, T. M. Sonbuchner, C. J. Cramer, and D. G. Truhlar. *Physchemcomm*, 5(18) :117–134, 2002.
- [60] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly. *Journal of Computational Chemistry*, 21(2) :132–146, 2000.
- [61] A. Jakalian, D. B. Jack, and C. I. Bayly. *Journal of Computational Chemistry*, 23(16) :1623–1641, 2002.
- [62] A. D. Becke. *Journal of Chemical Physics*, 88(4) :2547–2553, 1988.
- [63] J. C. Slater. *Journal of Chemical Physics*, 41(10) :3199, 1964.
- [64] J.C. Slater. *Quantum Theory of Molecules and Solids*, volume 2. McGraw-Hill, New York, 1965.
- [65] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch. *Chemical Physics Letters*, 257(3-4) :213–223, 1996.
- [66] A. J. Stone. *Chemical Physics Letters*, 83(2) :233–239, 1981.
- [67] M. J. Frisch et al. Gaussian 98, 1998. Gaussian Inc. : Pittsburgh, PA.
- [68] M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, and A. Thomas. *Journal of Computational Chemistry*, 21(12) :1088–1100, 2000.
- [69] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. *Journal of Computational Chemistry*, 4(2) :187–217, 1983.
- [70] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. *Journal of the American Chemical Society*, 117(19) :5179–5197, 1995.
- [71] W. L. Jorgensen, D. S. Maxwell, and J. TiradoRives. *Journal of the American Chemical Society*, 118(45) :11225–11236, 1996.
- [72] W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren. *Journal of Physical Chemistry A*, 103(19) :3596–3607, 1999.
- [73] N. L. Allinger. *Journal of the American Chemical Society*, 99(25) :8127–8134, 1977.
- [74] S. L. Price and A. J. Stone. *Journal of the Chemical Society-Faraday Transactions*, 88(13) :1755–1763, 1992.

- [75] U. Koch, P. L. A. Popelier, and A. J. Stone. *Chemical Physics Letters*, 238(4-6) :253–260, 1995.
- [76] C. F. Matta and R. F. W. Bader. *Proteins-Structure Function and Genetics*, 40(2) :310–329, 2000.
- [77] A. Warshel and M. Levitt. *Journal of Molecular Biology*, 103(2) :227–249, 1976.
- [78] U. C. Singh and P. A. Kollman. *Journal of Computational Chemistry*, 7(6) :718–730, 1986.
- [79] M. J. Field, P. A. Bash, and M. Karplus. *Journal of Computational Chemistry*, 11(6) :700–733, 1990.
- [80] P. Amara and M. J. Field. *Theoretical Chemistry Accounts*, 109(1) :43–52, 2003.
- [81] H. Lin and D. G. Truhlar. *Journal of Physical Chemistry A*, 109(17) :3991–4004, 2005.
- [82] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field. *Journal of Physical Chemistry A*, 102(24) :4714–4721, 1998.
- [83] N. Ferre, X. Assfeld, and J. L. Rivail. *Journal of Computational Chemistry*, 23(6) :610–624, 2002.
- [84] H. Lin and D. G. Truhlar. *Theoretical Chemistry Accounts*, 117(2) :185–199, 2007.
- [85] M. J. Field. *Journal of Chemical Theory and Computation*, 4(7) :1151–1161, 2008.
- [86] H. M. Senn and W. Thiel. *Angewandte Chemie-International Edition*, 48(7) :1198–1229, 2009.
- [87] P. Pulay. *Molecular Physics*, 17(2) :197, 1969.
- [88] M. Head-Gordon and J. A. Pople. *Journal of Physical Chemistry*, 92(11) :3063–3069, 1988.
- [89] M. J. Field. *Journal of Physical Chemistry*, 95(13) :5104–5108, 1991.
- [90] N. C. Shaner, G. H. Patterson, and M. W. Davidson. *Journal of Cell Science*, 120(24) :4247–4260, 2007.
- [91] G. U. Nienhaus and J. Wiedenmann. *Chemphyschem*, 2009. in press.
- [92] A. A. Voityuk, M. E. Michel-Beyerle, and N. Rosch. *Chemical Physics*, 231(1) :13–25, 1998.

- [93] W. Weber, V. Helms, J. A. McCammon, and P. W. Langhoff. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11) :6177–6182, 1999.
- [94] V. Helms, C. Winstead, and P. W. Langhoff. *Journal of Molecular Structure-Theochem*, 506 :179–189, 2000.
- [95] V. Tozzini and R. Nifosi. *Journal of Physical Chemistry B*, 105(24) :5797–5803, 2001.
- [96] M. E. Martin, F. Negri, and M. Olivucci. *Journal of the American Chemical Society*, 126(17) :5452–5464, 2004.
- [97] J. Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara, and H. Nakatsuji. *Journal of Computational Chemistry*, 28(15) :2443–2452, 2007.
- [98] R. Nifosi, P. Amat, and V. Tozzini. *Journal of Computational Chemistry*, 28(14) :2366–2377, 2007.
- [99] S. Olsen and S. C. Smith. *Journal of the American Chemical Society*, 130(27) :8677–8689, 2008.
- [100] A. Sinicropi, T. Andruniow, N. Ferre, R. Basosi, and M. Olivucci. *Journal of the American Chemical Society*, 127(33) :11534–11535, 2005.
- [101] I. Demachy, J. Ridard, H. Laguitton-Pasquier, E. Durnerin, G. Vallverdu, P. Archirel, and B. Levy. *Journal of Physical Chemistry B*, 109(50) :24121–24133, 2005.
- [102] B. Grigorenko, A. Savitsky, I. Topol, S. Burt, and A. Nemukhin. *Journal of Physical Chemistry B*, 110(37) :18635–18640, 2006.
- [103] L. V. Schafer, G. Groenhof, A. R. Kligen, G. M. Ullmann, M. Boggio-Pasqua, M. A. Robb, and H. Grubmuller. *Angewandte Chemie-International Edition*, 46(4) :530–536, 2007.
- [104] L. V. Schafer, G. Groenhof, M. Boggio-Pasqua, M. A. Robb, and H. Grubmuller. *Plos Computational Biology*, 4(3) :1–14, 2008.
- [105] V. Helms, T. P. Straatsma, and J. A. McCammon. *Journal of Physical Chemistry B*, 103(16) :3263–3269, 1999.
- [106] N. Reuter, H. Lin, and W. Thiel. *Journal of Physical Chemistry B*, 106(24) :6310–6321, 2002.

- [107] R. Nifosi and V. Tozzini. *Proteins-Structure Function and Genetics*, 51(3) :378–389, 2003.
- [108] M. A. Lill and V. Helms. *Proceedings of the National Academy of Sciences of the United States of America*, 99(5) :2778–2781, 2002.
- [109] R. Nifosi and V. Tozzini. *Chemical Physics*, 323(2-3) :358–368, 2006.
- [110] M. Andresen, M. C. Wahl, A. C. Stiel, F. Grater, L. V. Schafer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmuller, S. W. Hell, and S. Jakobs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37) :13070–13074, 2005.
- [111] R. Heim, D. C. Prasher, and R. Y. Tsien. *Proc Natl Acad Sci U S A*, 91(26) :12501–4, 1994.
- [112] R. Heim and R. Y. Tsien. *Curr Biol*, 6(2) :178–82, 1996.
- [113] R. Y. Tsien. *Annual Review of Biochemistry*, 67 :509–544, 1998.
- [114] M. V. Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov, and S. A. Lukyanov. *Nature Biotechnology*, 17(10) :969–973, 1999.
- [115] M. Chattoraj, B. A. King, G. U. Bublitz, and S. G. Boxer. *Proceedings of the National Academy of Sciences of the United States of America*, 93(16) :8362–8367, 1996.
- [116] J. W. Borst, M. A. Hink, A. van Hoek, and A. J. W. G. Visser. *Journal of Fluorescence*, 15(2) :153–160, 2005.
- [117] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien. *Nature Methods*, 2(12) :905–909, 2005.
- [118] V. V. Verkhusha and K. A. Lukyanov. *Nature Biotechnology*, 22(3) :289–296, 2004.
- [119] J. Wiedenmann and G. U. Nienhaus. *Expert Review of Proteomics*, 3(3) :361–374, 2006.
- [120] J. C. Tully. *Journal of Chemical Physics*, 93(2) :1061–1071, 1990.
- [121] A. Warshel and Z. T. Chu. *Journal of Physical Chemistry B*, 105(40) :9857–9871, 2001.
- [122] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. De Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark, and M. A. Robb. *Journal of the American Chemical Society*, 126(13) :4228–4233, 2004.

- [123] S. W. Hell. *Nature Methods*, 6(1) :24–32, 2009.
- [124] M. Sauer. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27) :9433–9434, 2005.
- [125] R. W. Zwanzig. *Journal of Chemical Physics*, 22(8) :1420–1426, 1954.
- [126] J.G. Kirkwood. *J. Chem. Phys.*, 3 :300, 1935.

Traitement des interactions électrostatiques dans les systèmes moléculaires

Les processus réactionnels des systèmes biologiques sont désormais modélisés par des méthodes de plus en plus précises de mécanique quantique (MQ) associées aux champs de force de mécanique moléculaire (MM). Cependant, cette amélioration induit généralement des problèmes dans le traitement habituel des interactions électrostatiques entre les atomes des régions MQ et MM d'une protéine. Dans cette thèse, nous proposons un modèle plus cohérent des interactions électrostatiques MQ/MM. Pour ce faire, une série de méthodes a été développée et testée pour le calcul de charges partielles des atomes de la région MQ. Ces méthodes ont ensuite été utilisées pour l'évaluation des interactions électrostatiques MQ/MM. L'application de notre approche à des systèmes moléculaires simples révèle un niveau de précision satisfaisant. Ces résultats valident notre démarche qui devrait permettre d'appliquer prochainement aux protéines des potentiels hybrides MQ/MM efficaces et précis sans altération de la modélisation des interactions électrostatiques.

Etude par simulation numérique de protéines fluorescentes

Une étude de dynamique moléculaire a permis de détecter les fines interactions de van der Waals qui conditionnent l'amélioration des propriétés de fluorescence dans la Cerulean qui est un mutant de l'ECFP (*Enhanced Cyan Fluorescent Protein*). Nous avons également étudié le mécanisme de photoconversion de la protéine fluorescente EosFP par l'utilisation de potentiels MQ/MM appropriés au traitement des états excités. L'analyse des interactions électrostatiques révèle le pouvoir catalytique déterminant de certains résidus polaires dans l'environnement du chromophore. Finalement, la stabilité thermodynamique des différents états structuraux de la protéine IrisFP a été évaluée. Là encore, les interactions électrostatiques jouent un rôle manifeste dans les propriétés photophysiques complexes de ce nouveau marqueur biologique. De façon générale, ces études de modélisation moléculaire améliorent notre compréhension des protéines fluorescentes afin de contribuer à leur développement pour l'imagerie cellulaire.

Treatment of electrostatic interactions in molecular systems

Reaction processes in biological systems are henceforth modeled by more and more advanced quantum mechanics (QM) methods coupled with force fields of molecular mechanics (MM). However, this improvement involves generally problems in the usual treatment of electrostatic interactions between atoms from QM and MM regions of a protein. In the present thesis, we propose a more consistent model for QM/MM electrostatic interactions. Thus, a set of methods has been developed and tested for the calculation of atomic charges in QM region. Then, these methods have been used for the evaluation of QM/MM electrostatic interactions. Application to simple molecular systems reveals a satisfactory accuracy level. These results validate our approach which may allow soon applications of precise and efficient QM/MM hybrid potentials to proteins without any problems in the modeling of electrostatic interactions.

Computational study of Fluorescent Proteins

A molecular dynamics study has identified weak van der Waals interactions which control the improved fluorescence for Cerulean, a variant of ECFP (*Enhanced Cyan Fluorescent Protein*). We have also investigated the photoconversion mechanism of the fluorescent protein EosFP using predictive QM/MM potentials for excited states. Analysis of electrostatic interactions reveals catalytic effects from polar residues in the chromophore environment. Finally, thermodynamic stability of different structures in IrisFP has been estimated. Electrostatic interactions are still critical for the complex photophysical properties of this new biological marker. In conclusion, these molecular modeling studies improve our knowledge about fluorescent proteins in order to develop more advanced highlighters for cell imaging.