HAL
open science

QM/MM modeling of the retinal chromophore in complex environments

Baptiste Démoulin

To cite this version:

Baptiste Démoulin. QM/MM modeling of the retinal chromophore in complex environments. Theoretical and/or physical chemistry. Université de Lyon, 2017. English. NNT : 2017LYSEN049 . tel-01661591v2

HAL Id: tel-01661591

https://theses.hal.science/tel-01661591v2

Submitted on 13 Dec 2017

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

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

Numéro National de Thèse : 2017LYSEN049

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée par
l'École Normale Supérieure de Lyon

École Doctorale N°206
École Doctorale de Chimie (Chimie, Procédés, Environnement)

Spécialité de doctorat : Chimie

Soutenue publiquement le 21/09/2017, par :

Baptiste DEMOULIN

QM/MM modeling of the retinal chromophore in complex environments.

Modèles QM/MM du chromophore rétinal dans des environnements complexes.

Devant le jury composé de :

DANIEL, Chantal	DR CNRS	LCQS	Rapporteure
NAVIZET, Isabelle	Professeure	MSME	Rapporteure
CERULLO, Giulio	Professor	Politecnico di Milano	Examinateur
LAURENT, Adèle	CR CNRS	CEISAM	Examinatrice
LOFFREDA, David	DR CNRS	LC ENS de Lyon	Examinateur
GARAVELLI, Marco	Professor	Universita di Bologna	Directeur de thèse
RIVALTA, Ivan	CR CNRS	LC ENS de Lyon	Co-encadrant de thèse

Abstract We have used our QM/MM interface to model the photochemical and photophysical properties of the retinal chromophore in several environments.

First, we proved that methylation of the retinal backbone, which converts a slow photochemistry to an ultra-fast protein-like behaviour in methanol solution, modifies the interplay between the retinal excited states, favouring the formation of a photo-active transient intermediate. Then, we have studied the direct effect of the environment in the case of rhodopsin mimics, where point mutations of a few amino-acids lead to systems that can absorb in the wide visible range. Combined with ultra-fast pump-probe spectroscopy, our method has shown that the electrostatic potential around the retinal can affect the shape of the excited potential energy surface, and is able to tune the excited state lifetime as well as the location of the photoisomerization. Next, we showed that the currently accepted protonation state of amino-acids in the vicinity of the retinal in bacteriorhodopsin leads to a strongly blue shifted absorption, while the protonation of Asp212 leads to accurate results; we now aim toward a validation of this protonation by computation of fluorescence and excited state lifetime. Finally, we have modeled the photophysics of the unprotonated Schiff base in a UV-pigment, where an original an previously unreported photochemistry takes place, especially with the direct involvement of a doubly excited state.

These studies have shown the reliability of our QM/MM potential for modeling a wide range of different environments.

Résumé Nous avons appliqué notre interface QM/MM pour modéliser les propriétés photophysiques et photochimiques du chromophore rétinal dans plusieurs environnements.

Nous avons commencé par montrer que la méthylation du squelette carbonné du rétinal, qui transforme une photochimie lente en un processus ultra-rapide, comme dans une protéine, dans une solution de méthanol, modifie l'interaction entre les états excités du rétinal, et favorise la formation d'une espèce transitoire réactive. Nous avons ensuite étudié l'effet direct de l'environnement dans le cas de mimiques de la rhodopsine, où des mutations ponctuelles de quelques acides aminés donnent des systèmes qui absorbent sur toute la gamme du visible. En combinaison avec la spectroscopie pompe-sonde ultra-rapide, notre méthode a montré que le potentiel électrostatique autour du rétinal peut affecter la forme des surfaces d'énergies potentielles excitées, et peut moduler le temps de vie de l'état excité ainsi que le lieu de photo-isomérisation. Ensuite, nous avons montré que l'état de protonation standard des acides aminés proches du rétinal dans la bacteriorhodopsine mènent à une surestimation de l'énergie d'absorption, alors que la protonation du résidu Asp212 donne des résultats plus précis ; nous souhaitons maintenant valider ce modèle par le calcule des propriétés de fluorescence et de temps de vie de l'état excité. Enfin, nous avons modélisé la photophysique de la base de Schiff non-protonée d'un pigment UV, où une photochimie originale, et non encore documentée, a lieu, impliquant notamment un état doublement excité.

Ces études ont montré la robustesse de notre potentiel QM/MM pour modéliser une large gamme d'environnements.

Remerciements

Une thèse est un document où un seul nom apparaît à la fin, occultant par là-même les nombreuses personnes qui participent à sa réussite, directement ou indirectement. Je tiens, en premier lieu, à remercier mon directeur, Marco Garavelli, et mon co-encadrant, Ivan Rivalta, pour avoir m'avoir accepté dans leur groupe. La liberté intellectuelle qu'ils m'ont accordée, leur soutien pour ma participation à différentes conférences, ainsi que leurs conseils et leur guidage m'ont permis de mieux cerner un sujet complexe mais passionant, et ont rendu mon expérience doctorale extrêmement enrichissante. Pour leur intérêt dans mon travail, j'aimerais aussi remercier Isabelle Navizet et Chantal Daniel, qui ont accepté de rapporter ce manuscrit, et Giulio Cerullo, David Loffreda et Adèle Laurent, qui ont participé au jury. Merci aussi à Antonio Monari pour sa participation dans les comités de suivi de thèse.

L'ensemble des résultats présentés dans ce manuscrit est issus avant tout d'un travail d'équipe. Je remercie donc l'ensemble du Laboratoire de Chimie de l'ENS de Lyon, et notamment Chantal Andraud, sa directrice, ainsi que l'axe de Chimie Théorique, animé par Carine Michel et David Loffreda. En particulier, je suis infiniment reconnaissant à Élise Dumont pour son accueil au sein du laboratoire, ainsi que pour son encadrement et ses conseils tout au long de la thèse, sur les plans scientifiques et professionnels. Je remercie aussi Tanguy Le Bahers pour son implication dans le projet sur les descripteurs de densité électronique, et pour toute l'aide fournie dans l'utilisation de Gaussian, ainsi que pour l'ensemble des discussions, scientifiques ou non, de ces quatre années. Je remercie chaleureusement toute l'équipe de Bologne, en particulier Francesco Segatta, Julieta Pepino, Irene Conti and Mohsen Taha El Tahawy, pour leur accueil au sein du laboratoire, et qui ont fait de ces deux mois passés en Italie une expérience très agréable. Merci aussi à Margherita Maiuri, de Milan, pour son travail expérimental sur les mimiques de rhodopsine, ainsi que pour l'aide à la réalisation de figures claires et explicatives. Je remercie aussi Salvatore Altavilla pour tout son travail sur le rétinal en solution. Durant ces quelques années, j'ai pu aussi participer à l'encadrement de stagiaires. Merci donc à Andrea Bonvicini, Léa Rochet, et Giulia Biffi, pour m'avoir forcé à clarifier mes pensées sur certains points obscurs de la chimie théorique. La

manipulation de codes de chimie quantiques n'est pas chose aisée, surtout quand un certain nombre d'entre eux sont intriqués à travers un code "fait maison". Ainsi, aucun résultat n'aurait pu être obtenu ou interprété sans l'implication d'Artur Nenov (de Bologne) et d'Oliver Weingart (de Düsseldorf) dans le développement de Cobram. Je leur adresse mes plus profonds remerciements pour leur disponibilité et leur aide sur les points les plus techniques de l'utilisation des différents codes manipulés, ainsi que sur les aspects fondamentaux des méthodes de chimie quantique. J'accorde également toute ma gratitude à l'équipe du PSMN, et particulièrement à Loïs Taulelle, pour sa disponibilité et tout son travail dans l'installation et la maintenance de nos clusters de calcul, et profite de ces lignes pour demander son pardon pour mon occupation du $HOME...

Ces quelques années ont aussi été l'occasion pour moi de m'essayer à l'enseignement. Je remercie donc le Département de Chimie de l'ENS de Lyon et son directeur, Cyrille Monneraud, pour m'avoir donné cette opportunité. Je suis très reconnaissant à l'ensemble des étudiants encadrés durant trois ans, qui m'ont donné l'occasion de reformuler, d'expliquer, et finalement de mieux comprendre certains aspects de mon travail. Merci aussi à eux pour toutes les discussions plus informelles sur les sciences, la recherche, ou autre, que nous avons pu avoir.

L'environnement joue un rôle primordial dans le déroulement de la thèse, et je remercie l'ensemble des thésards et des étudiants du laboratoire de Chimie pour la bonne ambiance, pour l'organisation des Vendredi'nkasi, des barbecues de fin d'année, des repas de Noël et autres événements festifs. Merci en particulier à Maelle Mosser, Romain Réocreux, Laure-Lise Chapellet, Margaux Galland, Corentin Gondrand, et aux nombreux autres. Merci aussi à mon co-bureau, Thomas, pour les discussions sur le sens de la vie et des sciences.

Sortir du laboratoire et se changer les idées a parfois été vital ces dernières années. Je remercie donc tous les groupes d'amis, rencontrés au cours de ces sept (!!!) années passées à l'ENS et qui se recoupent certainement, qui ont rendu cela possible. En particulier, merci d'abord aux volleyeurs : les jeunes (Corentin, Maëlan, Émilien, Julien...), les moins jeunes (Antoine, Chris 100%, Victor "le nettoyeur", Séb, Thibaut, Benti...) et les beaucoup moins jeunes (Neil, Baptiste, Oliv', Gaby, Léo, Greg), aux volleyeuses (Céleste, Chloé, Delphine, Marie, Elsa, Morgane, Margaux^2, Léa, Morgane 2, Anna, Sylvie...), et bien sûr à notre coach préféré, Richard. Les années passées en tant qu'étudiant à l'ENS ont aussi été l'occasion de participer à l'aventure associative, et je remercie bien évidemment l'ensemble des gens que j'ai pu y

côtoyer. Merci ensuite à notre fabuleux groupe de musique, Émilie Jaunie, pour les après-midi de répèt' si productives. J'aimerais aussi remercier les divers groupes de débiles pour tous ces moments où on a pleuré de rire tous ensemble, au Foyer, dans des bungalows, ou dans un certain bar près de la place des Terreaux. Merci donc à Groliv', Alex, Pinouz, Loulou, Antoine, Barbec', Raph', Ronan, Ralbol, Sylvain, Quentin, et tous les autres qui j'espèrent, se reconnaîtront. Ici, j'accorde une place particulière à Benji, pour toutes les pauses cafés, les bières, le schnaps et le punk. Merci aussi aux amis de promos, partis ou pas, mais qu'on fini toujours par retrouver conmme si rien n'avait changé : Paco, Nat', Jean-Loup, Amélie, Sarah, Élo, Maelle, Oliv', Arthur, Mathilde, Nico, Alice, Aude, Brice, Diane, Pauline... Vous citer tous prendrait certainement un chapitre entier de manuscrit, mais je vous remercie tous pour votre présence et les moments passés ensemble. J'aimerais tout particulièrement remercier Benji, Jean-Loup, Amélie, Paco et Nat', Aude et Brice, pour m'avoir supporté, aidé, recueilli au besoin, depuis toutes ces années. Merci pour votre soutien dans les (nombreux) moments de doute. Sans vous tous, cette thèse aurait certainement été bien plus difficile à mener.

Je tiens aussi à remercier ma famille, particulièrement mon père et ma mère, qui m'ont toujours soutenu dans tous mes projets. Si je suis arrivé ici, c'est en très grande partie grâce à eux et à leur aide tout au long de mes études. Merci aussi à mon frère, Rémi, pour les moments de relâche devant un bon vieux jeu vidéo.

Et enfin, merci à Cécile, pour avoir été là ces deux dernières années.

Table des Matières

1 Des processus photo-induits en biologie à la vision en couleurs et à la modélisation du rétinal **18**

1.1 Processus photo-induits en biologie 19

1.2 Vision en couleur et protéines contenant une molécule de rétinal 21

1.3 Études computationnelles sur la première étape de la vision 23

1.3.1 Modèles de rétinal en phase gaz 23

1.3.2 Modèles de paire d'ion 26

1.3.3 Photo-isomérisation dans la rhodopsin 28

1.4 Buts de l'étude et organisation du manuscrit 30

2 Modéliser des processus ultra-rapides dans des milieux biologiques et en environnement complexe **32**

2.1 Aspects théoriques de la photochimie 34

2.1.1 Équation de Schrödinger et approximation de Born-Oppenheimer 34

2.1.2 Chemin d'énergie minimale et intersections coniques 36

2.1.3 Dynamiques semi-classiques et saut de surfaces 39

2.2 Fonction d'onde électroniques 41

2.2.1 Approximation de Hartree-Fock pour résoudre le problème électronique 42

2.2.2 Interaction de configuration et fonctions d'onde multi-configurationnelles 45

2.2.3 Corrélation dynamique 48

2.3 Méthodes Mécanique Quantique / Mécanique Moléculaire (QM/MM) 54

2.3.1 Mécanique moléculaire et champs de force 55

2.3.2 Combinaison entre les approches MQ et MM 56

2.3.3 Implémentation de COBRAM 58

3 Transfert de charge intra-moléculaire photo-induit dans des mimiques du chromophore rétinal : des indices fondés sur la densité électronique aux niveaux TD-DFT et post-HF 63

4 Modulation fine des propriétés de l'état excité du chromophore rétinal en solution 65

4.1 Introduction : résultats expérimentaux 67

4.2 Création des modèles computationnels 69

4.2.1 Échantillonnage classique 69

4.2.2 Modèles MQ/MM 69

4.2.3 Détails computationnels 70

4.3 Optimizations de géométrie 71

4.3.1 Géométries en énergies d'absorption de l'état fondamental 71

4.3.2 Minima des états excités 73

4.3.3 Résumé 75

4.4 Relaxation des liaisons C−C et intersections coniques 76

4.4.1 Interpolation entre les géométries EBL et ABL 76

4.4.2 Analyse des différents minima 78

4.4.3 Chemin vers une intersection conique 84

4.5 Dynamiques semi-classiques 88

4.6 Conclusion et perspectives 90

5 Comprendre la modulation de la photophysique de mimiques de la rhodopsine par une approche combinée entre théorie et expérience 93

5.1 Introduction 95

5.1.1 Bio-ingénierie de la protéine hCRBPII 95

5.1.2 Systèmes considérés et spectroscopie pompe-sonde 97

5.2 Construction d'un modèle computationnel fiable 100

5.2.1 Modèle MQ/MM 100

5.2.2 Première tentative à partir des structures cristallographiques 102

5.2.3 Échantillonnage en dynamique classique 103

5.3 Propriétés de l'état fondamental 105

5.3.1 Géométries et absorptions verticales 106

5.3.2 Moments dipolaires et variations de charge 119
5.3.3 Effets de l'environnement 124
5.3.4 Résumé: propriétés de l'état fondamental 126
5.4 Optimisations de l'état excité 127
5.4.1 Géométries et émissions verticales 128
5.4.2 Chemins d'énergie minimale reliant les géométries EBL et ABL 131
5.4.3 Stabilités relatives et émissions verticales 132
5.4.4 Conclusion: géométries à l'état excité, et la cas particulier de M8 136
5.5 Interprétation des spectres expérimentaux 137
5.5.1 Spectres pompe-sonde 138
5.5.2 Étude des possibilités de photoisomérization par scans contraints 142
5.5.3 Une première étape vers l'interprétation des différents temps de vie d'états excités 145
5.6 Conclusion et perspectives 147

6 Des preuves pour la présence d'un Aspartate protoné dans le voisinage du rétinal dans deux opsines microbiennes 149

6.1 Introduction: des pompes à protons activées par la lumière 151
6.2 Détails computationnels 155
6.2.1 Modèles considérés 155
6.2.2 Modèles MQ/MM 156
6.3 Affinage du modèle de protonation 157
6.3.1 Protonation standard 157
6.3.2 Influence de certains amino-acides particuliers 161
6.3.3 Analyse de modèles avec Asp212 protoné 163
6.4 Caractéristiques de l'état excité 167
6.4.1 Relaxation de l'état excité 167
6.4.2 Dynamiques semi-classiques 170
6.4.3 Conclusion 174
6.5 Conclusion et perspectives 174

7 Vision dans l'ultraviolet : propriétés photophysiques de la base de Schiff non protonée du rétinal dans le pigment cône du hamster sibérien 176

8 Conclusion et perspectives 178

9 Conclusion et perspectives (version française) 182

List of Figures

1.1 Schematic view of several photo-induced processes. 20
1.2 Mechanisms of vision and photo-isomerization of retinal. 22
1.3 Computational results for retinal models in gas phase. 25
1.4 Computational results for the counterion effect. 27
1.5 Photoisomerization in rhodopsin. 30

2.1 Topologies of conical intersections. 37
2.2 Comparison between minimum energy path and dynamics. 38
2.3 Comparison between the CASSCF, RASSCF and GASSCF approaches. 48
2.4 Partition scheme used in COBRAM. 60
2.5 Boundary region and charge redistribution. 61

4.1 Structures of PSB and 10Me-PSB. 67
4.2 Experimental pump-probe results for PSB and 10Me-PSB. 68
4.3 Partitioning of the full system for PSB and 10Me-PSB. 70
4.4 Geometries of the optimized ground state geometries of PSB and 10Me-PSB. 71
4.5 Excited state of 10Me-PSB. 74
4.6 Excited state of PSB. 75
4.7 CASPT2 ordering of states in several excited state geometries of PSB. 76
4.8 MEP for skeletal relaxation in 10Me-PSB. 77
4.9 MEP scan for the skeletal relaxation for PSB. 78
4.10 Summary of the MEP of skeletal relaxations in PSB and 10Me-PSB. 80
4.11 Mulliken charges in PSB and 10Me-PSB. 81
4.12 Qualitative change in Mulliken charges along the skeletal relaxation. 82
4.13 MS-CASPT2 results. 83
4.14 Photoisomerization of 10Me-PSB. 85

4.15 Photoisomerization of 10Me-PSB. 86
4.16 Comparison of the conical intersections in PSB and 10Me-PSB. 87
4.17 Optimized conical intersections of PSB. 87
4.18 Optimized conical intersections of 10Me-PSB. 88
4.19 Semi-classical dynamics in PSB and 10Me-PSB. 89
4.20 Structural time-evolution in 10Me-PSB. 89
4.21 Qualitative summary of the conclusions of the study. 92

5.1 Mutations of the CRABPII protein. 96
5.2 Colors and absorption spectra of hCRBPII mutants. 97
5.3 Systems studied: M4, M8 and M10 structures. 98
5.4 Pump-probe principle and example. 99
5.5 Experimental pump-probe spectra for M4, M8 and M10. 100
5.6 QM/MM schemes. 101
5.7 Secondary structures of Rhodopsin and hCRBPII. 104
5.8 Classical dynamics setup. 105
5.9 Ground state geometries. 106
5.10 Comparison between M8 and M10. 108
5.11 Vertical absorptions in M4. 109
5.12 Vertical absorptions in M8. 113
5.13 Vertical absorptions in M10. 116
5.14 Permanent dipole moments along the trajectory. 120
5.15 Difference in dipole moment along the trajectory. 121
5.16 Averaged Mulliken charges. 123
5.17 Average charges per atom for the S_0, S_1 and S_2 states for M4, M8 and M10. 124
5.18 Electrostatic effect computed for the M8 protein. 126
5.19 Excited state geometries of M4. 128
5.20 Excited state geometries of M10. 129
5.21 Excited state geometries of M8. 130
5.22 MEP for skeletal relaxation for EBL-optimized samples. 132
5.23 MEP for skeletal relaxation for ABL-optimized samples. 133
5.24 Interpretation of the experimental pump-probe data for M4. 139

5.25 Interpretation of the experimental pump-probe data for M8. 140
5.26 Interpretation of the experimental pump-probe data for M10. 141
5.27 Constrained scans around $C_{13}-C_{14}$ in M4. 143
5.28 Constrained scans around $C_{11}-C_{12}$ in M8. 144
5.29 Constrained scans around $C_{13}-C_{14}$ in M8. 144
5.30 Constrained scans around $C_{11}-C_{12}$ in M10. 145
5.31 Tentative interpretation of the different lifetimes of each proteins. 146

6.1 Structure of bacteriorhodopsin. 152
6.2 Structure of xanthorhodopsin. 154
6.3 Ground state geometries for **model1**. 157
6.4 Excited state manifolds in the ground state geometry of **model1**. 158
6.5 Different models of bacteriorhodopsin. 159
6.6 Protonated setups of **model1**. 164
6.7 Ground state geometry of the protonated setups of **model1**. 166
6.8 Geometries of the excited state of the protonated setups of **model1**. 168
6.9 Geometries obtained for optimizations constrained to **O1-md**. 169
6.10 Geometries obtained for optimizations constrained to **O2-md**. 170
6.11 Dynamics of **O1-cry** and **O1-md**. 172
6.12 Dynamics of **O2-cry** and **O2-md**. 173

List of Tables

4.1 Summary of ground state transition energies. λ_{max} alues are reported in nm. Only the configurations that contribute for more than 0.095% to the wavefunction are presented. 73

4.2 Difference in dipolar moment for the different setups 73

4.3 Methylated retinal: CASSCF wavefunction character along the MEP scan. Only the CSFs with a contribution higher than 0.095 are shown. 77

4.4 Native retinal: CASSCF wavefunction character along the MEP scan. Only the CSFs with a contribution higher than 0.095 are shown. 79

4.5 Mulliken charges summed up on the two sides of the molecule for 10Me-PSB and PSB, for the different geometries computed. 80

4.6 ESA computed at the EBL and ABL geometries for PSB and 10Me-PSB. SA20-CASSCF wavefunctions were used, and vertical emission are indicated at the CASPT2/SA20-CASSCF level. The two most intense ESA signals obtained are written, and the wavelengths are printed in nm. In brackets are printed the corresponding oscillator strengths. 83

5.1 SS- and MS-CASPT2 energies (in eV) computed on top of CASSCF optimized geometries in M4, M8 and M10, and difference with respect to experimental data. 103

5.2 Average geometries for the three proteins: bond length alternation (left); distortions (right).107

5.3 CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M4 protein. 110

5.4 PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at the MS-CASPT2 level, obtained for the ten snapshots for M4 protein. 112

5.5 CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M8 protein. 114

5.6 PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at MS-CASPT2 level, obtained for the ten snapshots for M8 protein. 115

5.7 CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M10 protein. 117
5.8 PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at MS-CASPT2 level, obtained for the ten snapshots for M10 protein. 118
5.9 Average vertical absorptions obtained for each protein, at the CASPT2 and MS-CASPT2 levels. The values indicated for M8 are those obtained out of the matching set only. . . . 119
5.10 Dipole moment comparison between M8 "matching", M8 "non-matching", and M10 dipole moments, averaged over the ten snapshots. 122
5.11 Vertical absorptions obtained *in vacuo*, on top of QM/MM optimized geometries, averaged over the ten snapshots for each proteins. Results in protein are included for comparison (6-31G* basis set). 125
5.12 Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M4 protein. 134
5.13 Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M10 protein. 135
5.14 Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M8 protein. Bold values indicate samples that were optimized in ABL geometry. 136

6.1 CASSCF wavefunctions and CASPT2 energies and oscillator strengths obtained for the different models of bR considered. Energies are computed in eV. 160
6.2 CASSCF wavefunctions and CASPT2 energies and oscillator strengths obtained for the different models of xR considered. Energies are computed in eV. 161
6.3 CASSCF wavefunctions, and CASPT2 energies, obtained by switching off selected amino-acids in the quadrupole. 162
6.4 CASSCF wavefunctions and CASPT2 energies computed for the protonated models of bR **model1**. 165

Abbreviations used

- 10Me-PSB: 10-methylated protonated Schiff base
- ABL: Alternated bond lengths
- Arg: Arginine
- Asp: Aspartate
- BLA: Bond lengths alternation
- bR: Bacteriorhodopsin
- CASPT2: Complete Active Space Perturbation Theory at the Second Order
- CASSCF: Complete Active Space Self-Consistent Field
- CCW: Counter clockwise
- CI: Conical intersectionp
- CIS: Configuration interaction singles
- CRABPII: cellular retinoic acid binding protein type II
- CSF: Configuration state function
- CW: Clockwise
- DFT: Density functional theory
- EBL: Even bond lengths
- ES: Excited state
- ESA: Excited state absorption
- FC: Franck-Condon
- FS: Fluorescent state
- GAFF: Generalized Amber forcefield
- GASSCF: Generalized Active Space Self-Consistent Field
- Gln: Glutamine

- Glu: Glutamc acid
- GPCR: G protein-coupled receptor
- GS: Ground state
- GSB: Ground state bleaching
- HBN: Hydrogen bonding network
- hCRBPII: human celullar retinol-binding protein type II
- HF: Hartree-Fock
- His: Histidine
- IPEA: Ionization potential - Electron affinity
- Lys: Lysine
- MCSCF: Multi-Configurationnal Self-Consistent Field
- MD: Molecular dynamics
- MEP: Minimum energy path
- MM: Molecular Mechanics
- MP2: Møller-Plesset 2
- MS-CASPT2: Multi-State CASPT2
- PA: Photoabsorption
- PES: Potential energy surface
- PM-CASSCF: Perturbatively mofified CASSCF
- PSB: Protonated Schiff base
- QM: Quantum mechanics
- RASSCF: Restricted Active Space Self-Consistent Field
- Rh: Rhodopsin
- SA-CASSCF: State Averaged-Complete Active Space Self-Consistent Field
- SE: Stimulated emission
- SHUV: Siberian hamster ultraviolet
- SOC: Spin-orbit coupling
- TD-DFT: Time-dependent density functional theory
- TS: Transition state
- xR: Xanthorhodopsin

1 From biological photo-induced processes to color vision and retinal modeling

Résumé Dans ce chapitre introductif, nous commençons par présenter les différents processus déclenchés par une excitation lumineuse sur un chromophore, principalement dans des systèmes d'intérêt biologique. On s'intéresse particulièrement à l'évolution de l'état excité photo-induit, en séparant les processus radiatifs et non-radiatifs, avec notamment les intersections coniques. En effet, ces intersections coniques jouent un rôle primordial dans le processus de vision en couleur chez les vertébrés. Nous présentons ensuite plus en détail ce mécanisme de vision, et nous introduisons le chromophore rétinal, ainsi que la famille des opsines, qui sont des protéines qui peuvent lier et intégrer ce chromophore sous forme de base de Schiff protonée. Sous l'effet d'un rayonnement lumineux, le rétinal est isomérisé par rotation autour d'une des liaisons doubles. Cette famille de protéines, et ce mécanisme, se retrouve dans de nombreux organismes, au cœur de processus vitaux comme la régulation du potentiel électrostatique entre une cellule et le milieu externe, la transmission du signal lumineux, ou encore le mouvement photo-induit. Cette diversité justifie les nombreuses études, expérimentales et théoriques, sur le chromophore rétinal.

Nous nous penchons ensuite sur l'état de l'art en matière de modélisation du rétinal, en commençant par les modèles minimaux en phase gaz, puis en décrivant des modèles plus réalistes qui prennent en compte la présence d'un contre-ion, et enfin en étudiant des modèles de protéines complets en QM/MM. Pour chaque niveau de modélisation, nous présentons les résultats déjà publiés, en termes d'étude statique de surfaces d'énergie potentielle, et en terme de dynamique de photo-isomérisation.

Nous finissons par présenter les buts de la thèse, et l'organisation générale du manuscrit.

Interactions between light and matter are at the core of several crucial biological processes, such as photosynthesis, where the light energy is converted into chemical energy for oxidizing water molecules [1], or color vision, where light energy is converted into an electric signal that is conveyed in the optical nerve [2]. On the other side, some systems are also designed to avoid light induced reactions, such as DNA or RNA, where UV light can induce cyclizations between nucleobases that prevent the correct replication of the molecule. Collecting, storing and dissipating light-energy require highly evolved molecular mechanisms that can involve dozens of different chromophores and proteins. For instance, the photosystem II, which is the primary site for photosynthesis, contains a complex system of antennae (containing carotene molecules) that ensures a very effective collection of light at several wavelength, and transports the light-induced electronic excitation to the reaction center. The common point between all light-induced processes is that the reactions involved are all occuring on the excited state manifolds of the chromophores.

1.1 Photo-induced processes in biology

A summary of some possible photo-induced processes is displayed on figure 1.1. Upon absorption of light (photoabsorption, PA), a molecule is promoted from its ground state, usually a singlet for an organic molecule, S_0 (point **A** in figure 1.1), where the nuclei and the electrons are at equilibrium at a given geometry, to an excited state S_n (**B**) where this equilibrium is broken. The corresponding point of the S_n surface is called the Franck-Condon (FC) point, and corresponds to the response of the electronic cloud to the light-induced perturbation. Out of this particular geometrical configuration, the system has to relax from this non-equilibrated situation to a new equilibrium, involving the concomitant relaxation of nuclei and electrons. Thus, the geometry can be affected, and several processes , either radiative (*i.e.*, with emission of a photon) or non-radiative, can take place.

First, the promoted wavepacket can relax to a local minimum on the excited state surface (**C**), where it relaxes to the ground state through the emission of a photon. This process is called fluorescence, and occurs on the nano- to microsecond timescale. It is also possible that at some point of the evolution, a crossing with another energy surface corresponding to a different electronic spin occur (**D**), triggering a population of this new spin state provided that the transition is favored. Characterizing this population exchange is done by computing the spin-orbit coupling (SOC) between the two states. Following the path which lowers the global energy, the system can then relax to the ground state by emission of a

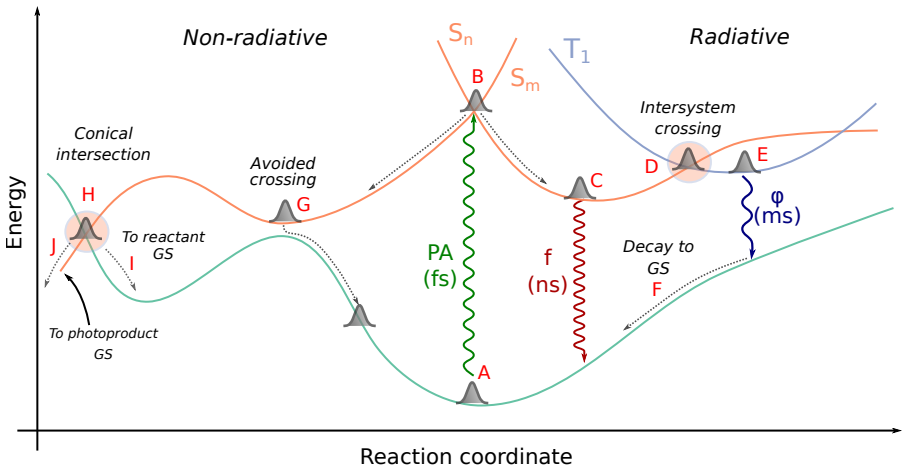


Figure 1.1: Schematic view of several photo-induced processes. Capital letters refers to particular phenomena and are described in the text.

photon (**E**). However, as this radiative process is spin-forbidden, it occurs on a much longer timescale than fluorescence. This phenomenon is called phosphorescence, and occurs on the millisecond timescale.

Non-radiative decays occur when no photon is emitted for decaying back to the ground state. This occur when the two potential energy surfaces come close to one another. Two topologies are possible in this case. In the first, the two surfaces "avoid" one another (point **G**). If the wavepacket can stay trapped long enough at this point, a transition to the lower energy state is likely to occur, and can be characterized by Landau-Zener type probabilities. Another possibility is that out of the FC region, the relaxation on the excited state surface leads to a geometry where the energies of both the excited state and the ground state are the same. This particular topology of both surfaces is called a conical intersection (CI) (**H**). Out of this point, the system can decay back to the initial geometry(**I**), or follow a further relaxation that leads to a different ground state geometry (**J**). In the latter case, a photo-product is formed, and the speed of this process is completely determined by the topology of the excited state surface [3]. For instance, a barrierless process on the excited state surface leads to ultra-fast formation of the photo-product, usually on a sub-picosecond timescale. Conical intersections have a critical importance in a wide variety of biological processes [4], such as light-harvesting [5], charge and proton transfer [6, 7], and also in atmospheric chemistry.

1.2 Color vision and retinal-containing proteins

Conical intersections play a primary role in the process of vision [8]. In the following, we will focus on human vision. When light penetrates the eye, it is collected in the retina, which is paved with different types of photoreceptors — the cones and the rods (figure 1.2). The cones are highly sensitive pigments that are used in bright light conditions and can absorb at different wavelengths — 420, 530 and 560 nm for the blue (short wavelength, **S**), the green (medium wavelength, **M**) and red (long wavelength, **L**) pigments — while rods are used in dim light condition, and absorb light only in the blue, at 440 nm. Despite their different absorption properties, cones and rods have a similar structure. In particular, their light-collecting part is composed of a folded membrane which embed opsin proteins (named photopsins in cones and rhodopsins in rods). This family of G protein-coupled receptors (GPCR) are trans-membrane proteins formed by seven α-helices that form a cavity, where a retinal molecule is covalently bounded to a nearby lysine, forming a protonated Schiff base — PSB.

This chromophore is composed of six double C=C bonds that separate a β-ionone ring and the terminal iminium. In visual pigments, the retinal is found in its 11-*cis* form in the dark-adapted state (figure 1.2). The model system for visual pigments is the rhodopsin protein, which is the only one for which the crystal structure has been resolved, in the dark (11-*cis*) and the (all-*trans*) photo-intermediates. In the following, we will focus on the behaviour of rhodopsins.

Upon light absorption, the chromophore undergoes an ultra-fast isomerization leading to its all-*trans* isomer [9]. During this process, the excited state of the retinal follows an ultra-fast decay with a lifetime of 150 fs [10], and the photoproduct is formed in 200 fs [11] with a quantum yield of 65% [12]. This process is vibrationally coherent [13], and recent studies have proven that this process occurs at the molecular limit speed, “that of a single period of the local torsion coordinate” (Johnson et al. [14]), and is thus a tremendous basis for the design of highly effective, and selective, molecular photoswitches. This isomerization induces steric perturbations in the protein, and constitute the primary event in color vision, leading to the cascade of thermal events that ends with the transmission of the signal and the reformation of the 11-*cis* molecule [2].

Opsin protein are not only used as the primary photoreceptor for vision in higher life forms. They are indeed found in a large variety of bacteria, where they are used as light-driven ion pump to maintain the electrostatic potential across membranes, or as light sensors for phototaxis [15]. The prototype system for this class of opsin proteins is the bacteriorhodopsin, which acts as a light-driven proton pump. In bacterial opsins, the dark-state adapted retinal adopts an all-*trans* conformation and absorbs at 570 nm [16, 17].

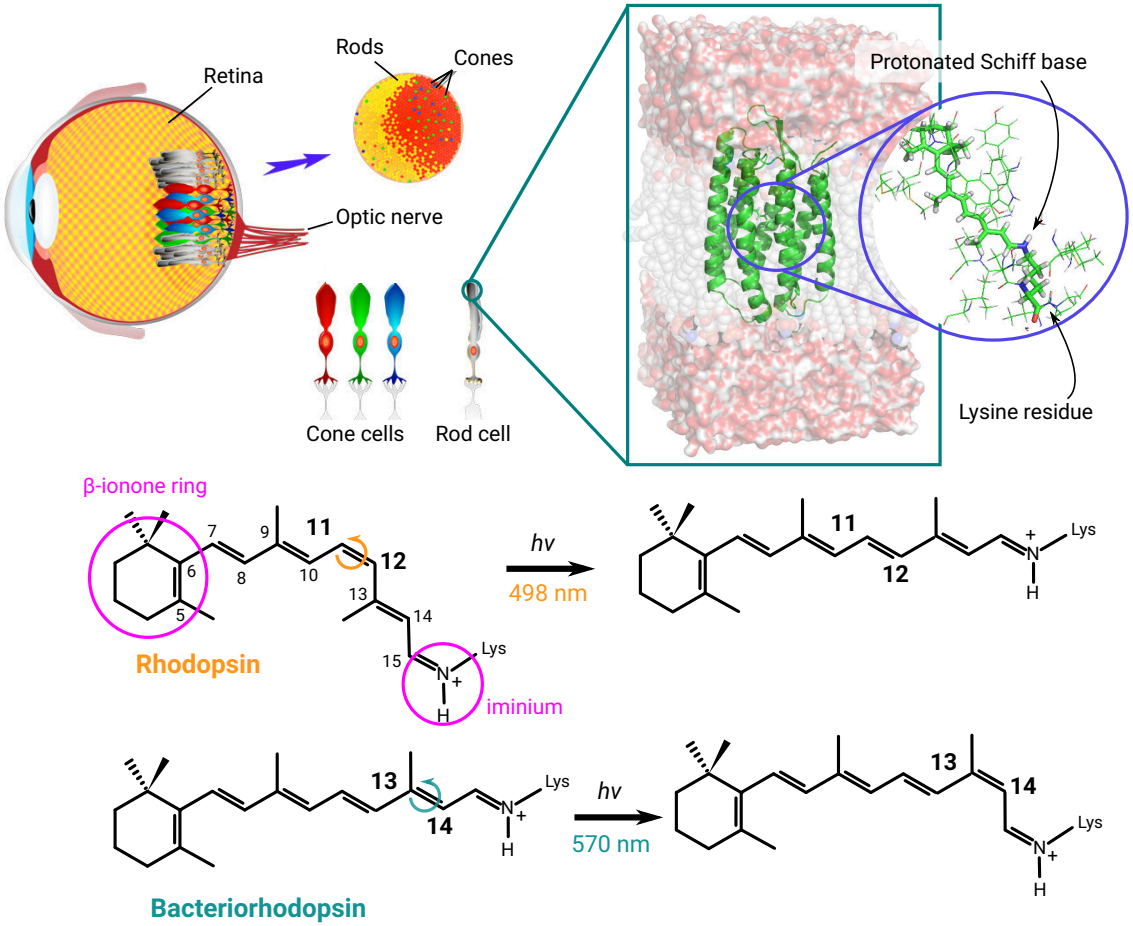


Figure 1.2: Processes of vision. Top panel: highlight of the photo-receptors that contain the photo-acitve opsin proteins. Bottom panel: photo-isomerization in rhodopsin and bacteriorhodopsin.

Upon light absorption, it undergoes an ultra-fast isomerization leading to an 13-*cis* conformation. This transformation is also ultra-fast, and is completed in around 500 fs, with an excited state lifetime decay of 200 fs and a quantum yield of 65% [18].

Although all opsin proteins share different isomers of the same chromophore, they display a variety of properties when it comes to their photophysical — vertical absorptions, excited state lifetimes — and photochemical properties [19]. For instance, retinal in gas phase absorbs light at 630 nm [20], considerably red-shifted with the value observed in most proteins. This effect is known as the *opsin shift* [21]. Moreover, the photochemistry of the retinal in methanol solution is significantly altered with respect to what happens in proteins, with considerably longer lived fluorescent state ($\approx$ 3 ps) and lower quantum yield ($\approx$ 0.20) [22]. These two facts stress the catalytic capacity of the protein environment, as well as its huge tuning effect. Thus, opsins are the "natural" reference systems for designing new

molecular photoswitches and photochromic devices [23]. Indeed, the retinal features a high tuning of its properties in a given environment, and can be reversibly switched between different absorption colors, or converted between light emitting and non-emitting states, with applications in optogenetics or optical memories [24, 25]. The relative simplicity of opsin proteins, where there is only one chromophore involved, also make them prototype systems for designing and calibrating new computational methods for modeling excited state reactivity.

1.3 Computational studies on the primary event of vision

Extensive computational studies have been carried out for unravelling the intrinsic properties of the retinal in gas phase [26–29], the interplay between the retinal and its environment [27, 30–34], and the characteristics of the photoisomerization in model proteins and solvents [8, 35–39]. Modeling photochemical processes require methods that can treat ground and excited states manifolds at an equal footing, and that can be transferred from gas-phase to molecule embedded in environments — solvents or proteins. Methods that combine geometries obtained with multi-configurational wavefunctions in the complete active space self-consistent field (CASSCF) [40] framework, and with energy corrections added by means of second order perturbation theory (CASPT2) [41], are considered reliable and appropriate for studying photochemical processes. This level of theory will be denoted as CASPT2//CASSCF (*i.e.* CASPT2 energies computed on top of CASSCF optimized geometries). In this part, we will focus on the computational studies of the chromophore found in rhodopsin protein, the 11-*cis* retinal molecule.

1.3.1 Retinal models in gas phase

Garavelli et al. have performed minimum energy path computations on model systems in gas phase, from minimal models [26] to more realistic ones [42–44]. Some of the results obtained are shown on figure 1.3. These studies have shown that the relaxation on the ionic S_1 (^1B_u-like) state surface, corresponding to a single $HOMO \rightarrow LUMO$ ($H \rightarrow L$) excitation, follows a two states / two modes (TSTM) mechanism. The driving force out of the Franck Condon region is first given by the symmetric stretching of the C=C bonds, up to a bifurcation point (**BP**), followed by the isomerization around the central bond — thus, two modes are involved — and where the S_2 state, corresponding to the double $(H \rightarrow L)^2$ was found to be high in energy and thus do not participate in the reaction — two states only are involved [45]. The computed path indicate a barrierless relaxation of S_1 that ends at a conical intersection point with a

twisted central bond ($\approx 80°$ for the twisting angle) (figure 1.3 panel a). This is accompanied by an initial charge transfer from the C–CH–$NH_2{}^+$ side to the CH_2–CH–C side of the molecule (figure 1.3 panel b), in agreement with what has been postulated from ultra-fast spectroscopy [46]. Upon relaxation on the S_1 surface, the charge is reditributed smoothly between the two sides, and results in a CI point where a net charge has been transfered. The decay to the ground state is accompanied by a back transfer of the charge on the $NH_2{}^+$ moiety. Increasing the length of the polyene chain conserves the charge transfer character [3], but results in steeper slopes and longer energy plateaus in the excited state surface due the larger flexibility of the systems, where skeletal relaxations allow for efficient dissipation of the torsional energy (figure 1.3 panel f). This is consistent with experiments that have proven the existence of of a fluorescent state (FS) [46]. Semi-classical dynamics have confirmed these conclusions, and have shown that skeletal relaxation occurs in the first 50 fs [47] (figure 1.3 panel d), followed by oscillations in the valley connecting the relaxed point to the point where the isomerization starts. Larger systems show a large number of such oscillations during the trajectory (figure 1.3 panel e), before the systems gains enough momentum to drive the *cis-trans* isomerization. Therefore, the computations in gas phase are in very good agreement with the experimental evidence of an ultra-fast photo-isomerization in visual pigments. Explicit inclusion of the β-ionone ring shows a planarization of the connection between the ring and the polyene chain, indicating that steric interactions in protein could affect the general shape of the potential energy surface [27].

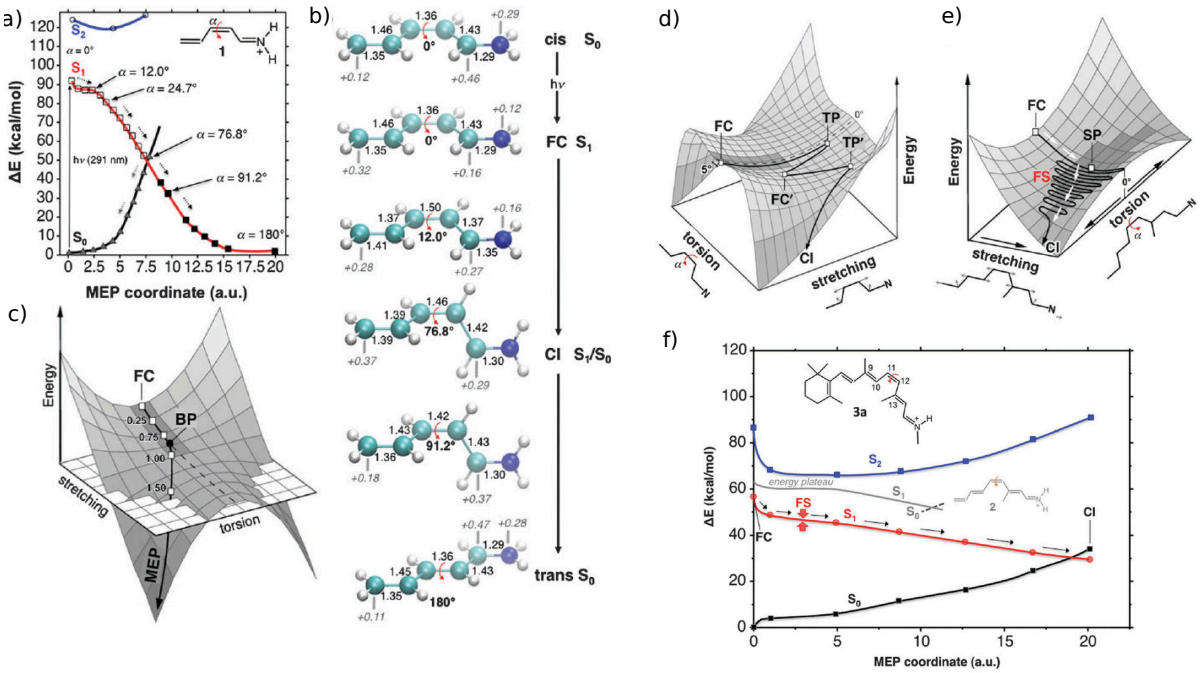


Figure 1.3: Computational studies on retinal models in gas phase. a) Minimum energy path computation on a minimal model. b) Evolution of the charge repartition on selected points of the minimum energy path. c) Representation of the two modes that lead to the photoisomerization; BP is a bifurcation point, where the skeletal relaxation has completed, and the system starts to move along the isomerization coordinate. d) Semi-classical trajectory on minimal model; FC' is the point reached after the first oscillation on the stretching coordinate. e) Schematic pathway followed by a larger system, based on Minimum Energy Path (MEP) commputations. f) MEP for realistic retinal models (plain lines); system without ionone ring is presented in dashed line; FS is the located fluorescent state. Reproduced from Ref [38].

The modeling of the intrinsic properties of the retinal allows a fine understanding of the effect of its surrounding environment, which can be thought of as perturbations of the intrinsic retinal behaviour. In particular, the interplay between a protonated Schiff base and its counterion are expected to have major effects on the properties exposed above. In particular, the different character of S_1 (ionic) with respect to S_0 and S_2 (covalent) states can lead to differential stabilization or destabilization, leading to modification of the relative order of these states [44]. As shown above, the positive charge is localized on the iminium moiety in the S_0 state, and is transfered upon excitation to the $CH_2{=}CH{-}C$ end. Thus,

a charge placed near the iminium is expected to stabilize the covalent S_0 and S_2 states, and destabilize the S_1 state, as shown on figure 1.4 panel a.

1.3.2 Models of ion pair

Computations carried out for a tight ion pair (where the protonated Schiff base and a chloride counter-anion are in close contact) in gas phase have proven that the S_1 surface is destabilized, while S_0 and S_2 are stabilized, leading to a blue-shifted absorptions, and to a wide region in of the conformational space where S_1 and S_2 are degenerated [48], as presented on figure 1.4 panel b for a model of *cis*-retinal with five double bonds. The presence of the counterion also affects the characters of S_1 and S_2, and both of them become mixed within ionic and covalent configurations (greyed region in figure 1.4 panel b). This could question the general validity of the two-modes / two-states model exposed above. However, inspection of the electronic structure of the twisted minimum obtained show that at this point, S_1 has fully recovered its ionic nature, and S_2 is found high in energy. Therefore, the two-state model remains valid [49]. The existence of an energy plateau indicates a photochemistry that would be significantly slower than what is obtained in gas phase, similar to what is observed in solution.

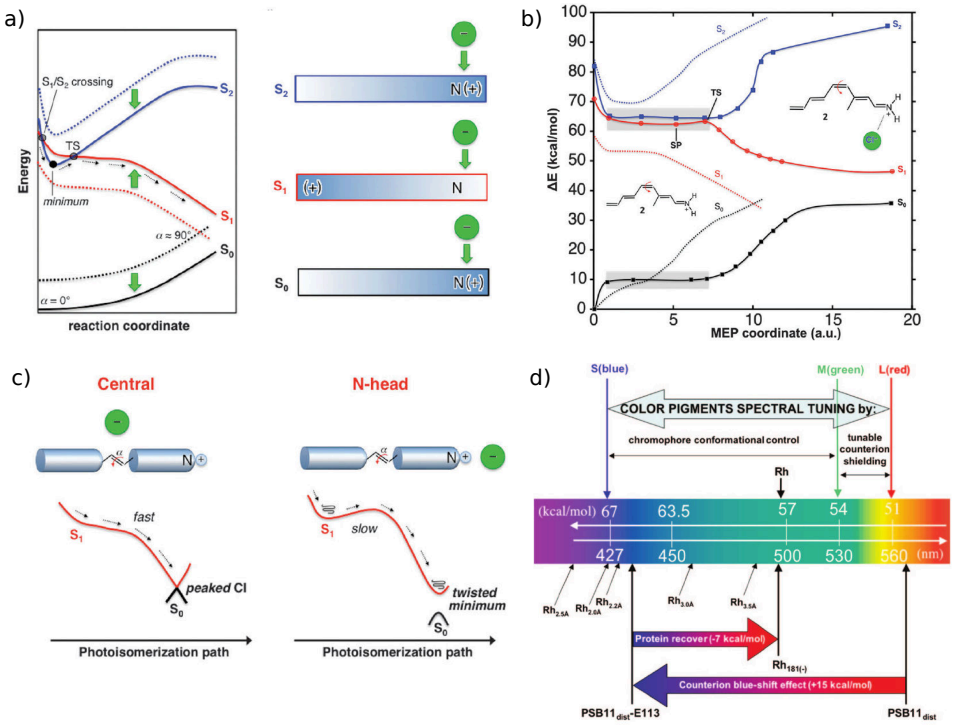


Figure 1.4: Modeling of environmental effect. a) Schematic representation of the effect of a counterion on the S_0, S_1 and S_2 potential energy surfaces. b) MEP for a tight ion pair, with the greyed surface correspond to region with a mixing of ionic and covalent characters for S_1 and S_2; in dashed line, the corresponding MEP without the counterion. c) Control of the photoisomerization by the position of the counterion. d) Schematic representation of the color tuning mechanisms in natural visual pigments. Reproduced from Ref [38].

The specific position of the counterion along the polyenic chain can also open different photoisomerization reactions. As shown in figure 1.4 (panel c), a charge placed in the vicinity of the isomerizing bond allows for the recovery of the peaked CI already computed in gas phase, while a tight ion pair leads to a twisted minimum [49]. Thus, controlling the position of the charge, *i.e.* controlling the shape of the electrostatic potential around the retinal, allows a fine control on absorption properties as well as on excited state nature, and even on the selectivity of the isomerization. These results suggest that the visual pigments have been evolutationary designed for finely tuning the properties of the retinal chromophore. However, the counterion quenching is not the only factor that can affect the absorption wavelength of the retinal. Indeed, the blue pigment is even more blue-shifted that the tight ion pair, indicating a very ineffective shielding of the protein, which in turn leads to a very inefficient photochemistry. Thus,

other mechanisms are involved, such as chromophore twisting, or β-ionone ring flexibility [21, 50]. In particular, a strong twisting between the β-ionone ring and the polyene chain leads to blue-shifted absorption.

However, retinal embedded in proteins are usually found in tight ion pairs with a carboxylate counterion [51, 52], with another carboxylate in close proximity with the isomerizing bond. This last group has been proposed as a "selector" for the site of isomerization [53], considering the results presented in figure 1.4 panel c. However, a tight ion pair is expected to lead to a very ineffective photochemistry due to the final twisted minimum structure obtained, and does not correlate well with the ultra-fast process occuring in proteins. This may be explained by noticing that in protein, the tight ion pair is itself embedded inside an electrostatic environment, and the effect of the ion pair itself can be quenched by the ensemble of other residues. Evaluating the protein shielding in rhodopsin has been done by coupling the CASPT2//CASSCF approach (Quantum mechanics, QM) highlighted above to a classical molecular mechanics (MM) description in the so-called QM/MM framework [31, 54], and it has been proven that the protein environment is responsible for shielding around 50% of the counterion effect [50]. This is especially due to fragments in the retinal binding pocket (around 3.5 Å from the retinal), and appears necessary for recovering an ultra-fast process from a tight ion-pair. The absorption of the distorted retinal model in gas phase and in tight ion pair is found at 435 and 570 nm respectively, and agrees nicely with the natural absorptions of visual pigments (425, 530 and 560 nm for the blue, green and red pigments respectively). As the spectral tuning of the visual pigments cannot be entirely explained from a quenching perspective, it has been proposed that the tuning from the blue to the green pigment is due to deconjugations of the polyene chain, with torsions of the β-ionone ring and the chain itself, while the tuning from the green to the red pigment would be due to more and more shielding of the counterion by the protein (figure 1.4 panel d).

1.3.3 Photoisomerization in rhodopsin

Having modeled the intrinsic properties of retinal models in gas phase, and the effects of the the presence of a counterion, the next step has been the modeling of the photochemistry of the retinal embedded in the protein, where the chromophore is highly distorted in the dark-adapted state. *Ab initio* molecular dynamics carried out on a model distorted chromophore have highlighted that the steric induced distorsion leads to ultra-fast and effective photoisomerization process, with an aborted double-pedal movement involving a clockwise rotation around the $C_{11}-C_{12}$ bond and counter-clockwise motions of

the C_9-C_{10} bond [55]. However, this last motion is aborted when the molecule returns to the ground state, and the reaction leads to an all-*trans* product (a full bicycle pedal movement would lead to the 9-*cis* product). This space saving mechanism, that minimizes the steric interactions with the rest of the protein (figure 1.5 panel b), is fully consistent with the earlier computational studies of Warshel et al [56, 57], where semi-empirical QM/MM potentials have been used [58]. *Ab initio* minimum energy paths, computed at the QM/MM level with electrostatic embedding [59] — accounting for the electrostatic interactions at the QM level — are consistent with the two-modes / two-states mechanism found in gas phase, and confirm the space saving mechanism with a clockwise motion around $C_{11}-C_{12}$ simultaneous with a smaller motion around C_9-C_{10} [50, 60]. This motion is accompanied by fast out of plane vibrational motion of hydrogen atoms in the vicinity of the double bond involved (HOOP mode) [61], which has been confirmed by femtosecond stimulated Raman spectroscopy (FSRS) [62], and helps the pyramidalization of the carbons to drive the isomerization reaction. Another finding was that the outcome of a trajectory is highly dependent on the moment of the hopping event. Indeed, trajectories that hop directly when reaching the CI lead to the *trans* product, while trajectories that "loop" in the excited state end up with only 50% probability of ending to a *trans*-product [61]. Eventually, ultra-fast pump-probe experiments combined with scaled CASSCF/AMBER dynamics have given evidence for a conical intersection in rhodopsin, reached 80 fs after excitation [8]. This full dynamics in protein has also shown the evidence for the space-saving bicycle pedal movement inside the protein cavity. The quantitative agreement with experimental data (see figure 1.5 panel a) obtained out of the QM/MM computation shows the relevance of this approach for modeling the behaviour of the excited state of the retinal in a protein environment.

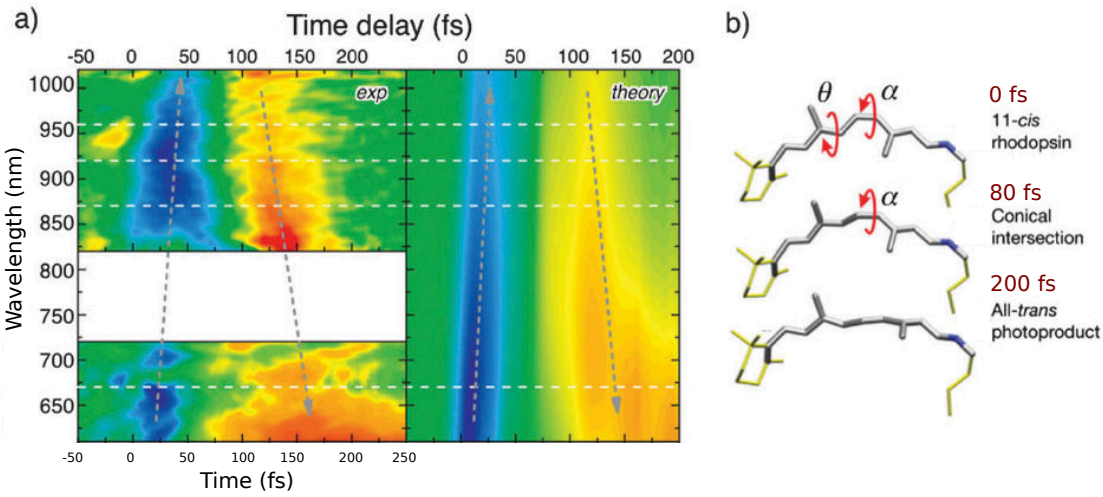


Figure 1.5: Dynamics of the photo-isomerization of the retinal in rhodopsin. a) Comparison of experimental pump-probe spectroscopy (left) and computed data (right) in the near IR region. b) Selected structures of the retinal during the isomerization, illustrating the double-pedal movement: Starting structure, conical intersection and all-*trans* product. Adapted from Ref [8].

1.4 Goals of the study and organization of the manuscript

The CASPT2//CASSCF/AMBER approach has proven itself to be a very effective tool for investigating on the primary photoevents in rhodopsin, for static minimum energy paths computations as well as semi-classical trajectories. It relies on an unbiased description of the excited state manifolds, and on explicit computation of electrostatic interactions at the QM level. However, the counterion effect, and its possible shielding by the protein or solvent environment, shows the crucial importance of the set of point charges that are included in the model. This concerns primarily the protonated states of amino-acids in the vicinity of the chromophore, but also their particular conformation, or the arrangement of the solvent molecules around the chromophore. Thus, every system requires a careful parametrization in order to get significant results out of the computational models. The aim of this thesis is to create QM/MM models for the retinal embedded in several complex environments, e.g. environment that are hard or impossible to characterize experimentally, due to their flexibility (solvent), or to the difficulty to obtain atomic-resolution information (protonation states in proteins), in order to model the photochemical and photophysical properties of the chromophore in these environment. In conjugation with available

experimental data, such as ultra-fast pump-probe spectroscopy, our computations will be used to complete our knowledge on the various effects imposed by the environment on a reaction that develops entirely in an excited potential energy surface. This manuscript is organized in six chapters. The first presents the theoretical aspects of photochemistry, especially the challenges for modeling reactions that goes through conical intersections, as well as the CASSCF and CASPT2 methods, and the particularities of the QM/MM COBRAM interface that we will use in all investigations. The second part is a recently published comparison of the intramolecular photo-induced charge transfer predicted by several functionals in the TD-DFT framework and by several post-HF techniques. The third part presents a computational study on the effects of the methylation of a protonated Schiff base in solution, in terms of minimum energy paths and dynamical behaviour. In the fourth part, we have modeled rhodopsin mimics, where point mutations in the retinal binding pockets experimentally lead to systems that can absorb in the wide visible range. The fifth part presents evidence for a new protonation model for Asp212 in the vicinity of the chromophore in bacteriorhodospin. The last chapter deals with a recently published work on the modeling of the unprotonated Schiff base of the Siberian hamster cone pigment, an important visual pigment containing an unprotonated Schiff base in the dark state and for which X-ray structure is not available.

2 Modeling ultra-fast processes in biological media and complex environment

Abstract In this chapter we will highlight the computational methods that will be used in the following parts. One of the main problems we have to deal with in photochemistry is that chemical reactions can involve several potential energy surfaces. The interplay between these surfaces creates regions where the usual approximations of quantum chemistry are not valid anymore. In particular, it has been shown that the ultra-fast decay of the retinal excited state in gas phase and in protein is due to a conical intersection between the ground and the first excited state. Therefore, we need methods that are able to treat in a balanced way all the relevant electronic states. Furthermore, as the properties of retinal are highly tuned by its environment, it is of utmost importance that the description of the environment is explicitly included in our models.

We will first present the main aspects of theoretical chemistry, and in particular the Born-Oppenheimer approximation and its limits. Then, we will describe the methods we use for computing the electronic potential energy surfaces. Eventually, our QM/MM setup, which allows for an accurate description of the external environment of the retinal, will be highlighted.

Résumé Dans ce chapitre, nous présentons les méthodes computationnelles utilisées dans le reste du manuscrit. En photochimie, l'un des principaux problèmes est lié au fait que les réactions étudiées peuvent faire intervenir plusieurs surfaces d'énergie potentielle. Les interactions entre ces différentes surfaces créent des régions dans lesquelles les approximations habituelles de la chimie quantique ne sont plus valables. En particulier, on a montré que le temps de vie ultra-court de l'état excité du rétinal en phase gaz et dans la protéine est dû à une intersection conique entre cet état excité et l'état fondamental. Par conséquent, nous avons besoin de méthodes qui sont capables de traiter sur un pied d'égalité tous les états excités impliqués. De plus, comme les propriétés du rétinal sont modulées par son environnement, il est important d'inclure explicitement la description de cet environnement dans nos modèles.

Nous présenterons d'abord les aspects principaux de la chimie théorique, et en particulier l'approximation de Born-Oppenheimer et ses limites. Nous décrirons ensuite les méthodes utilisées pour calculer les surfaces d'énergie potentielle électronique. Enfin, nous présenterons notre potentiel QM/MM, qui permet une description précise de l'environnement extérieur du rétinal.

2.1 Theoretical aspects of photochemistry

The behaviour of a quantum system is determined by the time-dependent Schrödinger equation, which is usually solved using the Born-Oppenheimer approximation. However, the validity of this approximation is not guaranteed in regions of conformational space where electronic potential energy surfaces become too close. This will be demonstrated in the first section. Next, two computational processes will be presented. The first is a static point of view and deals with the topologies of the electronic potential energy surfaces. We will also present a method for optimizing the geometries of conical intersection. The second is a molecular dynamics approach, which include a switching algorithm to allow the propagation of nuclei wavepackets on different electronic surfaces.

2.1.1 Schrödinger equation and Born-Oppenheimer approximation

A polyatomic system, containing n electrons at position $\vec{r}$ and M nuclei at positions $\vec{R}$, and with the hamiltonian $\hat{\mathscr{H}}$ is completely defined by its wavefunction $\Psi(\vec{r},\vec{R})$, which is in turn obtained by solving the time-dependent Schrödinger equation:

$$i\hbar\frac{d\Psi}{dt}(\vec{r},\vec{R}) = \hat{\mathscr{H}}(\vec{r},\vec{R})\Psi(\vec{r},\vec{R}) \tag{2.1}$$

The hamiltonian can be decomposed in the kinetic energy of the nuclei $\hat{\mathcal{T}}_N(\vec{R})$, the electrostatic potential between nuclei $\hat{\mathcal{V}}_{NN}(\vec{R})$, the kinetic energy of the electrons $\hat{\mathcal{T}}_e(\vec{r})$, the electrostatic potential between electrons and nuclei $\hat{\mathcal{V}}_{Ne}(\vec{R},\vec{r})$ and the electrostatic potential between electrons $\hat{\mathcal{V}}_{ee}(\vec{r})$, and we can thus write:

$$\begin{aligned}\hat{\mathscr{H}}(\vec{r},\vec{R}) &= \hat{\mathcal{T}}_N(\vec{R}) + \hat{\mathcal{V}}_{NN}(\vec{R}) + \hat{\mathcal{T}}_e(\vec{r}) + \hat{\mathcal{V}}_{Ne}(\vec{R},\vec{r}) + \hat{\mathcal{V}}_{ee}(\vec{r}) \\ &= \hat{\mathcal{T}}_N(\vec{R}) + \hat{\mathcal{H}}_e(\vec{r},\vec{R}),\end{aligned} \tag{2.2}$$

where $\hat{\mathcal{H}}_e(\vec{r},\vec{R})$ is the electronic hamiltonian. Solving this problem in the general case in incredibly complex because of the couplings between electrons and nuclei. However, the mass of electrons and nuclei differ by three orders of magnitude. Thus, we can assume that the electrons move much faster than the nuclei, and that the electrons rearrange instantaneously when the nuclei move. This is equivalent to assume that the nuclei are fixed, with $\hat{\mathcal{T}}_N(\vec{R}) \approx 0$, and we are left with the time-independent electronic

problem:

$$\hat{\mathcal{H}}_e(\vec{r};\vec{R})\phi(\vec{r};\vec{R}) = E\phi(\vec{r};\vec{R}) \tag{2.3}$$

where $\phi(\vec{r};\vec{R})$ is the *electronic* wavefunction, parametrized by the nuclei positions $\vec{R}$. Provided that this equation is solved accurately, a set of wavefunctions $\phi_n(\vec{r};\vec{R})$ with energies E_n are obtained, forming the electronic states of the system. Within these approximations, the general wavefunction of the system can be expressed as a linear combination of these eigenstates in the Born-Oppenheimer expansion [63]:

$$\Psi(\vec{r},\vec{R}) = \sum_n \phi_n(\vec{r};\vec{R})\chi_n(\vec{R}) \tag{2.4}$$

where the $\chi_n(\vec{R})$ are the *nuclear* wavefunctions. Assuming a fast rearrangement of the electrons with respect to nuclei motions means that the nuclei only feel the average potential energy of the electrons, leading to the following set of equations:

$$i\hbar\sum_n \frac{d\chi_n(\vec{R})}{dt} = \hat{\mathcal{T}}_N(\vec{R})\chi_n(\vec{R}) + E_n\chi_n(\vec{R}) \tag{2.5}$$

This last expression is the Born-Oppenheimer approximation [63]. The electronic states are uncoupled, so that a system starting in state n will stay in this state forever. This leads to the picture of nuclei moving on the *adiabatic* potential energy surfaces obtained by solving the electronic problem of equation (2.3).

Assessing the validity of this approximation can be done by injecting the general expression (2.4) into the general problem of equation (2.1), leading to

$$i\hbar\sum_n \frac{d\phi_n(\vec{r};\vec{R})\chi_n(\vec{R})}{dt} = \sum_n \chi_n(\vec{R})\hat{\mathcal{H}}_e\phi_n(\vec{r};\vec{R}) + \sum_n \hat{\mathcal{T}}_N(\vec{R})\chi_n(\vec{R})\phi_n(\vec{r};\vec{R}). \tag{2.6}$$

Projecting this equation on one of the electronic wavefunction, say ϕ_m, and integrating over the electronic positions $\vec{r}$ leads to the set of coupled equations determining the evolution of the nuclear wavefunction χ_m [64]:

$$\begin{aligned} i\hbar\frac{d\chi_m}{dt}(\vec{R}) = \hat{\mathcal{T}}_N(\vec{R})\chi_m(\vec{R}) + E_m\chi_m(\vec{R}) \\ -\hbar^2\sum_n\sum_i \frac{1}{2M_i}\left(2\langle\phi_m|\nabla_{R_i}|\phi_n\rangle\cdot\nabla_{R_i} + \langle\phi_m|\nabla^2_{R_i}|\phi_n\rangle\right)\chi_n(\vec{R}). \end{aligned} \tag{2.7}$$

We have used the general expression for $\hat{\mathcal{T}}_N(\vec{R})$

$$\hat{\mathcal{T}}_N(\vec{R}) = -\hbar^2 \sum_i \frac{1}{2M_i} \nabla^2_{R_i}$$

where the sum is carried out over the number of nuclei in the system, and M_i is the ratio of the masses of nuclei i with respect to the mass of an electron. The last term obtained in equation (2.7) contains the *non-adiabatic* coupling terms $\langle \phi_m | \nabla_{R_i} | \phi_n \rangle$ and $\langle \phi_m | \nabla^2_{R_i} | \phi_n \rangle$, which couples the electronic states to the nuclear motions. Thus, this general expression allows for transitions between electronic states. The Born-Oppenheimer approximation comes down to neglecting these terms. The non-adiabatic coupling terms can be further expressed in function of the derivative of the electronic hamiltonian $\hat{\mathcal{H}}_e$:

$$\langle \phi_m | \nabla_{R_i} | \phi_n \rangle = \frac{\langle \phi_m | \left(\nabla_{R_i} \hat{\mathcal{H}}_e \right) | \phi_n \rangle}{E_n - E_m} \tag{2.8}$$

Thus, we can conclude that the Born-Oppenheimer approximation is valid when:

- the electronic wavefunctions are slowly varying with respect to the atomic positions ;
- the electronic states are well separated in energy.

As highlighted in the introduction, photochemical processes often involve regions where two or more potential energy surfaces come close to one another, or even cross at conical intersections. Thus, the standard Born-Oppenheimer approximation is expected to fail next to these points, and non-adiabatic couplings will have to be taken into account. The consequence is that the method chosen for computational investigations should include an efficient algorithm for computing the non-adiabatic couplings.

2.1.2 Minimum energy path and conical intersections

Potential energy surfaces that are computed in the Born-Oppenheimer approximation are still useful to study reaction pathways in the minimum energy path approach. The shape of potential energy surfaces gives information about the structure accessibility of different reaction paths, by focusing on critical points like slopes, valleys, or funnel points. The paths that are found with this method can be viewed as the reaction path for a vibrationally cold molecule that has infinitesimal amount of kinetic energy [3, 44].

In particular, the geometrical parameters of a conical intersection can be uncovered by means of geometry optimizations. A conical intersection is defined by the intersection between two N-dimensional

energy surfaces [4]. The seam of the conical intersection is an $N-2$-dimension hyperline. Thus, one can distinguish two particular coordinates, usually called $\boldsymbol{g}$ and $\boldsymbol{h}$, which are combination of geometrical parameters and define the *branching space*. Considering two energy potential energy surfaces E_1 and E_2, with associated wavefunctions Ψ_1 and Ψ_2 respectively, it can be shnown that $\boldsymbol{g}$ corresponds to the difference gradient vector, and $\boldsymbol{h}$ corresponds to the coupling vector, defined as:

$$\begin{aligned} \boldsymbol{g} &= \nabla_{R_i}\left[E_1 - E_2\right] \\ \boldsymbol{h} &= \left\langle \Psi_1 \middle| \nabla_{R_i} \middle| \Psi_2 \right\rangle, \end{aligned} \tag{2.9}$$

with the same notations as in the previous section. Plotting the energy along these two coordinates lead to two conical surfaces, while the in the other directions, the energies of ground- and excited-state are equal. Two topologies exist for conical intersection [65] (see figure 2.1). In peaked conical intersection, the excited state (in blue on figure 2.1) is driven toward the seam in all possible directions. This topology leads to highly effective excited state decays, and thus to very efficient photochemistry. On the contrary, in sloped conical intersections, the excited state potential energy surface goes uphill in the vicinity of the seam. This leads to longer excited state lifetimes, and less efficient photochemistry.

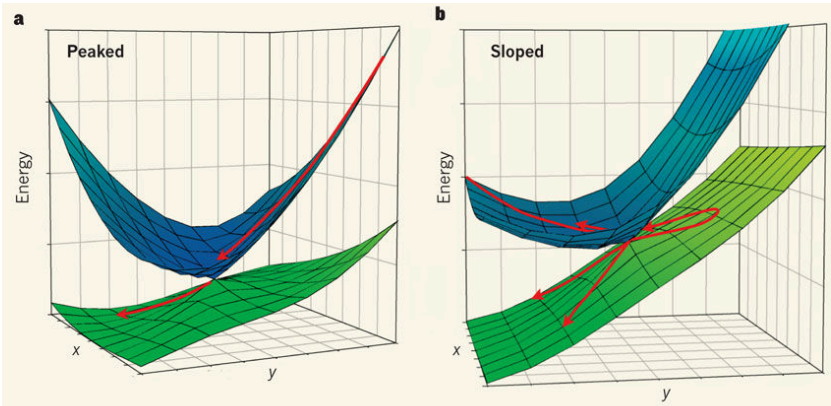


Figure 2.1: Two particular topologies of conical intersection. The excited state potential energy surface is represented in blue, while the ground state is in green. x and y are the coordinates of the branching space, corresponding to the $\boldsymbol{g}$ and $\boldsymbol{h}$ vectors described before (for instance, x can represent C−C bonds stretching, while y can represent a torsions in the polyene chain, in the case of a retinal chromophore). Left: peaked conical intersection ; Right: sloped conical intersection. Reproduced from Ref [65].

Several methods have been developed for optimizing a conical intersection [66]. The method we will focus on here is the one implemented in COBRAM [59] (see section 2.3.3), which is the gradient projection method from Bearpark et al. [67] As the energy changes rapidly in the branching space, the method proceeds by minimizing the energy difference $E_1 - E_2$ in the branching space, while the energy E_2 is minimized in the orthogonal space. The effective gradient used in the method is thus the projection defined as:

$$\bar{\boldsymbol{g}} = \mathcal{P}\frac{\partial E_2}{\partial \boldsymbol{q}_\alpha} \tag{2.10}$$

where $\mathcal{P}$ is the projector onto the space orthogonal to the branching space, and $\boldsymbol{q}_\alpha$ denotes the atomic coordinates.

The static approach described in this part is usually completed by means of semi-classical dynamics, where nuclei are propagated on adiabatic potential energy surfaces. Different behaviour can be observed, as already shown in previous chapter in the case of retinal. The differences between the two approaches is highlighted on figure 2.2.

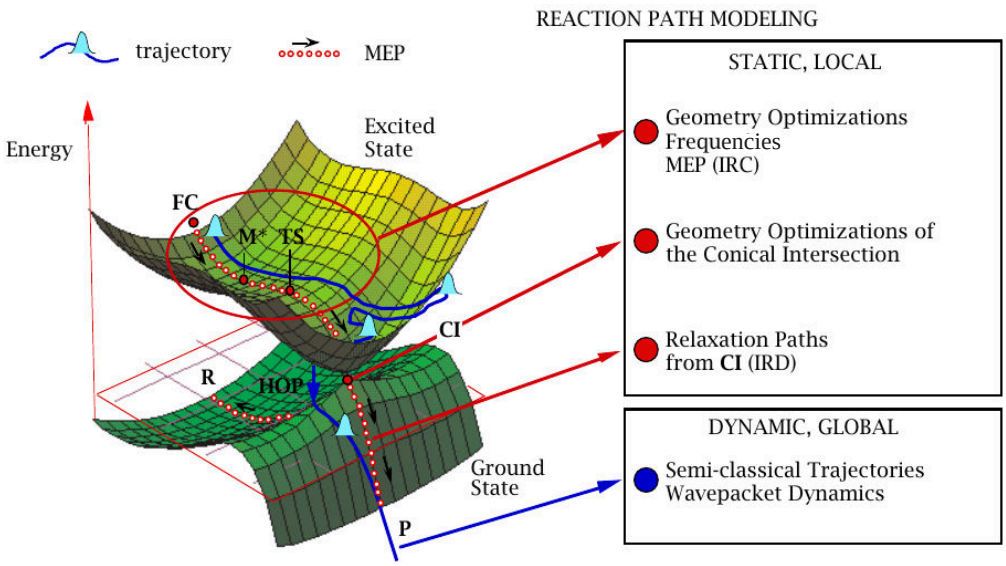


Figure 2.2: Minimum energy path approach (red path) and semi-classical dynamics (blue path). Reproduced from Ref [3].

2.1.3 Semi-classical dynamics and surface-hopping

The dynamical behaviour of a polyatomic system derives in principle from the resolution of the time-dependent Schrödinger equation. However, this is generally not feasible [64]. The Born-Oppenheimer expansion and approximation, derived in section 2.1.1, considerably simplify this picture by separating the electronic and nuclear part of the Hamiltonian. This approximation is based on the difference of masses between nuclei and electrons. The coupled equations obtained, described in equation (2.7), is still too demanding and cannot be solved in the general case, especially when several electronic states are involved. A convenient approximation is then to consider that the nuclei follow classical trajectories on the electronic potential energy surfaces [64, 68, 69]. This approach is called molecular dynamics, and the quantum part is then fully deported on the electrons in the system. The movements of the nuclei is then fully determined by integrating the Newton's equations of motion:

$$M_i \frac{d^2 R_i}{dt^2} = -\frac{\partial E}{\partial R_i} \tag{2.11}$$

where the energy of the system is defined as:

$$E = \sum_i \frac{\boldsymbol{P}_i}{2M_i} + \langle \hat{\mathcal{H}}_e \rangle \,. \tag{2.12}$$

The term $\langle \hat{\mathcal{H}}_e \rangle$ denotes the integration over the electronic wavepacket. The potential energy is then the expectation value of the electronic Hamiltonian. At each step of the simulation, $\hat{\mathcal{H}}_e$ is diagonalized, and the total electronic energy is computed, then injected in the set of equations (2.11). However, in regions of strong electronic coupling, the potential will be state-averaged over several electronic states, which breaks the simplicity of the classical description, and the physical interpretation of these trajectories. For instance, after a conical intersection, the system should be fully described by the ground state potential of the photo-product, and not anymore by the reactant excited-state surface.

This also raises the issue of the treatment of potential energy surface crossings that occur at conical intersections, where nuclei that hop from one surface to another can experience drastic changes in the forces. Proper treatment of the quantum effects induced by electrons on the classical nuclei is of paramount importance for accurate computations of photochemical events. A method that couples the formal simplicity of molecular dynamics and the treatment of the hopping events is the fewest switch surface-hopping algorithm from Tully [68, 69]. The method relies on a large number of trajectories, each

of them considering that nuclei move on a single potential energy surface. Defining a set of eigenfunctions of the electronic Hamiltonian φ_n, the electronic wavefunction of the system can be written as:

$$\phi\left(\vec{r};\vec{R}(t)\right) = \sum_n c_n(t)\varphi_n(\vec{r};\vec{R}(t)), \tag{2.13}$$

and we can also write the density matrix A with elements $a_{lm} = c_l c_m^*$. The diagonal elements a_{kk} correspond to the electronic population of state k, while the off diagonal elements a_{kl} corresponds to the coherence between states k and l.

Deriving the evolution of these coefficients show that electronic states are coupled through off-diagonal Hamiltonian matrix elements, and through non-adiabatic couplings. The general algorithm proceeds by computing the population of states at each step, and the coupling elements that rule the time evolution of these populations. The probability of a switch from state k to state l between steps t and $t + \Delta t$ is equal to the ratio between the change in population of state k in Δt, *i.e.* $\dot{a}_{ll}\Delta t$, and the population in state k at time t, *i.e.* a_{kk} [68]. Thus, the algorithm proceeds as follows, for a two-state system:

1. Initial populations are affected to each state; for instance, the dynamics starts with all the population in state 2: $a_{11} = 0$ and $a_{22} = 1$ initially;

2. the classical equation of motion are solved on the electronic potential energy surface of state 2, and the new coefficients a_{kl} are obtained;

3. the transition probabilities are computed, and compared to a generated random number, to decide if a switch occur ; if no switch occur, we go back to step 2;

4. else, the trajectory evolves on the potential energy surface of new state ; if the potentials are different, a velocity adjustment has to be made.

This algorithm ensures that the number of switches during a trajectory is minimal. As a switch can occur at any point during the trajectory, a large number of trajectories has to be obtained in order to obtain significant results. Finally, we can also propagate a trajectory by disabling the switching capacities. This case corresponds to a movement locked in one energy surface, and can be used to sample the conformational space around a point of interest — this can be especially useful to sample the seam of a conical intersection.

2.2 Electronic wavefunctions

Thus far, we have presented methods that rely on the computation of the matrix elements of the Hamiltonian matrix. This allows the computation of an external potential on which the nuclei can be geometry optimized, or move. In this section, we look in further detail the resolution of the electronic time-independent Schrödinger equation

$$\hat{\mathcal{H}}_e \phi(\vec{r};\vec{R}) = E\phi(\vec{r};\vec{R}) \tag{2.14}$$

In computational chemistry, two frameworks are available for solving this problem. The first is based on wavefunctions (post-HF methods), while the second relies on electronic density (DFT) and is based on the seminal work from Hohenberg and Kohn [70, 71]. This approach has become the most used in computational studies with the design of hybrid exchange-correlation functionals, which allow the inclusion of dynamical correlation in an efficient way. In its time-dependent framework (TD-DFT) [72], it can describe excited states of organic chromophores [73]. However, retinal molecules are the typical case where TD-DFT breaks down, although it gives accurate ground state geometries, in agreement with state-of-the-art wavefunction methods [29]. However, due the charge transfer character of the main electronic excitation of the retinal molecule, the excitation energies obtained in DFT are overestimated [74]. We have recently shown [75] (see chapter 3), based on charge density descriptors developed by Le Bahers et al. [76], that some functionals — especially the most recently developed — were able to perform better than other in this respect, although most of them lead to strong blue-shifts of the vertical absorption. Thus, it is advisable to use different functionals for getting different properties.

On the other hand, multi-reference wavefunctions, especially those based on the complete active space self-consistent field [40, 77] (CASSCF) approach, corrected with second order perturbation theory [41, 78, 79](CASPT2), lead to accurate and balanced description of the retinal in several environments. This method has been applied in gas phase [20, 26, 44, 45], as well as in rhodopsin protein in combination with the AMBER forcefield [8, 31, 54](see section 2.3). This is the approach that will be used throughout our work. Thus, we will only focus on the presentation of these wavefunction-based methods, which are extension of the Hartree-Fock theory.

2.2.1 Hartree-Fock approximation for solving the electronic problem

In the Born-Oppenheimer approximation, the nuclei are considered frozen. Thus, the $\hat{\mathcal{V}}_{NN}(\vec{R})$ term is constant, and will be forgotten in the following. We consider a polyatomic system with atoms at position $\{R_A\}$ and electrons at positions $\{r_i\}$. The distance between electron i and electron j is written r_{ij} and the distance between electron i and nucleus A is written R_{Ai}. The full electronic non-relativistic Hamiltonian, can then be defined, in atomic units, by [80]:

$$\begin{aligned}\hat{\mathcal{H}}_e &= -\frac{1}{2}\sum_i \nabla_{r_i} - \sum_i \sum_A \frac{Z_A}{R_{Ai}} + \sum_{i,j>i} \frac{1}{r_{ij}} \\ &= \hat{h} + \sum_{i,j>i} \frac{1}{r_{ij}}\end{aligned} \tag{2.15}$$

where Z_A is the atomic number of nucleus A, and $\hat{h}$ represents the one-electron part of this Hamiltonian. The electronic wavefunction of the system, denoted $|\Psi\rangle$, with energy E, has to respect the Schrödinger equation

$$\hat{\mathcal{H}}_e |\Psi\rangle = E |\Psi\rangle \tag{2.16}$$

Moreover, Fermi principle stands that $|\Psi\rangle$ has to be antisymmetric with respect to the exchange of two electrons. Each electron in the system is described by a one-electron wavefunction ϕ_i, called a *spin-orbital*, which is the product of a spatial function $\psi_i(r)$ and a spin function $\sigma_i(\omega)$. This general spin function will be written $\alpha(\omega)$ for spin *up* and $\beta(\omega)$ for spin *down*, and the generalized coordinate x gathers the spatial coordinate r and the spin coordinate ω. In summary:

$$\phi_i(x) = \psi_i(r)\sigma_i(\omega) = \begin{cases} \psi_i(r)\alpha(\omega) \\ \psi_i(r)\beta(\omega) \end{cases} \tag{2.17}$$

Given a set of spin-orbitals ϕ_n the most simple way to build an antisymmetric wavefunction is the Slater determinant defined by

$$|\Psi\rangle = \begin{vmatrix} \phi_i(x_1) & \phi_j(x_1) & \dots & \phi_k(x_1) \\ \phi_i(x_2) & \phi_j(x_2) & \dots & \phi_k(x_2) \\ \dots & \dots & \dots & \dots \\ \phi_i(x_n) & \phi_j(x_n) & \dots & \phi_k(x_n) \end{vmatrix} = |\phi_1\phi_2\dots\phi_n|, \tag{2.18}$$

The hypothesis in the Hartree-fock approximation is that a single Slater determinant of the form (2.18) can give a reasonable description of the ground state of the N-electrons system considered. The variational principle stands that the best possible approximation of the ground state wavefunction Ψ has to minimize the energy of the system, which can be written $\langle\Psi|\hat{\mathcal{H}}_e|\Psi\rangle$. Thus, the initial problem comes down to find the best set of spin-orbitals to form a Slater determinant that minimizes the energy of the system.

Applying the variational principle [80] shows that the optimal spin-orbitals have to be eigenvectors of the Fock operator f defined by its matrix elements f_{ij} by:

$$\begin{aligned} f_{ij} &= \langle\phi_i|f\left|\phi_j\right\rangle \\ &= h_{ij} + \left[\sum_b (ij|bb) - (ib|bj)\right] \end{aligned} \tag{2.19}$$

where h_{ij} is an element of the matrix representing $\hat{h}$, the one-electron part of the Hamiltonian, and where we have used the standard notation for the two-electron integrals:

$$(ab|cd) = \int\int \phi_a^*(x_1)\phi_c^*(x_2)\frac{1}{r_{12}}\phi_b(x_1)\phi_d(x_2)dx_1dx_2. \tag{2.20}$$

The optimal spin-orbitals can thus be defined as verifying:

$$f\,|\chi_n\rangle = \epsilon\,|\chi_n\rangle\,, \tag{2.21}$$

where ϵ_n defines the energy of the spin-orbital n. The Slater determinant formed out of the N lowest in energy spin-orbitals is then called the Hartree-Fock wavefunction, and will be noted Ψ_0. Its total energy is:

$$E = \sum_i h_{ii} + \frac{1}{2}\sum_{ij}(ii|jj) - (ij|ji), \tag{2.22}$$

where the factor $1/2$ prevents the double-counting of the electron-electron interaction. The N spin-orbitals with the lowest energy are *occupied*. As equation (2.21) has an infinite numer of solutions, the remaining spin-orbitals will be referred to as *unoccupied* or *virtual*. As the expression of f depends on a set of spin-orbitals, the problem of finding the eigenvectors in equation (2.21) is done iteratively. However, no method is generally available to solve numerically this equation in molecules.

The formulation of Roothaan [81] simplifies the picture, and allows the computation of molecular orbitals through introducing a set of known basis functions. In the case of closed-shell systems, where all valence electrons are paired, the set of N spin-orbitals can be built with $N/2$ spatial orbitals following:

$$\begin{aligned} \chi_{2i} &= \psi_i(r)\alpha(\omega) \\ \chi_{2i-1} &= \psi_i(r)\beta(\omega) \end{aligned} \tag{2.23}$$

Introducing a set $\{\varphi_k\}$ of known basis functions, we can then expand the unknown molecular orbitals ψ_i in the basis of these known functions:

$$\psi_i = \sum_k C_{ki}\varphi_k \tag{2.24}$$

In principle, the set $\{\varphi_k\}$ is infinite. However, in practice, a finite set of K functions is used. Thus, K spatial functions with the form given in equation (2.24) can be built, leading to a set of $2K$ spin-orbitals. Solving equation (2.21) for an N-electron system thus amounts to finding the optimal set of coefficients $\{C_{ki}\}$ that give the optimal set of spin-orbitals, and the variational principle then translates to the set of linear combinations known as the Roothaan equations:

$$\boldsymbol{f}\boldsymbol{C} = \boldsymbol{S}\boldsymbol{C}\boldsymbol{\epsilon}, \tag{2.25}$$

where $\boldsymbol{f}$ is the matrix of the closed-shell Fock operator, $\boldsymbol{C}$ is the matrix of the coefficients, $\boldsymbol{S}$ is the overlap matrix with elements $S_{ij} = \langle \varphi_i \,|\, \varphi_j \rangle$, an $\boldsymbol{\epsilon}$ is the diagonal matrix containing the molecular orbital energies as defined in equation (2.21). The elements of the Fock matrix can be expressed with the one-particle density matrix $\boldsymbol{D}$:

$$\begin{aligned} D_{rs} &= 2 \sum_{occupied} C_{ra}C^*_{sa} \\ f_{pq} &= h_{pq} + \sum_{rs} D_{rs}\left[(pq|rs) - \frac{1}{2}(pr|sq)\right], \end{aligned} \tag{2.26}$$

This problem is non-linear, and has to be solved iteratively. The resulting procedure is called the *self-consistent field* (SCF) method.

Expanding the basis of K spatial functions will lead to more flexibility in solving the Fock equations, and thus lead to lowering the total energy. The energy obtained for an infinite set is called the Hartree-Fock limit E_{HF}. Slater determinants correlate electrons with the same spin, but do not include correlation

effect for electrons with opposite spin. This results in an overestimation of the Hartree-Fock energy, and the difference with the exact energy is called the correlation energy E_{corr}, which is a negative quantity:

$$E_{exact} = E_{HF} + E_{corr} \tag{2.27}$$

The correlation energy highlighted in equation (2.27) can be artificially split in two parts. The first part in the dynamic correlation, arising from the instantaneous electron-electron interaction, that is completely absent from Hartree-Fock theory — as one electron only sees the average electronic cloud created by the others. The second part is the static correlation, and has no classical equivalent. This part arises from the interaction between multiple Slater determinants. Therefore, in cases where the correlation energy is non-negligible, the Hartree-Fock theory is expected to fail. This happens, in particular, in cases where a single determinant is not sufficient to describe the electronic structure of the molecule, for instance in transition metals, or for the description of excited state. Recovering the correlation energy is thus a crucial part in developing quantum chemistry methods.

2.2.2 Configuration interaction and multi-configuration wavefunctions

Configuration interaction

Recovering static correlation requires the introduction of more than one Slater determinant. The Hartree-Fock wavefunction is built by creating a Slater determinant with the N spin-orbitals that are eigenvectors of the Fock operator f with the lowest energy. As there are in principle an infinite number of solution — in practice, we are limited by the number of basis functions used in the expansion in equation (2.24) — we can define wavefunctions obtained by promoting an electron from an occupied spin-orbital χ_a to an unoccupied spin-orbital χ_r in the Hartree-Fock wavefunction $|\Psi_0\rangle$. Such wavefunctions will be noted $\left|\Psi_a^r\right\rangle$. In the following, the indices $a, b, c, \ldots$ refer to occupied orbitals, $r, s, t, \ldots$ refer to unoccupied orbitals and $i, j, k \ldots$ refer to arbitrary orbitals. Provided that the Hartree-Fock wavefunction is a reasonable approximation, the *exact* wavefunction $|\Phi\rangle$ can then be defined as the sum of all possible excitations obtained from $|\Psi_0\rangle$ by forming the configuration interaction (CI) expansion:

$$|\Phi\rangle = c_0 + |\Psi_0\rangle + \sum_{a,r} c_a^r \left|\Psi_a^r\right\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs} \left|\Psi_{ab}^{rs}\right\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst} \left|\Psi_{abc}^{rst}\right\rangle + \ldots \tag{2.28}$$

For a complete determination, the CI coeffcients $\{c_a^r\}, \{c_{ab}^{rs}\}, \ldots$ have to be determined variationally. This is done by computing the matrix elements of the Hamiltonian in the basis of the N-electron wavefunctions. Even with simplifications from the Brillouin theorem — namely, determinants that differ by more than two spin-orbitals do not interact, and single excited determinants do not interact with the ground state — and taking into account spatial and spin symmetry, the size of the full CI matrix is usually way above computational capacities. Indeed, exciting n out of N electrons leads to $\binom{N}{n}$ possibilities ; if there are $2K$ spin-orbitals, this leads to $2K - N$ virtual orbitals, so $\binom{2K-N}{n}$ ending points for the excitation. Overall, we can build $\binom{N}{n}\binom{2K-N}{n}$ determinants. The full CI approach leads to exact correlation energy, within the subspace spanned by the one-electron basis set chosen for the Hartree-Fock expansion. The size of the full CI matrix makes this method prohibitive for most applications in quantum chemistry, save for the smallest systems.

Multi-configurational self-consistent field

The complexity of the multi-configurational expression in equation (2.28) can be reduced if one only performs it in a selected number of configurations. This is the approach called Multi-Configurationnal Self-Consistent Field (MCSCF). With this method, one forms a linear combination of the configurations $|\mu\rangle$ — which are Slater determinants — that are expected to contribute significantly to static correlation:

$$|\Psi_{MCSCF}\rangle = \sum_{\mu} C_{\mu} |\mu\rangle . \tag{2.29}$$

The Hamiltonian matrix is then diagonalized in the new basis formed by these new wavefunctions, and the CI coefficients C_μ as well as the molecular spin-orbitals are optimized variationally. This method has the inconvenient of forcing the user to manually select electronic configurations.

However, this approach can be extended to be complete within a subset of molecular spin-orbitals. This is the Complete Active Space Self-Consistent Field [40, 77, 82](CASSCF) method, which relies heavily on the construction of configuration state functions (CSFs). These configuration state functions are symmetry-adapted linear combinations of Slater determinants. The molecular spin-orbitals space is divided in three part:

- *inactive*: spin-orbitals that are doubly occupied in all CSFs;
- *active*: spin-orbitals with varying occupation number:
- *external*: spin-orbitals that are unoccupied in all CSFs.

Considering n electrons in M spin-orbitals is called CASSCF(n, M). The CASSCF wavefunction is then built as a linear combination of all the CSFs obtained. The active orbitals are manually selected as those that are expected to give significant contributions to static correlation. The main advantage of the CASSCF method is that once the active space has been selected, the CASSCF wavefunction is fully characterized. As in MCSCF, the CI coefficients as well as the molecular spin-orbitals are optimized using the variational principle. The state-averaged formulation of CASSCF (SA-CASSCF) allows the simultaneous optimization of the ground and several electronic excited states using a single set of averaged orbitals.

The main limitations of the CASSCF approach is the size of the active space. Indeed, it can be shown that the number of CSFs that can be created for an active space (n, M) with total spin S is

$$N_{CSFs} = \frac{2S+1}{M+1}\binom{M+1}{\frac{n}{2}-S}\binom{M+1}{\frac{n}{2}+S+1} \tag{2.30}$$

and thus has an approximate factorial grow with the size of the active space. As the active space selection is crucial for accurate prediction [83], or even for the convergence of the algorithm, this can raise issues for large systems. Approaches have been developed, where the active space is further divided in three part, within the restricted active space approach (RASSCF) [84], where the total number of holes (in space RAS1) and total number of particles (in space RAS3) can be tuned, thus improving the size of the effective active space. Another development is the generalized active space [85] (GASSCF), in which the number of subspaces is in principle arbitrary, and the number of inter-space excitations is fixed. These different approaches are summarized on figure 2.3.

In Hartree-Fock theory, molecular orbitals were defined as the eigenvectors of the Fock operator in equation (2.21). This concepts survives in CASSCF procedure in terms of Natural Orbitals. These orbitals are obtained from the one-particle density matrix $\gamma(x_1, x_1')$:

$$\gamma(x_1, x_1') = N \int dx_2 \dots dx_N \Psi(x_1, x_2, \dots x_n)\Psi^*(x_1', x_2, \dots x_n), \tag{2.31}$$

which can be expanded in an orthonormal basis of Hartree-Fock spin-orbitals to define its matrix elements γ_{ij}:

$$\gamma(x_1, x_1') = \sum_{i,j} \gamma_{ij}\chi_i(x_1)\chi_j^*(x_1') \tag{2.32}$$

As this matrix is hermitian, it can be diagonalized by means of unitary transformations of the Hartree-Fock spin-orbitals to find its eigenvectors η_i, with corresponding eigenvalues λ_i. These new spin-orbitals

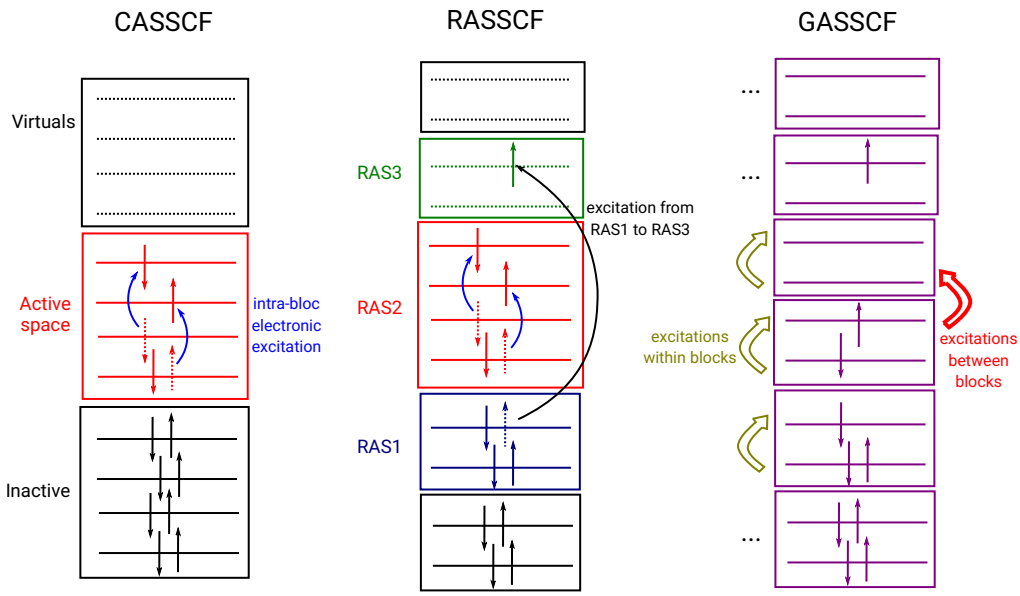


Figure 2.3: Comparison between the CASSCF, RASSCF and GASSCF approaches.

are called the *natural orbitals*, and the corresponding eigenvalues are called the *occupation number*, which are taken as the formal equivalent of spin-orbitals and orbital energies for HF methods [86].

2.2.3 Dynamical correlation

Recovering the dynamical correlation requires to re-inject the instantaneous electron-electron interaction that is averaged in the variational treatment of Slater determinants. This can efficiently be done by means of perturbation theory.

Rayleigh-Schrödinger perturbation theory

Considering a system with Hamiltonian $\hat{\mathscr{H}}$, we want to solve the associated Schrödinger equation

$$\hat{\mathscr{H}} \left|\Psi\right\rangle = \mathcal{E} \left|\Psi\right\rangle, \tag{2.33}$$

Let's assume that we can separate $\hat{\mathscr{H}}$ in two parts: $\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \mathcal{W}$, so that the eigenstates and eigenvalues of $\hat{\mathscr{H}}_0$ are known, *i.e.* we have solved the problem:

$$\hat{\mathscr{H}}_0 \left|i\right\rangle = E_i^{(0)} \left|i\right\rangle \tag{2.34}$$

Our aim is to find the perturbation on $E_i^{(0)}$ and $|i\rangle$ induced by the addition of the operator $\mathcal{W}$, assuming that the perturbation is reasonably small. Therefore, $\mathcal{E}$ and $|\Psi\rangle$ are expected to be close to $E_i^{(0)}$ and $|i\rangle$. One way to do this is to to connect the unperturbed problem from equation (2.34) to the original one from equation (2.33). Thus, we introduce a parameter λ, with $0 \leq \lambda \leq 1$, so that a new Hamiltonian $\hat{\mathscr{H}}_\lambda$ is defined as [87]:

$$\hat{\mathscr{H}}_\lambda = \hat{\mathscr{H}} + \lambda \mathcal{W} \tag{2.35}$$

The exact eigenfunctions and eigenvalues can now be expanded in terms of powers of λ:

$$\mathcal{E} = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \tag{2.36}$$

$$|\Psi\rangle = |i\rangle + \lambda \left|\Psi^{(1)}\right\rangle + \lambda^2 \left|\Psi^{(2)}\right\rangle + \dots \tag{2.37}$$

The coefficients $E_i^{(n)}$ are the n-th order correction to the unperturbed energies, and the fonctions $\left|\Psi^{(n)}\right\rangle$ are the n-th order correction to the wavefunctions. Moreover, we assume that the eigenvectors of $\hat{\mathscr{H}}_0$ are orthonormal, and we impose the intermediate normalization, meaning that we impose $\langle i|\Psi\rangle = 1$. Projecting equation (2.37) on $|i\rangle$ leads to a polynomial expression in λ that has to be zero, and hence leads to the conclusion that all the perturbed wavefunction terms $\left|\Psi^{(n)}\right\rangle$ are orthogonal to $|i\rangle$.

Injecting equations (2.36) and (2.37) into equation (2.33), and equating the terms in λ^n on each side of the resulting equation, we obtain the following results:

$$\hat{\mathscr{H}}_0 |i\rangle = E_i^{(0)} |i\rangle \tag{2.38}$$

$$\hat{\mathscr{H}}_0 \left|\Psi^{(1)}\right\rangle + \mathcal{W} |i\rangle = E_i^{(0)} \left|\Psi^{(1)}\right\rangle + E_i^{(1)} |i\rangle \tag{2.39}$$

$$\hat{\mathscr{H}}_0 \left|\Psi^{(2)}\right\rangle + \mathcal{W} \left|\Psi^{(1)}\right\rangle = E_i^{(0)} \left|\Psi^{(2)}\right\rangle + E_i^{(1)} \left|\Psi^{(1)}\right\rangle + E_i^{(2)} |i\rangle \tag{2.40}$$

$$\dots \tag{2.41}$$

The perturbated energies can be obtained by projecting these expressions on $|i\rangle$. This leads to:

$$E_i^{(0)} = \langle i| \hat{\mathscr{H}}_0 |i\rangle \tag{2.42}$$

$$E_i^{(1)} = \langle i| \mathcal{W} |i\rangle \tag{2.43}$$

$$E_i^{(2)} = \langle i| \mathcal{W} \left|\Psi^{(1)}\right\rangle \tag{2.44}$$

$$\dots \tag{2.45}$$

Finding the corrections to the energy then comes down to solving these equations. Among these, equation (2.42) gives the same eigenvalue problems as equation (2.34). The first order term is simply the expectation value of the perturbation $\mathcal{W}$ on state $|i\rangle$. Finding higher order corrections , however, require the computation of the perturbated wavefunctions. In general, the calculation of perturated energies at order n requires the calculation of the wavefunction at order $n-1$. We will only compute the second order energy, but the process is the same for the higher order terms.

Projecting equation (2.39) on any eigenfunctions $|n\rangle$ of $\hat{\mathscr{H}}_0$ different from $|i\rangle$ gives the components of $\left|\Psi^{(1)}\right\rangle$ in the basis $\{|n\rangle\}$ with $n \neq i$, which allows to write:

$$\begin{aligned}\left|\Psi^{(1)}\right\rangle &= \sum_n \langle n|\Psi^{(1)}\rangle\, |n\rangle \\ &= \sum_{n\neq i} \frac{\langle n|\,\mathcal{W}\,|i\rangle}{E_i^{(0)} - E_n^{(0)}}\, |n\rangle\end{aligned} \tag{2.46}$$

This expression can then be injected in equation (2.44) to find

$$E_i^{(2)} = \sum_{n\neq i} \frac{|\,\langle n|\,\mathcal{W}\,|i\rangle\,|^2}{E_i^{(0)} - E_n^{(0)}} \tag{2.47}$$

The computation of the other terms is done in exactly the same way. For our purpose, however, second order energies will be sufficient.

Møller-Plesset methods

In the previous section we have exposed the Rayleigh-Schrödinger theory. We did not make any assumptions on the wavefunctions used, and in particular, the results and the general expressions obtained remain valid in the case of N-electron wavefunctions. In the Hartree-Fock approximation, we have solved the eigenvalue problem for the one-electron operator f as:

$$f\,|\chi_n\rangle = \epsilon_n\,|\chi_n\rangle\,, \tag{2.48}$$

leading to a set of N occupied spin-orbitals. In other terms, the Slater determinant $|\Psi_0\rangle$ built out of these N occupied spin-orbitals is an eigenvector of the Hartree-Fock Hamiltonian $\hat{\mathcal{H}}^{HF}$ with eigenvalue E_0^0,

defined by:

$$\hat{\mathcal{H}}^{HF}|\Psi_0\rangle = E_0^0|\Psi_0\rangle$$
$$\sum_{n=1}^{N} f(n)|\Psi_0\rangle = \sum_{n=1}^{N}\left(h(n)+v^{HF}(n)\right)|\Psi_0\rangle = \sum_{n=1}^{N}\epsilon_n|\Psi_0\rangle \tag{2.49}$$

Møller-Plesset theory applies the results obtained in the previous section to a perturbation $\mathcal{W}$ that is equal to the difference between the real two-electron interaction and the Hartree-Fock potential v^{HF} [80]:

$$\mathcal{W} = \sum_{i<j}\frac{1}{r_{ij}} - \sum_{n} v^{HF}(n) \tag{2.50}$$

Applying the formula derived above for first order correction to the energy, we find:

$$E_0^{(1)} = -\frac{1}{2}\sum_{ab}(aa|bb)-(ab|ba), \tag{2.51}$$

which, when summed up with $E_0^{(0)}$, gives back the Hartree-Fock energy (see equation (2.22)). The second order correction involves terms of the form $\langle\Psi_0|\mathcal{W}|n\rangle$ for $|n\rangle$ that are eigenvectors of $\hat{\mathcal{H}}^{HF}$ different from $|\Psi_0\rangle$. Thus, they are excited determinants. Due to Brillouin theorem and the fact that we consider wavefunctions that are eigenvectors of the Fock operator, the single excited determinants will not contribute to the sum; furthermore, as the perturbation is a two-electron operator, triple and higher order excited states will not contribute either. Thus, only the doubly excited determinants play a role in the second order corrections, leading to the following expression, where the indices $a, b, \ldots$ range over occupied orbitals and indices $r, s, \ldots$ range over virtual orbitals:

$$E_0^{(2)} = \frac{1}{4}\sum_{abrs}\frac{|(ar|bs)-(as|br)|^2}{\epsilon_a+\epsilon_b-\epsilon_r-\epsilon_r} \tag{2.52}$$

Perturbation theory for CASSCF reference wavefunction

In the previous section, we developed a perturbation theory with a single Slater determinant as a starting point. Thus, this method is valid for improving the description of systems where a single determinant is a good approximation of their electronic wavefunctions. As pointed out above, in section 2.2.2, this is not the case for many chemical systems, especially when static correlation plays an important role. In particular, excited states of organic molecules are problematic. However, perturbation theory is an easy way to include dynamical correlation effects. Therefore, much effort has been devoted to extending the perturation theory to multi-configurational wavefunctions. In particular, Andersson et al. have developed

a method to introduce second order corrections to a reference state described by a multi-configurational CASSCF wavefunction [41, 78, 79]. The detailed implementation of the method are out of the scope of this document, but we will highlight its main characteristics.

We assume that a CASSCF wavefunction $|0\rangle$ has been optimized, with energy E_0. We want to obtain the second order energy correction. The configurational space is separated in four subspaces, associated with projection operators:

- V_0 is the one-dimensional space spanned by the reference CASSCF wavefunction, associated to projector $\hat{P}_0$;
- V_K corresponds to states generated by excitations within the active space, associated to $\hat{P}_K$;
- V_{SD} is the space spanned by all single and double excitations generated from the reference wavefunction, associated to $\hat{P}_{SD}$ — it is worth to mention that this subspace contains V_K:
- $V_{TQ...}$ is the space containing all higher order excitations, associated to $P_{TQ...}$.

As seen above, in section 2.2.3, the second order energy is completely characterized by the first-order wavefunctions, which in turn can only be composed of elements that are orthogonal to the reference state. As the Hamiltonian contains only one- and two-electron operators, the first order wavefunction can thus be decomposed exclusively on the elements of V_{SD}. Furthermore, the generalized Brillouin theorem stands that single excited states do not interact with the reference state through the Hamiltonian. Therefore, only the double excitations from V_{SD} are considered. This subspace is then further divided in groups of internal, semi-internal and external subspace when the double excitation involve none, one or two secondary orbitals respectively. To summarize, we can develop the first order wavefunction $\left|\Psi^{(1)}\right\rangle$ as:

$$\left|\Psi^{(1)}\right\rangle = \sum_j C_j \left|j\right\rangle, \qquad \left|j\right\rangle \in V_{SD}, \tag{2.53}$$

where the $\{C_j\}$ are solutions of:

$$\sum_j C_j \left\langle i\right| \hat{\mathcal{H}}_0 - E_0 \left|j\right\rangle = -\left\langle i\right| \hat{\mathscr{H}} \left|0\right\rangle, \qquad \left|i\right\rangle \in V_{SD}, \tag{2.54}$$

resulting from equation (2.39). Two problems remain. The first is the definition of $\hat{\mathcal{H}}_0$; the second is the size of the matrix in left hand side of equation (2.54).

Defining the reference Hamiltonian is done with analogy to the Møller-Plesset method (see section 2.2.3). In fact, the implementation of CASPT2 is strictly equivalent to MP2 when the reference state is a closed-shell determinants. This is done by choosing a one-electron operator $\hat{F}$ so that:

$$\hat{\mathcal{H}}_0 = \hat{P}_0\hat{F}\hat{P}_0 + \hat{P}_K\hat{F}\hat{P}_K + \hat{P}_{SD}\hat{F}\hat{P}_{SD} + \hat{P}_{TQ...}\hat{F}\hat{P}_{TQ...}, \tag{2.55}$$

where the operator $\hat{P}_\alpha$ are the projectors defined above. The chosen one-electron operator is a generalized Fock operator, of the same form as the one obtained in the Roothaan equations, with matrix elements defined in second quantization as:

$$\begin{aligned} \hat{F} &= \sum_{pq} f_{pq}\hat{E}_{pq} \\ \hat{E}_{pq} &= \sum_{\sigma} \hat{a}^\dagger_{p\sigma}\hat{a}_{q\sigma} \\ f_{pq} &= h_{pq} + \sum_{rs} D_{rs}\left[(pq|rs) - \frac{1}{2}(pr|sq)\right], \end{aligned} \tag{2.56}$$

where the indices p, q run over all orbitals, and D_{rs} are the elements of the density matrix. $\hat{E}_{pq}$ is the spin-averaged substitution operator that promotes an electron from orbital q to orbital p. Unlike the case of single reference Slater determinants, the off-diagonal elements of f cannot all be eliminated, since only rotation within active, inactive or secondary orbitals leave the wavefunction unchanged. This matrix is thus composed of 3x3 blocks corresponding to inactive, active and secondary subspaces. With analogy to the single determinant closed-shell case, the diagonal elements of f are called "orbital energies" and are noted $f_{pp} = \epsilon_p$.

In the expressions developed in section 2.2.3, the denominators of the second order energies are the differences between energies of the reference state and energies of the states in the interacting space. For single Slater determinants, there is usually — at least for organic molecules — a significant separation in energy between the reference and the first-order interacting states. However, this does not hold in the case of CASSCF wavefunctions, where some wavefunctions in the interacting space can be close in energy to the reference wavefunction. Such states are called *intruder states*. To avoid the unphysical divergence in energy, shift methods have been developed. The most used is the "imaginary shift" [88], which is added to the zeroth-order Hamiltonian $\hat{\mathcal{H}}_0$, and the effects on the computed energy are removed when the couplings with intruder states have been handled. A value of 0.2 has been shown to be effective for correcting the problem. The "ionization potential–electronic affinity" (IPEA) shift has been introduced

in order to correct an unbalanced treatment of closed-shell and open-shell CSFs, leading to systematic errors in the treatment of open-shell systems [89]. However, it was recently shown that IPEA shift should not be used for computations on excited states of organic chromophores [90]. Thus, this will not be developed further.

The last extension of the CASPT2 method is the multi-state CASPT2 (MS-CASPT2) treatment [91], where the reference space V_0 is now spanned by several CASSCF wavefunctions obtained from a state-averaged computation. This method is especially suitable when several electronic states are mixed at the CASSCF level. The main difference with single-state CASPT2 (SS-CASPT2) is that an effective Hamiltonian is formed, where the diagonal elements are the SS-CASPT2 energies, and the off-diagonal elements are the coupling between CASSCF states and their first-order corrections *via* the full Hamiltonian. Diagonalizing this effective Hamiltonian leads to MS-CASPT2 energies and perturbatively modified CASSCF (PM-CASSCF) wavefunctions.

In this section, we have presented the methods used for obtaining accurate electronic wavefunctions for the retinal chromophore. However, we want to model the retinal chromophore, for which the external environment has a significat role on its electronic states, as presented in the introduction. The next section deals with the inclusion of this environment in the computation.

2.3 Quantum Mechanics / Molecular Mechanics schemes

Biological reactions often occur in burried sites inside large molecular systems such as proteins, which are in turn embedded in extended media, like membranes or solvent. Even when this environment is not directly involved in the electronic processes, it can affect the reactions by imposing constraints on the shape of the fragments involved, or by favoring particular intermediates *via* electrostatic interactions [92]. Therefore, modelling the environment in of utmost importance to get reliable computational models. However, quantum mechanics (QM) can only treat a limited number of atoms, even if hardware and software developments — such as linear scaling approaches [93] — allow the full treatment of larger and larger systems. Another problem comes with the high number of conformations that are thermally accessible in proteins, and the timescales involved. The description of these large systems is conveniently done in the molecular mechanics (forcefield) approach, although this does not allow the description of electronic processes. Combining the two approaches allows the description of the electronic processes embedded in large molecular systems, leading to the Quantum Mechanics / Molecular

Mechanics (QM/MM) framework [94, 95]. In this section, we will first describe the forcefield approach. Then, we will review the practical aspects for developing a QM/MM scheme. Finally, we will describe our own software implementation with the COBRAM interface.

2.3.1 Molecular mechanics and forcefield approach

The description of large biomolecular systems is usually done by means of parametrized forcefields, where atoms are considered as semi-rigid spheres, interacting *via* bonded interactions (based on chemical connectivity) including bond vibrations, angle stretchings and torsional potentials, and non-bonded interactions, such as electrostatic potential and van der Waals forces. As their is no unique way to define each of these interactions, several forcefield families have been developed.

Each forcefield relies on building blocks that are parametrized independently, and may be used together to build a complete systems. The building blocks depend on the forcefield, and on the level required for the description. For most biomolecular applications dealing with reactions occuring in proteins, all-atom forcefields are used, where the building blocks are amino-acids in protein, and conjugated bases in DNA. These building blocks are composed of several "atom types", which are entities whose properties depend on their bonding environment. The full energy of the system is obtained as a sum of the individual contributions coming from each atom in the system. One of the most simple approach is the one used in AMBER forcefields, where the system total energy is written as [96]:

$$\begin{aligned} E^{tot} = & \sum_{bonds} k_D(d-d_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\Phi[1+cos(n\Phi+\delta)] \\ & + \sum_{A,B}\left(\epsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right] + \frac{1}{4\pi\epsilon_0}\frac{Q_A Q_B}{r_{AB}}\right) \end{aligned} \tag{2.57}$$

In this expression, the bond vibrations and angle stretchings are described with harmonic potentials; torsional potentials are described with a multiplicity n and a phase δ; A and B are non bonded atoms, with van der Waals interactions described with a Lennard-Jones potential.

Special attention has to be devoted to the electrostatic term, since this is a long-range interaction that can have sensitive impact on biological systems. Atomic charges are artificial quantities derived from QM electrostatic potential, and that are carefully parametrized and tested. Thus, they only have meaning inside a given forcefield. This last comment is also true for all the other parameters, and transferring them from one forcefield to another is not trivial. Therefore, the combination of molecular mechanics with quantum mechanics has to be carefully calibrated and tested [94].

2.3.2 Combining QM and MM approaches

The combination of quantum mechanics and molecular mechanics stems from the seminal work of Warshel et al. [58], where an enzymatic reaction has been modeled in the protein site by means of semi-empirical quantum mechanics appproach. The first step in a QM/MM study is the partition of the full (real) system [94], denoted $\mathcal{R}$, in two parts: the first is the inner part $\mathcal{I}$ that contains the fragment that are expected to undergo significant electronic rearrangement, and is thus treated at the quantum mechanics (QM) level ; the outer part $\mathcal{O}$ contains the remaining part of the system, and it is treated at the molecular mechanics (MM) level. The separation between the two layers can be straightforward in case of absence of covalent connections among them, such as a molecule in solution, but other cases such as the reaction center in a protein, where the relevant molecules are indeed covalently bonded to the rest of the protein, are more problematic. In this case, the separation between the inner and outer system involves cutting a covalent bond. Thus, special care has to be taken, because simply cutting a covalent bond results in a radical in the QM region, and an ill-defined fragment in the MM region. In general, all bonded interactions involving an MM atom will be treated at the MM level. Several schemes have been designed to fill the valency of the dangling bond in the QM region, such as the frozen orbital method [97, 98], where a strictly localized orbital is used to describe the the electrons of the frontier bond, or the atom-link method [99, 100], where an atom, usually an hydrogen, is added to the QM region and is transparent for the MM system. The link region will be denoted $\mathcal{L}$.

Both methods have advantages as well as inconvenients. Frozen orbitals have to be parametrized for each system, and for each type of computations (theoretical level and basis set) [101]. However, it describes accurately the properties of the broken bond. Several problems arise with the link-atom method. Indeed, this atom will be very close to the group covalently linked, which may lead to very high electrostatic potential in this region. Second, this adds several degrees of freedom to the QM region. Finally, a C−H bond is chemically different from the original covalent bond, although the method has been found relatively accurate when the original covalent bond was between two C(sp^3) atoms [94]. However, the implementation of such a method is very easy, and transferable, as it does not depend on external parametrization. The electrostatic problem may be overcome with a redistribution of the charges of the MM-dangling bond to the surrounding MM atoms [59, 94]. The additional degrees of freedom may be removed by constraining the position of the link atom to be on the QM/MM bond, and by fixing the QM-H distance.

Two main schemes have been developed for computing the total QM/MM energy $E^{\mathcal{R}}_{QM/MM}$ [95]. In the *substractive* scheme, mainly used for link-atom methods, the energy of the real system is computed at the MM level, and the energy of the inner system, including the atom-link is computed at the QM and at the MM level, leading to the following expression of the energy:

$$E^{\mathcal{R}}_{QM/MM} = E^{\mathcal{R}}_{MM} + E^{\mathcal{I}+\mathcal{L}}_{QM} - E^{\mathcal{I}+\mathcal{L}}_{MM}. \tag{2.58}$$

This scheme effectively amounts to a pure MM computation, where a part of the system is cut out and treated at the QM level. The computation of $E^{\mathcal{I}+\mathcal{L}}_{QM}$ is closely related to the embedding scheme implemented (see below). This method requires MM parameters for all atoms in the system, which can be problematic, especially when chemical reactions occur in the inner system. The problem is that strictly speaking, different sets of MM parameters should be obtained for reactants and products; however, the transition from one to another is not trivial. The second problem with this approach is that all the QM/MM interactions are treated purely at the MM level. However, this method is very simple to implement due to the absence of explicit coupling terms, and due to software development that allows the derivation of MM parameters for a wide range of organic residues — namely, the GAFF forcefield for AMBER [102].

In the *additive* scheme, the energy of the outer system is computed at the MM level, while the energy of the inner system is computed at the QM level. Some coupling terms are then added to the total, and the total energy can be written as:

$$E^{\mathcal{R}}_{QM/MM} = E^{\mathcal{O}}_{MM} + E^{\mathcal{I}+\mathcal{L}}_{QM} + E^{\mathcal{O},\mathcal{I}}_{QM-MM}. \tag{2.59}$$

In this scheme, the coupling term $E^{\mathcal{O},\mathcal{I}}_{QM-MM}$ contains the interactions between the inner and outer part, namely the van der Waals, bonded and electrostatic interactions.

Most substractive schemes currently used, such as ONIOM [103] or COBRAM [59] (see section 2.3.3), go beyond the strict approach described above. Particular care has been taken in the treatment of the QM-MM electrostatic contribution, and these schemes have been improved to use QM-derived charges for the inner region. Inclusion of this long-range QM-MM electrostatic interaction is crucial for the viability of the method, and has lead to several embedding schemes [94, 101], described in the next paragraph.

The most basic embedding scheme is called *mechanical embedding*. In this approach, the electrostatic interaction is completely handled at the MM level. The charges of the inner region do not change during the chemical process, and have to be obtained consistently with the set of point charges of the forcefield

used for the real system. Obtaining these charges is not trivial, in particular because of the changes in chemical nature of the atoms expected during chemical reactions, which drastically can change the electrostatic potential in the inner region. This effect is completely absent in this embedding. Another problem is that the QM computation is done in gas phase, *i.e.* the QM computation does not take into account the external set of point charges.

Overcoming these problem can be conveniently done with an *electrostatic embedding*, where the QM-MM electrostatic interaction is handled by including terms in the QM Hamiltonian of the form:

$$E^{el}_{QM-MM} = -\sum_{i}^{N}\sum_{J\in\mathcal{O}}^{L}\frac{q_J}{|r_i - R_J|} + \sum_{a\in\mathcal{I}+\mathcal{L}}^{M}\sum_{J\in\mathcal{O}}^{L}\frac{q_J Q_a}{|R_a - R_J|}, \tag{2.60}$$

for a set of N electrons, L atoms in the outer region, and M atoms in the inner region ; q_J is the point charge of the MM atom located at R_J, and Q_a is the nuclear charge of the QM atom at R_a ; r_i is the position of electron i. The set of point charges q_J have been derived for a given forcefield, and is not designed to provide a faithful representation of the charge density, but rather to reproduce structural data. Thus, the use of such point charges in the QM Hamiltonian is questionnable. However, such implementation is common practice, given the simplicity of using an available set of point charges, and experience has shown that this approach yields reliable results [59, 94]. With this method, the QM wavefunction is effectively polarized by the external set of point charges.

The next step is to include the polarization of the charges in the outer region by the QM region. This can be done by using a flexible set of charges in the MM computation in a *polarizable embedding* approach, which can be further divided in models where the MM polarized charges do not act back on the QM charges, and models where the full charge model is obtained self-consistently. However, these models require general MM polarizable forcefield, which are currently not mainstream, although recent developments in this direction may change this picture in the future — namely, with the abvailability of polarizable solvent forcefields or the general AMOEBA [104] forcefield for biomolecular applications.

2.3.3 COBRAM implementation

The QM/MM potential we used thoughout our studies is implemented in the COBRAM [59] code. It consists in several Python scripts that combine computational chemistry softwares to efficiently use the ability of different codes for different property computations, and can be applied to static computations

— for instance, geometry optimization or conical intersection search — and molecular dynamics. For both types of studies, three steps are involved:

- the computation of the total energy E;
- the computation of the forces based on the first derivative of the energy;
- the construction of a new geometry based on energy and forces.

The steps for partitioning the system, treating the boundary region and computing the energy and forces are described in the following.

Partition of the system and boundary region

Partition of the system The same notations as introduced previously, in section 2.3.2, will be used here. Namely, the real system $\mathcal{R}$ is split in an inner region $\mathcal{I}$, treated at the QM level, and an outer region $\mathcal{O}$, treated at the MM level. If the partition goes through a covalent bond, the atom-link method is used to fill the valency, leading to a boundary region $\mathcal{L}$. The modularity of COBRAM allows for independent computations of energies and forces in each region. Once this step is completed, the information is processed internally for obtaining the information required to obtain a new geometry. In this step, a new partition is used, which does not necessarily correspond to the previous one. The larger system **Opt1** is either frozen or treated with a basic algorithm, such as steepest descent. The remaining parts of the system **Opt2/MD**, which comprise the $\mathcal{I} + \mathcal{L}$ region, but can also include elements from $\mathcal{O}$, will be moved with more accurate algorithms, such as BFGS (optimization) or Velocity Verlet (MD). This results in an effective three layer scheme, which can be summed up as (see figure 2.4):

1. *high* layer (H): computed at QM level and moved with advanced algorithm, and corresponds to the $\mathcal{I}$ region;
2. *low* layer (L): computed at the MM level and frozen or moved with basic algorithm;
3. *medium* layer (M): computed at the MM level and moved with advanced algorithm.

In this approach, large movement around the reactive center can be efficiently computed with almost no extra computational effort, since the energy and forces in the *medium* layer will be treated at the MM level. This method is thus able to take into account a polarization of the outer region by the QM region, and can vastly improve the results obtained from geometry optimization or molecular dynamics. The

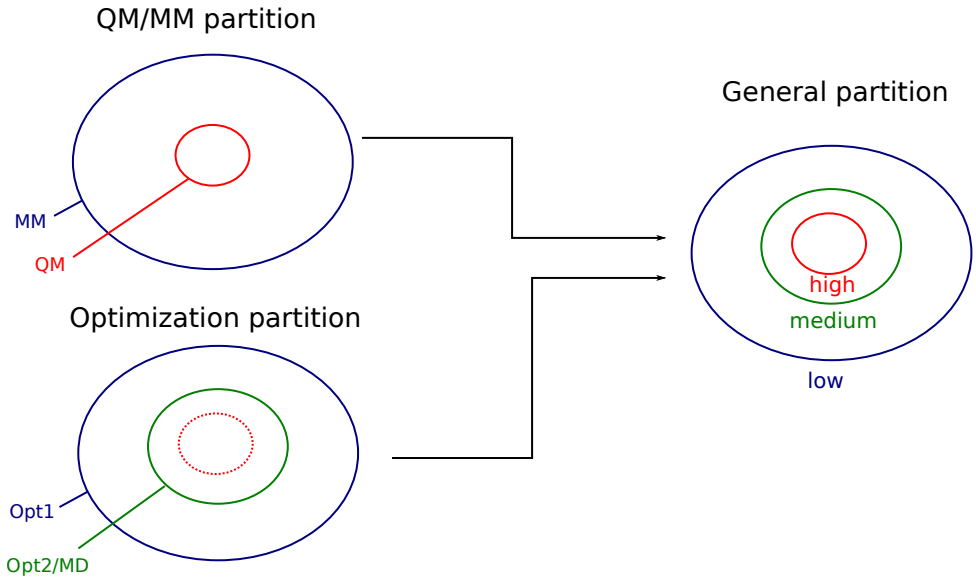


Figure 2.4: Partition scheme used in COBRAM.

layers can be combined according to the goal of the computational study. For instance, a chromophore in solvent can be treated in an HML scheme, where the chromophore is included in the high layer, and the closest solvent molecules are treated in the medium layer. The rest of the system is treated in the low layer. This also applies to the relaxation of a chromophore in a protein, where a covalently bound chromophore (namely, the retinal) is treated in the high layer, and the remaining part of the lysine residue is included in the medium layer, as well as fragments that are expected to interact with this chromophore. The rest of the protein can be kept frozen, and is treated in the low layer.

Boundary region The atom-link method is used when the partitioning goes through a covalent bond Q1-M1, as highlighted on figure 2.5. In that case, the valency in the QM region is filled by adding an hydrogen atom. The position of this hydrogen atom is constrained on the Q1-M1 bond, and the Q1-H distance is kept frozen at 1.09 Å, which is the standard value for a C(sp^3)-H bond. This method effectively removes the additional degrees of freedom from the additional H atom. The bond between Q1 and M1 is handled at the MM level.

To avoid the overpolarization at the Q1-H bond induced by the proximity of the MM M1 atom, the charge of M1 is redistributed on the neighbouring M2 atoms (see figure 2.5), and set to zero on M1. Each M2 atom receives a fractionnal charge that is proportional to the original MM charge hosted by this particular atom. Thus, two sets of MM charges, that differ only in the boundary region, are obtained:

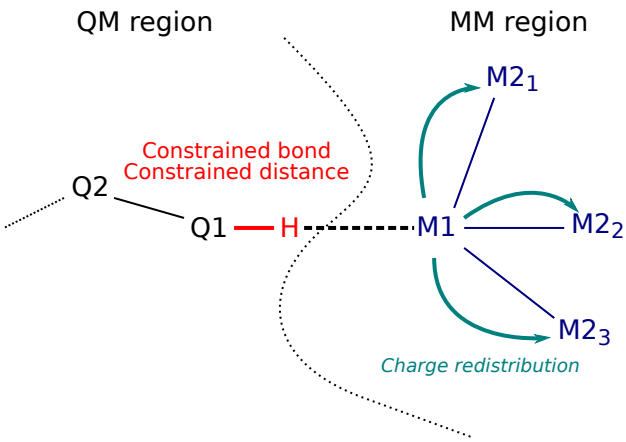


Figure 2.5: Boundary region and charge redistribution.

- the *pod* set refers to the original set of MM charges;
- the *emb* set refers to the modified set of MM charges.

The system is now fully described, and split into two or three layers. The next section deals with the computation of energies and forces.

Embedding scheme and QM/MM formalism

We use a substractive scheme with an electrostatic embedding. All energy components in the high layer $\mathcal{I}+\mathcal{L}$ are computed at the QM level, while all the energy components of the outer system $\mathcal{O}$ are computed at the MM level. The non-bonding van der Waals and bonded cross terms are all handled at the MM level, and the electrostatic interactions between QM and MM atoms is computed at the QM level. The expression of the total energy is then:

$$E^{\mathcal{R}}_{QM/MM} = E^{\mathcal{I}+\mathcal{L}}_{QM} + E^{\mathcal{R}}_{MM} + E^{el}_{QM}(\mathcal{I}+\mathcal{L}/emb) - E^{el}_{MM}(\mathcal{I}+\mathcal{L}/pod) - E^{\mathcal{I}+\mathcal{L}}_{MM}. \tag{2.61}$$

In this expression, $E^{\mathcal{I}+\mathcal{L}}_{QM} + E^{el}_{QM}(\mathcal{I}+\mathcal{L}/emb)$ corresponds to the energy of the high layer and the link atoms computed at the QM level embedded in the bath of the external modified MM charges (*emb* set), and are computed in a single step. Thus, the wavefunction is polarized by the external charges. The other terms are computed at the MM level, and can be seen as the MM energy of the *pod* set, plus the contribution coming from the van der Waals and bonded interactions between *pod* and the $\mathcal{I}$ region.

The analytical derivative of the total energy is used for the optimization of the high layer, while the medium layer is moved with pure MM computation, where the electrostatic interaction between QM and MM atoms is computed using atomic point charges derived from the QM wavefunction. The charge carried by the link atom is redistributed over the QM atoms to ensure that the total charge of the real system remains constant.

Practical implementation

The modular nature of COBRAM allows the use of different codes for computing the QM wavefunctions, the MM energies, and to handle the optimization or the dynamics of high/medium layers and low layer. Throughout our studies, the AMBER code [105] has been used for computing the MM energies, due to the availability of the GAFF [102] forcefield and the AmberTools [106] suite, where the Antechamber software [107] gives MM parameters that are fully compatible with other parametrized AMBER forcefields. Thus, AmberTools has been used to obtain the GAFF MM parameters of the retinal chromophore. The rest of the protein has been handled by the standard ff99SB forcefield [108].

As explained above, in section 2.2, treating the excited state of the retinal along the photoisomerization coordinate requires multi-reference wavefunctions (CASSCF), corrected by means of second order perturbation theory (CASPT2). This CASPT2//CASSCF (CASPT2 on top of CASSCF wavefunction) has been efficiently implemented in the Molcas 8 software [109]. Thus, unless otherwise stated, all QM computations will be handled by Molcas.

The energies and gradients are collected and processed by COBRAM accordingly to what has been said above (section 2.3.3). Geometry optimizations are handled by Gaussian [110] through the *external* function, where gradients previously computed are injected to the Gaussian optimizer. For semi-classical dynamics, Tully's fewest switches algorithm (section 2.1.3) is internally implemented, and combined with the Velocity Verlet algorithm.

3 Intramolecular photo-induced charge transfer in visual retinal chromophore mimics: electron density-based indices at the TD-DFT and post-HF levels

Published in Theoretical Chemistry Account

Link: https://link.springer.com/article/10.1007/s00214-016-1815-y

DOI: 10.1007/s00214-016-1815-y

Résumé Nous avons simulé la première transition brillante d'une série de dix molécules basées sur le rétinal en utilisant la TD-DFT, avec plusieurs fonctionnelles hybrides, et avec plusieurs méthodes post-Hartree-Fock. Les résultats CASPT2 ont été utilisés comme référence pour comparer les résultats. Nous avons comparé non seulement les énergies de transition, comme c'est généralement le cas dans ce type d'étude, mais aussi les variations de densité électronique depuis l'état fondamental vers l'état excité, et de plusieurs indices, liés à la densité, développés récemment. Parmis les résultats obtenus, il est clairement apparu que les corrections PT2 à la fonction d'onde CASSCF augmentent avec le caractère de transfert de charge de la transition. La littérature disponible indique déjà que la TD-DFT ne parvient pas à reproduire les énergies de transition de cette famille de molécules. Cette étude confirme cette observation pour la variation de densité électronique, différente entre TD-DFT et CASPT2. Néanmoins, certaines fonctionnelles donnent de meilleurs résultats que d'autres. Les fonctionnelles qui ont une fraction de 40-50 % d'échange exact donnent une meilleure description de la variation de densité électronique que les autres. De plus, la meilleure fonctionnelle pour le calcul d'énergies de transition n'est pas la meilleure pour simuler la densité électronique de l'état excité. Cela confirme une observation disponible dans la littérature pour une autre famille de molécules.

Abstract The first bright transition of a series of ten retinal-based dyes was simulated using TD-DFT, with several hybrid functionals, and with several post-Hartree–Fock methods. The CASPT2 results were used as reference to compare the results. The comparisons were not only based on the transitions energies, that is generally the case for such works, but were also based on the variation of the electron density from the ground to the excited along with several density-based indices recently developed. Among all the results obtained, it clearly appears that the PT2 correction to the CASSCF wavefunction is increasing with the increase in the charge transfer character of the transition. As already highlighted in the literature, TD-DFT poorly reproduces transition energies of this family of molecules whatever the functional. This work also confirms this observation for the variation of the electron density that is different between TD-DFT and CASPT2. Nevertheless, some functionals give better results than others. It appears that functionals having a fraction of 40–50 % of exact exchange give a better description of the electron density variation than the other functionals. Moreover, the best functional to compute the transition energies is not the best to simulate excited electron density. That confirms an observation done in the literature on a different family of molecules.

4 Fine tuning of the excited state properties of retinal chromophores in solution

Part of this work has been published in The Journal of Physical Chemistry Letters:

B. Demoulin, S. F. Altavilla, I. Rivalta, and M. Garavelli, Fine Tuning of Retinal Photoinduced Decay in Solution, *J. Phys. Chem. Lett.*, **2017**, *8* (18), pp 4407–4412

Link: http://pubs.acs.org/doi/10.1021/acs.jpclett.7b01780

DOI: 10.1021/acs.jpclett.7b01780

Résumé Nous avons étudié le comportement de la base de Schiff protonée du rétinal (PSB) et de son équivalent méthylé en position 10 (10Me-PSB) dans une solution de méthanol. En particulier, nous avons créé un modèle computationnel pour ces deux espèces, pour interpréter des données provenant d'expériences de spectroscopie ultra-rapide. Nous avons calculé les propriétés des états fondamentaux et excités, et nous avons montré qu'une interconversion entre deux géométries, peuplées de façon séquentielle durant la relaxation de l'état excité et caractérisées par différentes alternations de longueurs de liaisons dans la chaîne polyénique, est possible. Ces deux populations expliquent la présence de deux signaux d'émission stimulée observés expérimentalement pour chaque espèce. Une interpolation linéaire entre les deux géométries a montré que la transition vers une géométrie réactive présente une barrière énergétique de 3 kcal/mol pour PSB, alors que cette transition se fait sans barrière dans 10Me-PSB. Des trajectoires semi-classiques nous ont permis de modéliser l'évolution temporelle de l'état excité de chaque système, et nous avons montré que la méthylation accélère la population du mode de torsion, et accélère la décomposition de l'état excité par un mouvement de double-pédale asynchrone. Par ailleurs, nous avons démontré que la topologie de l'intersection conique est différente pour les deux systèmes. En effet, nous avons obtenu une intersection conique "en pointe" pour 10Me-PSB, ce qui indique une conversion de l'état excité à l'état fondamental plus rapide que dans PSB, où nous avons trouvé une intersection conique "en pente".

Abstract In this chapter we focus on the photochemistry of a native retinal protonated Schiff base (PSB) and its 10-methylated derivative (10Me-PSB) in methanol solution. In particular, we set up computational models for the two species, in order to interpret ultra-fast spectroscopy experiments data. The study involves the computation of both ground and excited- state properties, and we show that an interconversion between two geometries, populated sequentially during the excited state evolution and involving differences in the bond length alternation (BLA), is possible. This can account for the presence of two stimulated emission signals observed experimentally for both species. Constrained scans show that reaching a reactive geometry involves a barrier of 3 kcal/mol in PSB, whereas this transition is barrierless in 10Me-PSB. Semi-classical dynamics give an insight on the time evolution of the excited state dynamics, and the methylation accelerates the population of the torsion mode, allowing for a more effective excited state decay through an asynchronous double-bicycle pedal movement. We also show that the topology of the conical intersection (CI) is different in the two system, 10Me-PSB having a peaked CI, making the conversion from excited to ground state faster than in native PSB, where the CI appears sloped.

4.1 Introduction: experimental results

The tunability of vertical absorptions, isomerization yields, and excited state lifetime have made protonated Schiff bases (PSBs) become a paradigm to understand the parameters that allow for fast and effective photochemistry. In particular, an understanding of the steric and dielectric effects of binding pockets in retinal-containing proteins can be achieved by comparison between the photo-reactions that happen *in vivo* and those obtained with solvated PSBs. Ultra-fast pump-probe spectroscopy and fluorescence spectrocopy have been widely used to unravel the properties of excited states in proteins [111–114] as well as in solvated PSBs in different solvents [115–119]. Control of the photochemical and photophysical properties has also been investigated through synthetic modifications to the retinal backbone [120–123].

In particular, recent studies by Kukura et al. [121] have shown that adding a methyl group in position 10 on all-*trans* retinal (see figure 4.1) in solution makes the system have similar excited state lifetime as protein-embedded retinals, although with a reduction of the quantum yield.

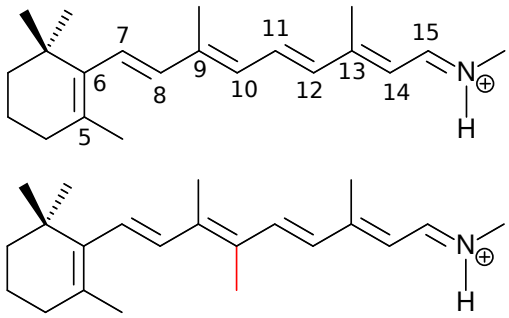


Figure 4.1: PSB (top) and 10Me-PSB (bottom) structures. The numbering used throughout this work is highlighted on the top molecule.

Ultra-fast pump-probe spectroscopy has been used to probe the excited state lifetime of the native (PSB) and 10-methylated retinal (10Me-PSB), and significant differences have been observed, as can be seen on figure 4.2.

In PSB, two stimulated emission (SE) signals are visible at 690 nm and 800 nm , with a biexponential decay with lifetimes of 2 ps and 8 ps. These signals are red-shifted in 10Me-PSB at 710 nm and 860 nm, with an excited state lifetime around 0.7 ps. The overall appearance of the stimulated emission (SE) peak has been attributed to an overlapping excited state absorption (ESA) signal peaking at 760 nm, based on previous studies by Ruhman et al. [117] The photoproducts were characterized to be only 11-*cis*

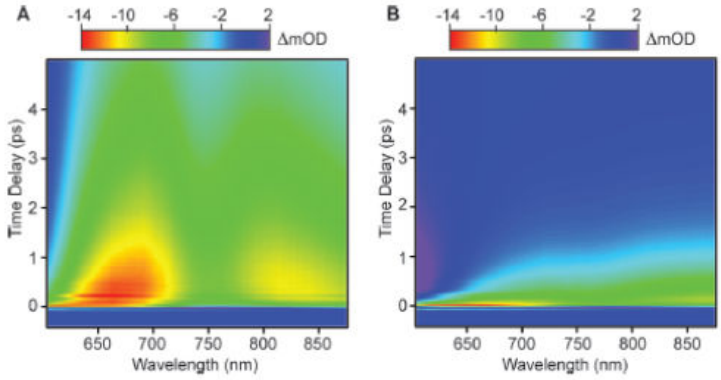


Figure 4.2: Experimental pump-probe experiments from Kukura et al. [121]: differential optical density (OD) obtained, with negative signals (in red) corresponding to stimulated emissions. (A) PSB ; (B) 10Me-PSB.

molecules, although with low quantum yields (0.16 and 0.09 for PSB and 10Me-PSB respectively), thus demonstrating that fast reaction does not necessarily correlates with high efficiency, as was previously believed . The low quantum yield observed was attibuted to a conical intersection (CI) located on the reactant side, making the isomerization less likely to happen. Going further on this matter, a systematic screening of methylation and de-methylation, and of Schiff base substitution was carried out, finding that opsin shift and suppression of isomerization channels (Shiff base substitution), or tuning of quantum yields, excited state lifetimes and opening of isomerization channels could be achieved [122]. Eventually, the methylation at position 10 has also been done on 11-*cis* PSB, leading to the same decrease in excited state lifetime, but no effect on the quantum yield, indicating that quantum yields may be an intrinsic property of PSBs, which depend on the chosen conformer ; thus, the role of the protein would be to select exclusively the most reactive conformer, leading to high quantum yields in the reaction *in vivo* [123].

In this chapter, we will focus on the modelling of the ground and excited state properties of the all-*trans* native PSB and 10Me-PSB. In particular, we are interested in accouting for the appearance of the pump-probe spectra, the different excited state lifetimes, and the low quantum yields observed. This model could serve as a basis for further investigation on similar solvated systems.

4.2 Setup of the computational models

4.2.1 Classical sampling

To obtain accurate representative starting geometries and arrangements of solvent molecules around the retinal chromophore, classical molecular dynamics sampling has been performed for both PSB and 10Me-PSB using the Amber 11 suite [105], with the ff99SB force field parameters [108].

The PSB and 10Me-PSB were first optimized at the CASSCF level [40], with an active space comprising the full π system of the molecules (*i.e.* 12 electrons in 12 orbitals) in gas phase. Their first excited state (S_1 bright state, corresponding to the $H \rightarrow L$ transition) has also been optimized. The resulting excited state structure was then solvated in a box of explicit methanol molecules. Parameters for the chromophore have been obtained with Antechamber [107], using the GAFF force field [102]. As we are mostly interested in the behaviour of the excited state of the molecules, the MM charges were set to be the Mulliken charges computed at the CASSCF level. The chromophore has been kept frozen to its gas phase excited state geometry during all molecular dynamics computations.

The system has been heated from 0 to 300 K for 1 ns at constant volume and constant pressure (1 atm). Then, a 50 ns production run was produced, with the frames recorded every 200 fs.

4.2.2 QM/MM models

Out of the classical MM sampling, the snapshot with the lowest total energy has been selected for further studies. The COBRAM [59] interface allows a multi-scale description of the system, combining a high layer (QM level), a medium layer (movable layer, MM) and a low layer (frozen, MM) (see section 2.3.3 in chapter 2). To reduce the computational cost, only atoms involved in the conjugated chain of the retinal molecule were considered in the high layer, while the remaining atoms of the β-ionone ring were included in the movable medium layer. As solvent molecules are expected to play a role in the excited state properties of the retinal molecule, some of them have to be included in this movable medium layer. This is especially important when it comes to constrained scans, and particularly when probing for possible isomerization (see section 4.4.3). For comparison to be possible between the two system, similar medium layer size for both PSB and 10Me-PSB are mandatory. The main problem is that subtle changes in the hydrogen bonding network (HBN) around the retinal may have huge impact on the MM energy computed, and thus lead to discontinuities during optimizations.

To assess the size of the medium layer, preliminary scans around the $C_{11}-C_{12}$ bonds have been carried out with a large number of solvent molecules included in the medium layer. Structural comparisons have been made on various point of the energy surface obtained, and concluded that 7 or 8 solvent molecules are most affected. This analysis was carried out for both PSB and 10Me-PSB, and the solvent molecules that undergo the biggest changes have been selected for inclusion in the medium layer. The resulting systems used are displayed on figure 4.3.

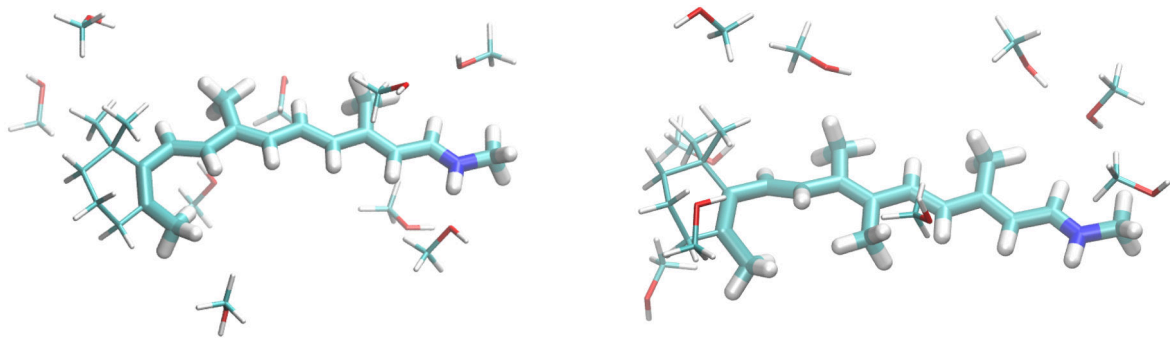


Figure 4.3: Layer partitionning of the two PSB systems considered. The High layer (QM part) is represented with large sticks, while the movable Medium layer (MM) is represented with thinner lines. Low layer (MM) solvent molecules are omitted for clarity. Left: Native retinal ; Right: Methylated retinal.

4.2.3 Computational details

All the geometry optimizations have been carried out at the CASSCF level, with an active space comprising the whole π system of the molecule, resulting in 12 electrons in 12 orbitals. Unless otherwise stated, the first three states (*i.e.* the ground and the first two excited states) were considered in the state average. All optimizations have been handled by the COBRAM interface, where energies and gradients are computed by Molcas 8 [109], and the solvent atoms are modelled with the ff99SB parameters of Amber [108]. Gradients are then injected in the Gaussian optimizer [110]. The CASSCF energies are corrected perturbatively with dynamical correlation within the CASPT2 framework. To avoid the problem of intruder states, a so-called “imaginary shift” of 0.2 was used. This value is widely used in similar computation. We set the IPEA value to 0, as it was recently shown shown [90] that this parameter has little to no influence for computing vertical excitations on organic chromophores. The CASPT2 energies were then used for computing oscillator strengths. Both single- and multi-state (SS

and MS) CASPT2 energies have been computed throughout the study. The 6-31G* basis set was used for all the computations.

Constrained scans have been performed using frozen redundant coordinates of Gaussian 09. For conical intersection (CI) search, the projected gradient method of Bearpark et al. [67] has been used, as implemented in COBRAM. Excited state absorptions have been computed using SA20-CASSCF wavefunctions.

Semi-classical dynamics have been performed with the Velocity Verlet algorithm as implemented in COBRAM. Tully's fewest switch algorithm [68] has been disabled for all trajectories computed.

4.3 Geometry optimizations

4.3.1 Ground state geometry and vertical absorptions

The geometrical parameters obtained after geometry optimizations are displayed on figures 4.4. The bond length alternation pattern obtained is the typical one expected after a CASSCF optimization for conjugated polyenes, and is similar for both systems, although sligthly more pronounced for PSB. The presence of of a methyl group seems to induce some distorsions near the ionone ring, in the region between C_8 and C_{10} for 10Me-PSB, making it less planar than PSB. This effect may be due to the proximity of a methyl group of the ionone ring (see figure 4.1), which leads to an unfavorable planar conformation. Thus, this distorsion at this position is not unexpected for 10Me-PSB.

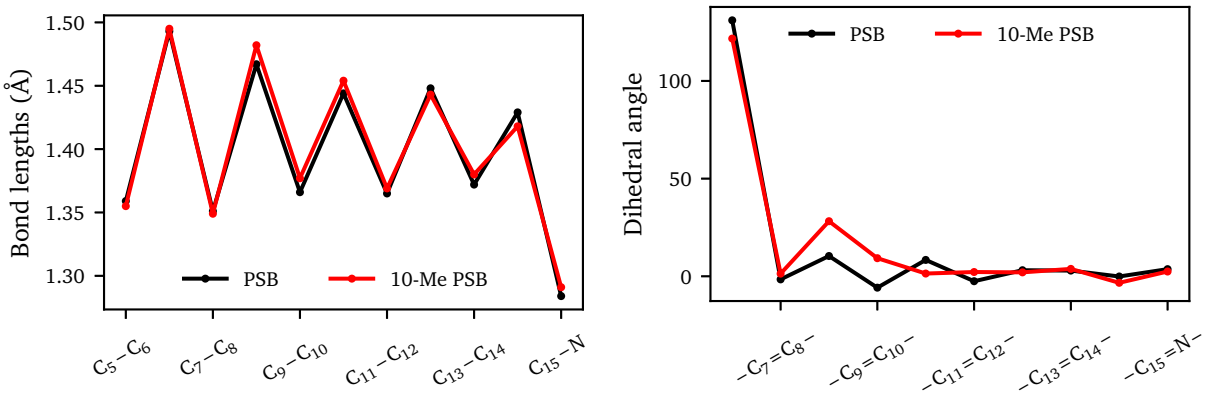


Figure 4.4: Geometrical parameters for the ground state geometries. Left: bond length alternation (BLA) ; right: dihedral angles. The bottom axis is labelled with names from 4.1.

On top of these optimized geometries the vertical absorption have been computed at the CASPT2 level, both in its single-state (SS) and multi-state (MS) formalisms. The results for SS-CASPT2 are presented on table 4.1. We found vertical absorptions of 465.3 nm (2.66 eV) and 497.58 nm (2.49 eV) for PSB and 10Me-PSB respectively. These values are in reasonable agreement with the experiments reported by Kukura et al. (2.78 eV and 2.76 eV for PSB and 10Me-PSB respectively). The difference may be attributed to the different Schiff base linkage, as experiments were conducted with *n*-Butylamine, whereas a CH_3 group was used in our model to reduce computational cost. Furthermore, the selection of one snapshot only can bias the result, as slight changes in the H-bonding network around the retinal may have sensitive effect on vertical transitions, with differential effects on the covalent S_0 state and the ionic S_1 state. However, the red-shift observed for 10Me-PSB is reproduced by our computation.

For 10Me-PSB, the wavefunction is cleaner than for PSB. Indeed, 10Me-PSB displays an S_1 state with a clear $H \rightarrow L$ character, and a S_2 state dominated by the double excitation $(H \rightarrow L)^2$ and the single excitation $H-1 \rightarrow L$. However, in solvated native PSB, the S_1 and S_2 states, although well separated in energy (they are separated by 0.83 eV), have similar character. Indeed, both are dominated by the single $H \rightarrow L$ excitation, followed by double $(H \rightarrow L)^2$ excitation contribution. The single excitation character is slightly more pronounced in S_1 than in S_2 (0.28 and 0.32 respectively). This may explain the higher oscillator strength found for S_2 with respect to S_1 (0.59 and 0.55 respectively). Given this mixing, an accurate assignment of the spectrocopic state is hard to perform. To get rid of the possible mixing, we carried out computations by state-averaging the ground and the first excited state only (SA2). This results in an S_1 state clearly dominated by the $H \rightarrow L$ excitation, and a vertical absorption of 424 nm (2.92 eV), still in reasonable agreement with the experimental value.

However, inspection of the difference in permanent dipole moments for SA3 and SA2 wavefunctions for both PSB and 10Me-PSB (see table 4.2) shows that the computed SA2-S_1 state in PSB is very similar to the SA3-S_2 state. The difference dipole moment has lower intensity for these two states when compared to the SA3-S_1 state. Comparison can also be made between PSB and 10-Me PSB. In the case of 10Me-PSB, SA2 and SA3 computations result in similar S_1 state, with a high change in dipole moment. Previous computational study on the retinal have proven that the $S_0 \rightarrow S_1$ transition is characterized by a high change in the permanent dipole moment [26, 31]. As a consequence, we think that the bright state, corresponding to a $H \rightarrow L$ transition and a sensitive change in the permanent dipole moment should be assigned to S_1 in the SA3 computation.

	Native			Methylated		
State	Nature	λ_{max}	f	Nature	λ_{max}	f
S_0	GS 0.68			GS 0.71		
S_1	$H \rightarrow L$ 0.28 $(H \rightarrow L)^2$ 0.19 $H \rightarrow L+1$ 0.12	465.30	0.55	$H \rightarrow L$ 0.54	497.58	0.97
S_2	$H \rightarrow L$ 0.32 $(H \rightarrow L)^2$ 0.13 $H-1 \rightarrow L$ 0.10	355.06	0.59	$(H \rightarrow L)^2$ 0.26 $H-1 \rightarrow L$ 0.20	333.28	0.18

Table 4.1: Summary of ground state transition energies. λ_{max} alues are reported in nm. Only the configurations that contribute for more than 0.095% to the wavefunction are presented.

	Native		Methylated	
Transition	$\Delta\mu$ (SA3)	$\Delta\mu$ (SA2)	$\Delta\mu$ (SA3)	$\Delta\mu$ (SA2)
$S_0 \rightarrow S_1$	12.995	7.985	16.499	16.471
$S_0 \rightarrow S_2$	7.965	-	2.586	-

Table 4.2: Difference in dipolar moment for the different setups

As a last note, we can notice that MS-CASPT2 gives a more accurate representation of the PSB ground state. Indeed, the S_1 gets a clear $H \rightarrow L$ character (weight 0.62), and S_2 becomes a mixed $(H \rightarrow L)^2$ (0.28) and $H-1 \rightarrow L$ (0.17) excitations. The $S_0 \rightarrow S_1$ transition has an energy of 2.62 eV (472 nm), with an oscillator strength of 1.69. The value of the transition energy is in line with the one found in SS-CASPT2 (2.66 eV), which strengthen our assumption to take the SS-CASPT2 S_1 as the bright state.

4.3.2 Excited state minima

10Me-PSB Starting from the ground state optimized structure, the bright S_1 state has been optimized at the SA3-CASSCF level. After relaxation, the first excited state exhibits inverted bond length with respect to the ground state (see figure 4.5) from C_7 to C_{15}. This geometry will be refered to as ABL (for Alternated Bond Length) in the following. We can notice that the C_9-C_{10} and $C_{11}-C_{12}$ bonds are the most elongated, making a rotation likely to occur in these regions. The S_1 wavefunction is dominated by the single $H \rightarrow L$ transition (0.48), with contributions from GS configuration (0.21). The computed vertical emission was found around 824 nm (1.50 eV), in line with what is found in experiments. However, it should be noticed

that the experimental emission is a broad signal mixed with ESA, making the assignment of a peak more difficult.

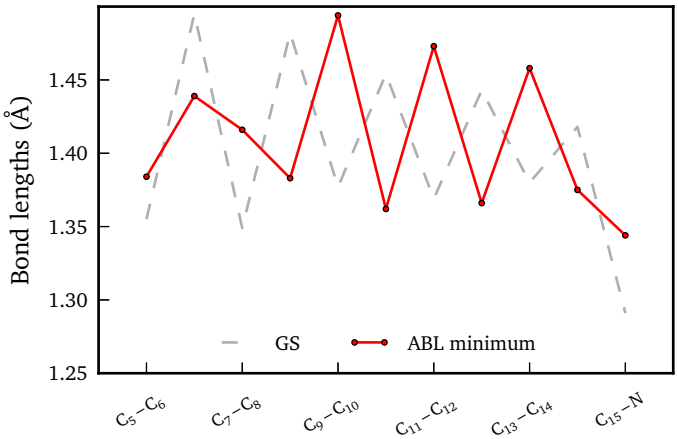


Figure 4.5: Comparisons of the bond length alternations of ground and excited state for 10Me-PSB.

PSB This case has been more problematic. As stated before (section 4.3.1), the S_1 and S_2 states are mixed at the GS geometry of PSB. Thus, several excited state optimizations were carried out from this point:

- SA3-CASSCF optimization of S_1, referred to as **SA3-S1**;
- SA3-CASSCF optimization of S_2, referred to as **SA3-S2**;
- SA2-CASSCF optimization of S_1, referred to as **SA2-S1**.

Both **SA3-S1** and **SA2-S1** display similar behaviour, and after a very few optimization steps the optimized state is already dominated by the double $(H \rightarrow L)^2$ excitation. After full optimization, an ABL-like geometry is obtained, but with a lower alternation than the one found in 10Me-PSB. This geometry will be referred to as **ABL*** (see figure 4.6). At these optimized points, as seen on figure 4.7, there is a near degeneracy between the $H \rightarrow L$ and $(H \rightarrow L)^2$ states.

On the contrary, the **SA3-S2** optimization leads to a geometry where all the bonds have a similar length. This geometry will be referred to as **EBL**. The minimum found is dominated by the $H \rightarrow L$ excitation.

To further investigate on the energy ordering in these geometries, CASPT2 energies were computed, and are reported on figure 4.7. For **SA2-S1**, a SA5-CASSCF computation was required to find the root

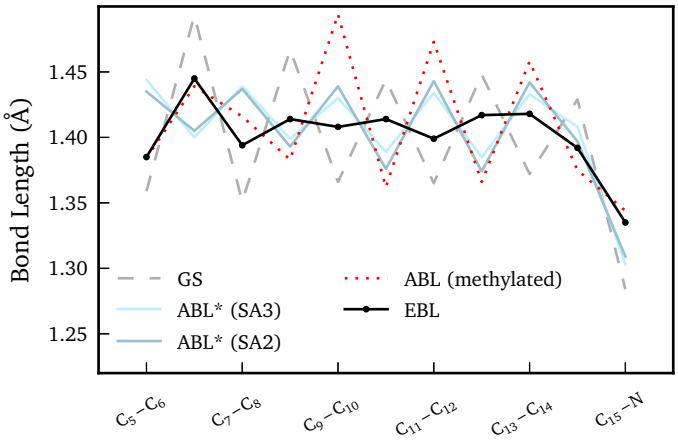


Figure 4.6: Comparison of the bond length alternation patterns obtained for several excited state optimizations carried out on PSB. The ABL geometry from 10Me-PSB is included for reference.

dominated by the single $H \rightarrow L$ excitation (located as S_2). On this picture, the lowest minimum is found to be at the EBL geometry, found after optimizing **SA3-S2**. Thus, this geometry should be the true minimum in the excited state surface, and features an emission computed at 2.04 eV (607 nm). This point will be further analyzed in the following section (see section 4.4).

4.3.3 Summary

The ground and excited state geometries of both PSB and 10Me-PSB have been characterized at the CASSCF level. Although the two systems have similar geometries in the ground state, they differ by their electronic structures. Indeed, two clearly separated excited states are found for 10Me-PSB, whereas S_1 and S_2 are strongly mixed in PSB. Optimizing the bright excited state further extends this difference, and two different minima are characterized. In 10Me-PSB, a geometry where the bond lengths are inverted with respect to those of the ground state was found (ABL geometry), whereas a geometry were all bond lengths are equalized was found in PSB (EBL geometry). In the following section, we will study the possible interconversion between these two geometries in both systems, and investigate on the photoisomerization pathways.

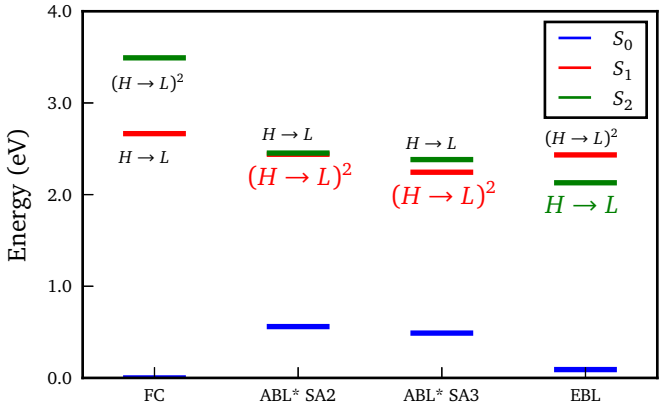


Figure 4.7: Ordering of states (CASPT2 energies) at the different minima found after geometry optimization. The state highlighted in color and bigger font is the one that has been optimized. SA3-CASSCF wavefunctions were used for all points except ABL*SA2, where SA5-CASSCF was required.

4.4 Skeletal relaxation and conical intersection

4.4.1 Scan from EBL to ABL geometries

In this section, we use the CASSCF optimized geometries of 10Me-PSB (ABL) and PSB (EBL) as starting point. For both systems, another geometry is built by modifying the bond length alternation of the two minima, replacing the original optimized ones by those coming from the other minimum (*i.e.* an EBL is built for 10Me-PSB, and an ABL is built for PSB). By linear interpolation, five intermediate geometries are produced between the two limit geometries. All points are then optimized at the CASSCF level, with frozen bond lengths in the conjugated chain.

10Me-PSB The results obtained for 10Me-PSB are presented on figure 4.8. The optimization has been carried out on the S_1 state, which is the bright $H \to L$ state for all points. The CASSCF and CASPT2 surfaces are qualitatively similar, and show a steady decrease in energy in the first excited state (line with round markers), which is consistent with the results of the previous optimization (see section 4.3.2). The EBL emission signal computed at the CASPT2 level is found at 654.6 nm (1.89 eV). The wavefunctions obtained are recorded in table 4.3.

As can be seen in this table, the $H \to L$ character of S_1 decreases along the scan. This may be due to the increased GS contribution, resulting from S_0 and S_1 coming closer in energy. This mixing is also

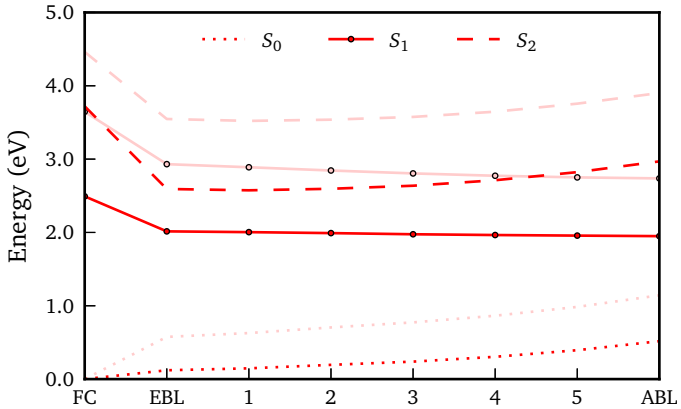


Figure 4.8: 10Me-PSB CASSCF (pink) and CASPT2 (red) energy surfaces. The labellings S_0, S_1 and S_2 refer to CASSCF states ordering.

present, obviously, in the evolution of the S_0 wavefunction. The S_2 state loses its dominant $(H \rightarrow L)^2$ character, with the $H \rightarrow L+1$ transition becoming its principal component.

	FC	EBL	1	2	3	4	5	ABL
S0	GS 0.71	GS 0.69	GS 0.68	GS 0.68 $H \rightarrow L$ 0.10	GS 0.64 $H \rightarrow L$ 0.12	GS 0.63 $H \rightarrow L$ 0.13	GS 0.61 $H \rightarrow L$ 0.15	GS 0.59 $H \rightarrow L$ 0.18
S1	$H \rightarrow L$ 0.54	$H \rightarrow L$ 0.58 GS 0.10	$H \rightarrow L$ 0.56 GS 0.12	$H \rightarrow L$ 0.53 GS 0.14	$H \rightarrow L$ 0.50 GS 0.16	$H \rightarrow L$ 0.48 GS 0.17	$H \rightarrow L$ 0.46 GS 0.19	$H \rightarrow L$ 0.44 GS 0.21
S2	$(H \rightarrow L)^2$ 0.26 $H-1 \rightarrow L$ 0.20	$(H \rightarrow L)^2$ 0.31 $H-1 \rightarrow L$ 0.17 $H \rightarrow L+1$ 0.16	$(H \rightarrow L)^2$ 0.29 $H \rightarrow L+1$ 0.17 $H-1 \rightarrow L$ 0.16	$(H \rightarrow L)^2$ 0.27 $H \rightarrow L+1$ 0.17 $H-1 \rightarrow L$ 0.15	$(H \rightarrow L)^2$ 0.24 $H \rightarrow L+1$ 0.17 $H-1 \rightarrow L$ 0.15	$(H \rightarrow L)^2$ 0.22 $H \rightarrow L+1$ 0.17 $H-1 \rightarrow L$ 0.15	$(H \rightarrow L)^2$ 0.19 $H \rightarrow L+1$ 0.17 $H-1 \rightarrow L$ 0.15	$H \rightarrow L+1$ 0.17 $(H \rightarrow L)^2$ 0.16 $H-1 \rightarrow L$ 0.15

Table 4.3: Methylated retinal: CASSCF wavefunction character along the MEP scan. Only the CSFs with a contribution higher than 0.095 are shown.

PSB As highlighted above, this case is more complicated when it comes to the choice of the state to optimize. We found that for each geometry, at the CASSCF level, the bright state is always S_2, even when this state has a $H \rightarrow L$ component lower than the one of S_1. For consistency, we decided to always optimize the state with the highest oscillator strength. The resulting CASSCF and CASPT2 energy surfaces are displayed on figure 4.9. As can be seen, the two surfaces differ significantly. At the CASSCF level, the energy of the optimized S_2 state increases from the EBL point, consistently with the results of the excited state optimization (section 4.3.2). However, when corrected with dynamical correlation at the CASPT2 level, another picture appears. Indeed, a surface crossing appears between the geometries **2** and

3. We can notice that at these points, we have similar bond lengths as in the ABL* geometry obtained before, where we found that the configurations dominated by the $H \rightarrow L$ and the $(H \rightarrow L)^2$ excitations are degenerated. The vertical emission obtained out of the ABL geometry has been computed at 1.59 eV (780 nm).

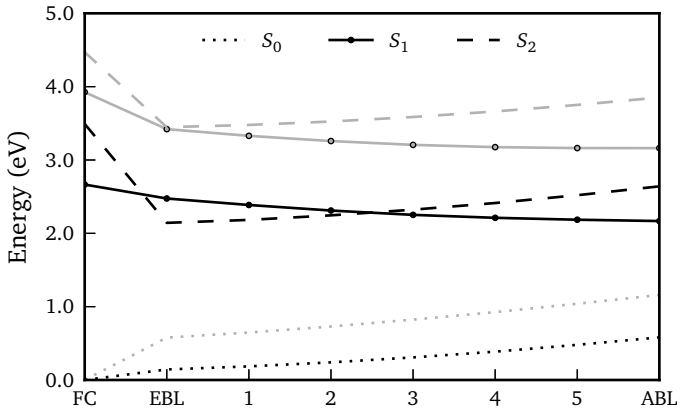


Figure 4.9: PSB CASSCF (gray) and CASPT2 (black) energy surfaces. The labellings S_0, S_1 and S_2 refer to CASSCF states ordering.

The wavefunctions obtained along the scan are displayed in table 4.4. All along the scan, S_1 and S_2 have similar configurations, dominated by the $H \rightarrow L$ excitation. The main difference is that the double excitation character disappears in S_1 around point **4**, while a GS character becomes more significant. In S_2, the double excitation character does not disappear, but loses its importance to other contributions. We can notice that the lowest excited state (at CASPT2 level) for all the points is the one where the $H \rightarrow L$ character is the highest, while this is not necessarily the one with the highest oscillator strength. These results show that the conversion from EBL to ABL has a barrier of 0.13 eV (around 3 kcal/mol) associated with a mixing of the S_1 and S_2 states and a typical avoided crossing, which can trap the system in the non reactive EBL geometry. This is consistent with the longer excited state lifetime observed in PSB than in 10Me-PSB.

4.4.2 Analysis of the different minima

Comparison of the MEP The two paths computed exhibit sensitive differences, and a summary picture can be seen in figure 4.10 which shows the overlapped CASPT2 surfaces. In 10Me-PSB, the S_1 surface decreases steadily from EBL to ABL, leading to a fast transition between the two, and thus

	FC	EBL	1	2	3	4	5	ABL
S0	GS 0.68	GS 0.59 $H \to L$ 0.14	GS 0.57 $H \to L$ 0.16	GS 0.55 $H \to L$ 0.18	GS 0.52 $H \to L$0.20	GS 0.50 $H \to L$ 0.23	GS 0.47 $H \to L$ 0.25	GS 0.44 $H \to L$ 0.28
S1	$H \to L$ 0.28 $(H \to L)^2$ 0.19 $H \to L+1$ 0.12	$H \to L$ 0.21 $(H \to L)^2$ 0.16 $H-1 \to L$ 0.11 GS 0.11	$H \to L$ 0.21 $(H \to L)^2$ 0.15 GS 0.13 $H-1 \to L$ 0.11	$H \to L$ 0.22 GS 0.15 $(H \to L)^2$ 0.14 $H-1 \to L$ 0.10	$H \to L$ 0.22 GS 0.18 $(H \to L)^2$ 0.12 $H-1 \to L$ 0.10	$H \to L$ 0.22 GS 0.21 $(H \to L)^2$ 0.11	$H \to L$ 0.24 GS 0.22	GS 0.27 $H \to L$ 0.21
S2	$H \to L$ 0.32 $(H \to L)^2$ 0.13 $H-1 \to L$ 0.10	$H \to L$ 0.29 $(H \to L)^2$ 0.14 $H \to L+1$ 0.10	$H \to L$ 0.27 $(H \to L)^2$ 0.14 $H \to L+1$ 0.11	$H \to L$ 0.24 $(H \to L)^2$ 0.14 $H \to L+1$ 0.11	$H \to L$ 0.21 $(H \to L)^2$ 0.14 $H \to L+1$ 0.11	$H \to L$ 0.19 $(H \to L)^2$ 0.14 $H \to L+1$ 0.11 GS 0.10	$H \to L$ 0.17 $(H \to L)^2$ 0.13 $H \to L+1$ 0.13 GS 0.10	$H \to L$ 0.15 $H \to L+1$ 0.13 $(H \to L)^2$ 0.13 GS 0.10

Table 4.4: Native retinal: CASSCF wavefunction character along the MEP scan. Only the CSFs with a contribution higher than 0.095 are shown.

allowing a faster transition from the FC region to a geometry that enables the possibility of barrierless isomerization. On the contrary, in PSB, the molecule is trapped in the EBL geometry, and has to overcome a barrier to reach the reactive geometry. This barrier arises from the mixing between the two lowest exited states, which are close in energy in the whole scan, and share similar configurations. This barrier is consistent with the findings of figure 4.7, where the BLA found in ABL* geometries can be seen as an intermediates between the EBL and ABL ones. On these geometries, we found that the $H \to L$ excitation is higher in energy than in the EBL geometry. To reach an ABL state, the system has to pass through ABL*-like structures, which is in line with the findings of this section.

Mulliken charges analysis A further analysis of the difference between the two systems can be drawn from looking at the Mulliken charges for the different geometries obtained. This analysis is computed by summing up the Mulliken charges collected out of a CASSCF computation on two sides of the retinal, separated by cutting the $C_{11}-C_{12}$ bond. The two parts will be reffered to as the *ionone side* for the part going from C_5 to C_{11} and the *nitrogen side* for the remaining atoms.

The results are presented on table 4.5. In the FC geometry, the charge is more localized on the nitrogen side in PSB than in 10Me-PSB. When exciting from S_0, the transfer is more effective in 10Me-PSB, as around 50% of the charge is transferred, whereas only 29% of the charge is transferred upon excitation in PSB. Thus, the presence of the methyl group in position 10 helps to stabilize the charge on the ionone ring side, and improves the charge transfer between the two sides of the retinal, leading to a geometry where most of the charge is on the ionone ring side. On the contrary, in PSB, in all the geometries studied, the charge stays mainly localized on the nitrogen side of the molecule.

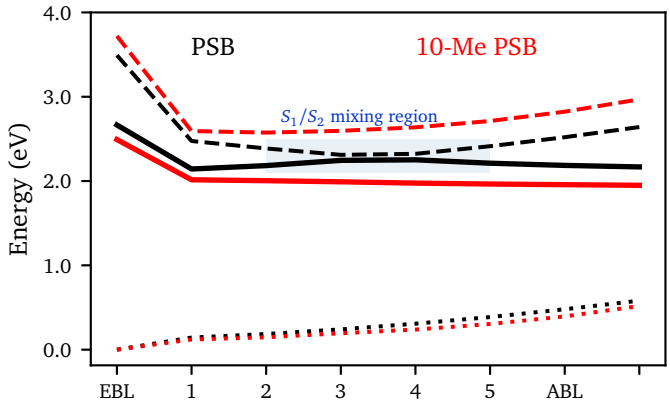


Figure 4.10: Path from EBL to ABL at the CASPT2 level for PSB (black) and 10Me-PSB (red). Dotted line: S_0 ; plain line: S_1 ; dashed line: S_2. The mixing zone for PSB is highlighted in the blue rectangle.

Table 4.5: Mulliken charges summed up on the two sides of the molecule for 10Me-PSB and PSB, for the different geometries computed.

System	Side	FC			EBL			ABL		
		S_0	S_1	S_2	S_0	S_1	S_2	S_0	S_1	S_2
10Me-PSB	Nitrogen	0.79	0.41	0.76	0.63	0.35	0.73	0.59	0.43	0.72
	Ionone	0.21	0.59	0.24	0.34	0.65	0.27	0.41	0.57	0.28
PSB	Nitrogen	0.87	0.62	0.65	0.79	0.66	0.51	0.79	0.56	0.51
	Ionone	0.13	0.38	0.35	0.21	0.34	0.49	0.21	0.44	0.39

The observation of the exact repartition of the charge is also useful to better understand the effect of methylation at position C_{10}. For this, we consider only the atoms of the conjugated chain of the retinal molecule, where the charge of each methyl group has been summed with the one of the carbon that bears that group. The charge of C_ϵ has been summed with the one of N. In particular, we highlight here the differences in charge repartition in the spectroscopic state (with highest oscillator strength) at each geometry studied for the two systems. The repartition of charges is highlighted on figure 4.11.

At the FC geometry, we can notice that in PSB, the charge is mostly located on the $C_{15}-N$ part of the molecule, but that it also extends on the ionone side. The $C_{11}-C_{12}$ bond also appears polarized, with a + charge on C_{12} and a - charge on C_{11}. After relaxation, at the EBL geometry, the picture is quite

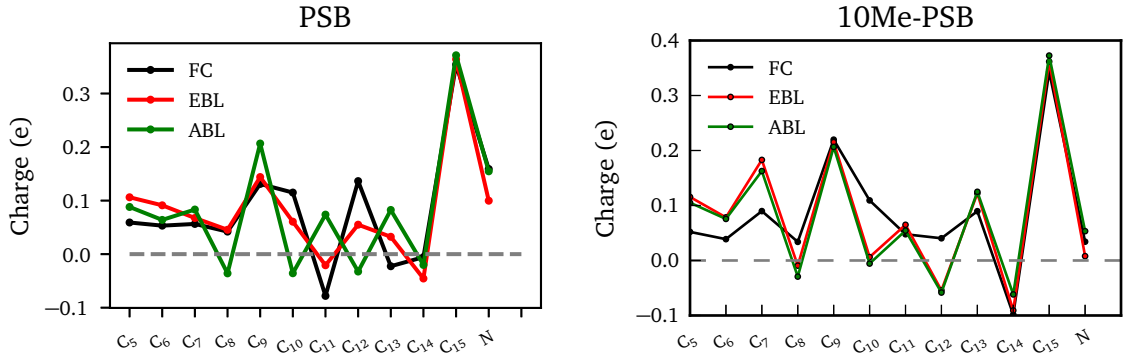


Figure 4.11: Mulliken charges along the conjugated side chain for PSB (left panel) and 10Me-PSB (right panel).

different, especially in the region between C_{14} and C_8. Indeed, the polarization highlighted before has been quenched, and the charges on C_{12} and C_{10} have been considerably reduced, while on C_{13} and C_{11} it has increased. This equalization of charges is expected, as all the bonds in this region now are comparable in size. At the ABL geometry, we have a strong polarization of all the bonds, and we can notice that the polarization of $C_{11}-C_{12}$ has been reversed compared to the FC geometry. Most changes occur, here again, in the central region of the molecule (from C_{14} to C_8). Again, this is expected, given the alternated bond lengths obtained in this region. Therefore, the EBL geometry can be interpreted as a transition structure between the FC and ABL geometries. The same kind of study can be carried out for 10Me-PSB, and useful information come out of the comparison of trends between the two cases. At the FC geometry, the polarization of $C_{11}-C_{12}$ completely disappears, and the positive tail observed in native PSB now extends from C_{13} to C_5 without interruption, with a bump at the C_9 position. It is noticable that the methylation does not affect the positions C_{10} and C_8, which is quite counter-intuitive ; instead, a huge effect is seen at odd positions on the chain, where the charge systematically increases. At the EBL geometry, we have a picture that is quite close to the one obtained at the ABL geometry for native PSB, with fully polarized bonds. The discrepancy between this charge repartition and the geometry of the EBL point could be the reason why this geometry is unstable in 10Me-PSB: at this geometry, the charge distribution is already the one expected for an alternated structure. Thus, contrary to native PSB, no intermediate structure can be stabilized, and the molecule is directly driven to the reactive ABL geometry. Finally, the comparison of the ABL points for native and 10Me-PSB shows that the two are pretty similar, except for a bump in charges at position C_7 in 10-Me PSB. This atom is the last one where the structural

changes affecting the conjugated chain are noticeable. An interpretation of these results seems to be that the EBL state can be destabilized by allowing an effective charge transfer toward the ionone ring. As the ABL geometry displays the bond inversion that allow the isomerization, this can lead to lower excited state lifetimes. Mulliken charges are known to be affected by the choice of the basis set. While here we report the results obtained using the 6-31G* basis set, we have observed the same trend (and thus draw the same conclusions) with the ANO-L (VDZP) basis set.

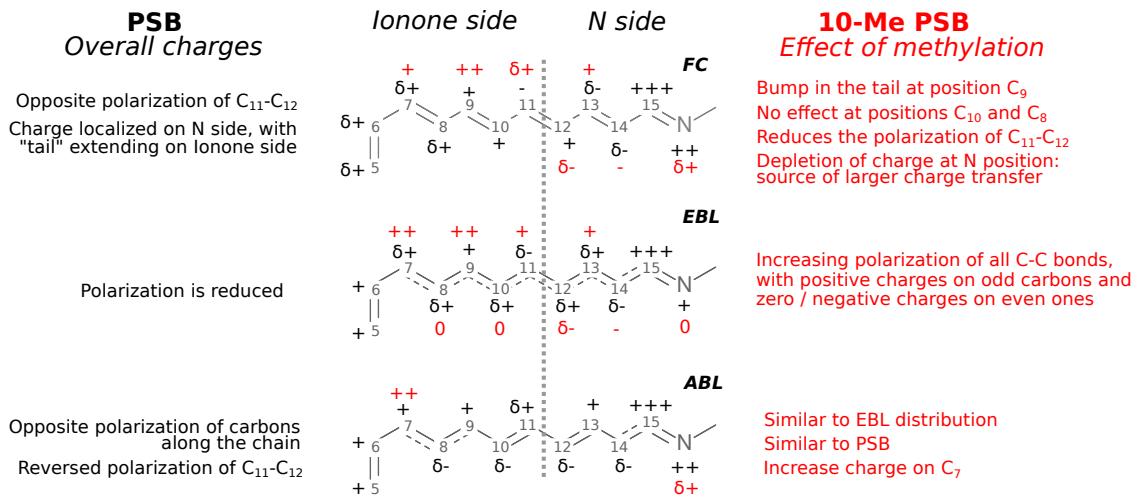


Figure 4.12: Qualitative description of the charges on the PSB backbone (in black) and of the modifications induced by the methylation at position C_{10} (in red).

MS-CASPT2 results For reference, we also include the MS-CASPT2 energies computed along the scan in figure 4.13. As expected, the S_1/S_2 mixing disappears in PSB, but the scan looks qualitatively similar in 10Me-PSB. However, MS-CASPT2 does not offer a good agreement when it comes to vertical emissions, as the EBL and ABL geometries now have similar signals (2.23 eV and 2.03 eV for EBL and ABL respectively) in PSB, and 2.23 eV and 1.98 eV resp. in 10Me-PSB), while it is clear from experimental data that the two signal are much separated. The MS-CASPT2 treatment is not recommended when the off-diagonal elements of the multistate PT2 Hamiltonian are too large [124], and in our cases they have been found to be an order of magnitude larger than the upper limit (0.002 a.u.).

Excited state absorption (ESA) The excited absorption has also been computed for all the minima. The results are gathered on table 4.6.

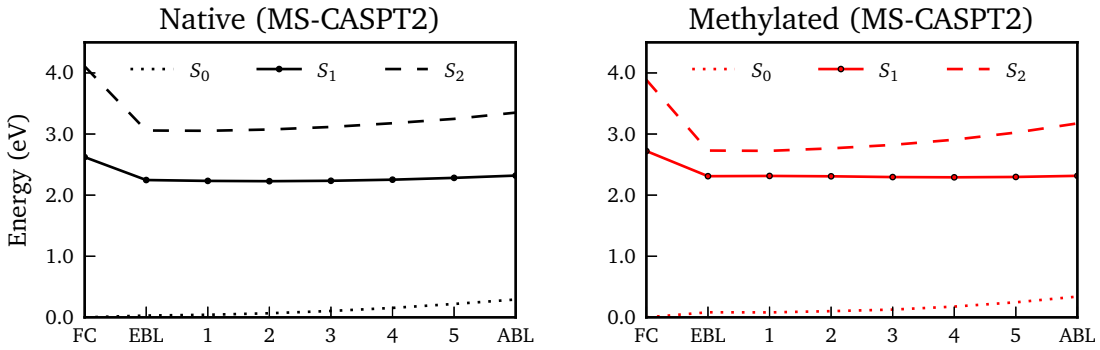


Figure 4.13: MS-CASPT2 energy surfaces for PSB and 10Me-PSB along the linear interpolation.

	PSB			10Me-PSB		
	Vert. em.	Most intense ESA	WF	Vert. em.	Most intense ESA	WF
EBL	599.7 (1.11)	488.7 (1.25)	$H \to L+1$ 0.35, $(H \to L)^2$ 0.15	645.3 (1.73)	471.0 (1.38)	$H \to L+1$ 0.36, $(H \to L)^2$ 0.14
		630.2 (0.51)	$H-1 \to L$ 0.26		679.0 (0.63)	$H-1 \to L$ 0.22, $H-3 \to L$ 0.13
ABL	773.7 (0.34)	479.4 (0.44)	$H-1 \to L+1$ 0.15, $H-1 \to L$ 0.11	842.3 (1.33)	365.6 (0.58)	$H \to L+1$ 0.22
		576.4 (0.27)	$H-1 \to L$ 0.29		542.5 (0.59)	$H-1 \to L$ 0.23
					(462.7 (0.11))	$(H-1 \to L)^2$ 0.13, $H-1 \to L+1$ 0.10)

Table 4.6: ESA computed at the EBL and ABL geometries for PSB and 10Me-PSB. SA20-CASSCF wavefunctions were used, and vertical emission are indicated at the CASPT2/SA20-CASSCF level. The two most intense ESA signals obtained are written, and the wavelengths are printed in nm. In brackets are printed the corresponding oscillator strengths.

In PSB, both geometries feature an intense ESA signal blue shifted with respect to the vertical emission. This ESA signal is found at 488.7 nm and 479.4 nm for EBL and ABL geometries respectively, and is less intense for the latter (oscillator strength are 1.25 and 0.44, resp.). This intense signal was also observed by Ruhman et al. [117], and was related to the ESA signal arising at 460 nm in the reactive I_{460} excited state intermediate in bacteriorhodopsin [112]. Another less intense ESA signal has been computed at higher wavelength. At the EBL geometry, this signal was computed at 630.2 nm (1.97 eV) ; a similar signal is computed at 576.4 nm (2.15 eV) at the ABL geometry. The shallow nature of the stimulated absorption peak has been attributed to an overlapping ESA signal around 760 nm (1.63 eV) by Kukura et al., following similar assumptions from Ruhman et al. [117] made out of an analogy with bacteriorhodopsin, where a discrepancy between the observed fluorescence spectra and stimulated emission has been proven [111–113]. In our model, this ESA could correspond to the signal arising from the EBL geometry, which is nicely placed in between the computed emissions for EBL and ABL (2.03 and 1.59 eV respectively). We should keep in mind here that this computation is carried out by

averaging 20 states ; the quantitative difference between the experimental ESA and the computed signal is around 0.33 eV, which is not an unreasonable value regarding the approximations made in our model, the principal one being that only one frame out of the dynamics has been considered.

In 10Me-PSB, we have more or less the same picture at the EBL geometry, where an intense blue-shifted signal has been computed at 471.0 nm (2.63 eV), and a less intense red-shifted signal has been found at 679.0 nm (1.82 eV). As in the case of PSB, this signal has a main $H-1 \rightarrow L$ component and it is found at wavelengths in between the two ABL/EBL emissions and provides a nice agreement with experimental data. However, the ESA at the ABL geometry are different, with two equally intense signals at 365.6 nm (3.39 eV) and 542.5 nm (2.28 eV). For reference, the blue excitation corresponding to the one found for the ABL point in PSB (with a major $H-1 \rightarrow L+1$ component) is also indicated in the table, although it has a significantly lower oscillator strength (0.11 *vs.* 0.44 in PSB). These signals are harder to interpret in 10Me-PSB, due to the lack of extensive experimental results, contrary to PSB. In this system, the ABL state is expected to rapidly decay to a conical intersection (CI) (see below, section 4.4.3), and the overall appearance of the pump-probe experiment published by Kukura et al. does not allow for more in depth analysis in this regard.

4.4.3 Path toward a conical intersection

Starting from the optimized ABL geometry for both systems, the dihedral angle around the $C_{11}-C_{12}$ bond has been rotated by increment of 10 degrees. Experiments show that *cis* $C_{11}-C_{12}$ products are obtained exclusively for 10Me-PSB, and that the quantum yield is 3-4 times higher for the 11-*cis* product than for the 9-*cis* product in PSB. As a consequence, this rotation seems to be the prefered one for both systems and was the only one considered in this work. For PSB, both the clockwise (CW) and the counter-clockwise (CCW) directions have been computed, due to the planarity of the molecule. In 10Me-PSB, the optimized ABL structure shows a pre-twist in the counter-clockwise direction. Therefore, we only computed the scan in this direction. All the geometries have been optimized at the CASSCF level, with the bond lengths of the conjugated chain frozen at the ABL geometry, and the dihedral angle around the $C_{11}-C_{12}$ bond frozen to its initial rotated value.

10Me-PSB The results are presented on figure 4.14. Both CASSCF and CASPT2 S_1 energy surfaces go downhill, and cross the S_0 surface at 111°(CASPT2) and 101°(CASSCF). This topology is reminiscent of a peaked conical intersection, which are typical of fast and effective photo-isomerization processes.

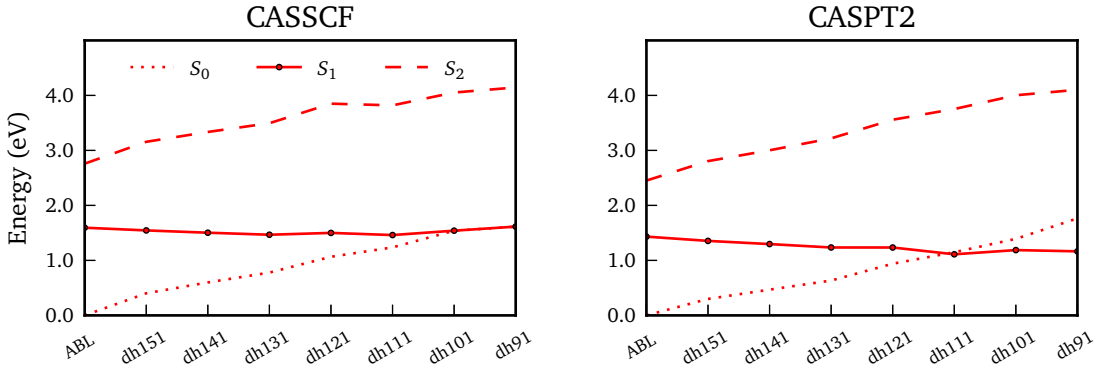


Figure 4.14: Constrained scan around the $C_{11}-C_{12}$ bond in 10Me-PSB. Only the clockwise direction has been computed.

Moreover, the connexion between the ABL geometry and the conical intersection seam is barrierless, which indicates that the transition from ABL to the seam is very fast. Thus, this finding is consistent with the low lifetime observed experimentally for 10Me-PSB.

PSB The results concerning PSB for both directions computed are presented on figure 4.15. In the two directions the qualitative picture is the same at both CASSCF and CASPT2 levels. At the CASSCF, no intersection between the S_1 and S_0 surfaces is observed, and the S_1 surface goes uphill in both directions computed. At the CASPT2 level, the two surfaces cross for a dihedral angle around $C_{11}-C_{12}$ around 97° (CCW) and −102° (CW), with a flat, slightly uphill, S_1 surface. This topolgy is reminiscent of a sloped conical intersection, indicating a less efficient population transfer, and thus to longer excited state lifetime. This is consistent with recent computational QM/MM dynamics [39], as well as with the experimental data.

Structure comparison In this part we highlight some of the features of the conical intersections that were located. For the geometries that were found to be degenerated at the CASPT2 level, we perform a CI optimization (see section 4.2.3), where all constrains on BLA or dihedral angles were removed. The resulting geometrical parameters are presented on figure 4.16.

As seen, the PSB CW and 10Me-PSB conical intersections share a lot of common parameters, which is not surprising given that the rotation is of the same nature. Concerning bond lengths, no difference is observed for the structures arising for the different rotations tested in PSB. In particular, we can notice that the $C_{11}-C_{12}$ bond keeps the same value as in the reference ABL structure, while the length of the

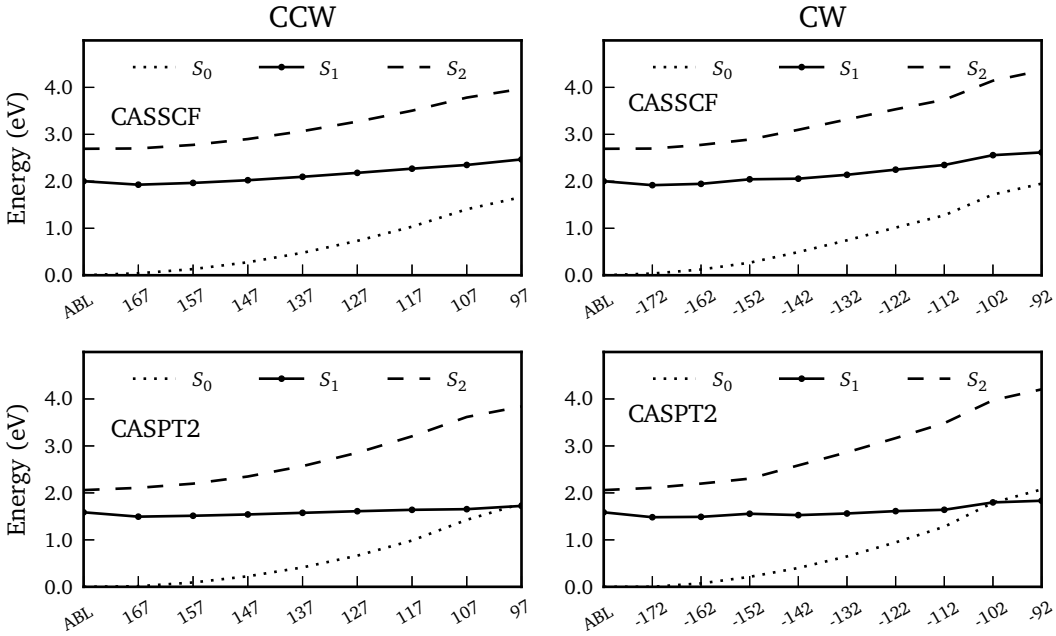


Figure 4.15: Constrained scan around the $C_{11}-C_{12}$ bond in PSB. Left panel: Counter-clockwise rotation ; right: Clockwise rotation.

C_9-C_{10} bond is considerably decreased. The isomerization around $C_{11}-C_{12}$ induces deformation in the surrounding, as both C_9-C_{10} and $C_{13}-C_{14}$ bonds are modified in the CI. The remaining of the retinal molecule is not affected by the rotation, as can be seen on the right side of figure 4.16. Thus, the rotation around $C_{11}-C_{12}$ is accompanied by two concomitant rotations, in the reverse direction, of C_9-C_{10} and $C_{13}-C_{14}$. This mechanism is reminiscent of the double-bicycle pedal already computed for the all-*trans* retinal of bacteriorhodopsin [125]. It is noticeable though that the $C_{13}-C_{14}$ bond is not affected is 10Me-PSB. Thus, the isomerization in 10Me-PSB seems to follow a different mechanism if one looks just at the CI geometries found within this static approach. The semi-classical dynamics show, however, a different point of view (see section 4.5). The isomerization of 10Me-PSB involves only rotations around $C_{11}-C_{12}$ and C_9-C_{10} in the optimized ("static") CI, and thus resembles the bicycle-pedal movement observed in rhodopsin [8]. In the two cases, the isomerization follows a space-saving mechanism, as it is the case in proteins, to limit the steric interactions with the surrounding environment.

Finally, the most stable conical intersection structure for PSB is the one arising from the counter-clockwise direction (stabilized by around 2.5 kcal/mol) with respect to the conical intersection found in the clockwise direction. This could indicate that the direction of the rotation is also affected by the

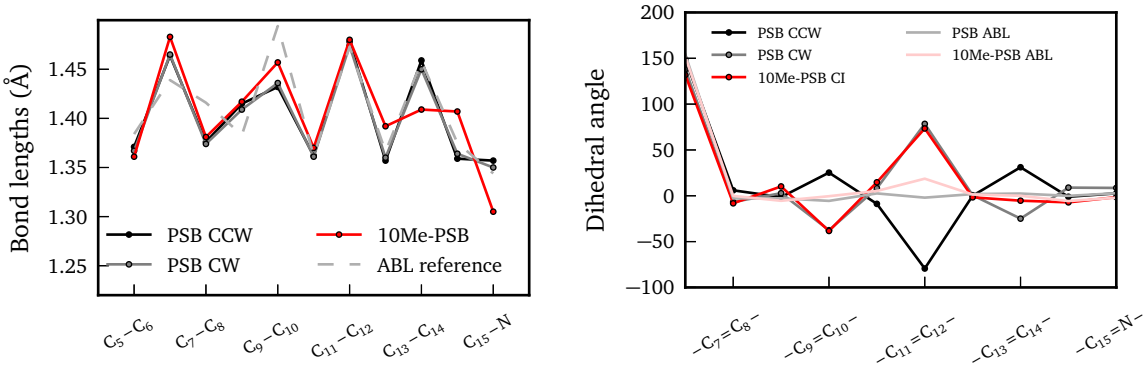


Figure 4.16: Geometrical parameters for the CI. Left: bond length alternation (BLA) ; right: dihedral angles. The bottom axis is labelled with names from 4.1. The reference BLA of ABL comes from 10Me-PSB optimization ; the references dihedral angles come from the corresponding contrained optimizations performed in section 4.4.1.

presence of the methyl group in 10Me-PSB, which tends to invert the direction of isomerization. However, this is observation could not really be conclusive since the results might be significantly affected by the choice of a selected snapshot determining the solvent molecules arrangements around the chromophore. The CI structures can be seen on figures 4.17 (PSB) and 4.18 (10Me-PSB). It can also be noticed that the torsion of the $C_{11}-C_{12}$ bond at the CI point is around 79.4°for PSB, while in 10Me-PSB it is found slightly lower, at 73.4°. This is consistent with the low quantum yield observed experimentally, indicating that the CI point is located on the reactant side of the energy surface, as proposed by Kukura et al. [121] Moreover, in 10Me-PSB this dihedral is farther from the fully distorted bond (90°), which may account for the lower quantum yield observed in this case.

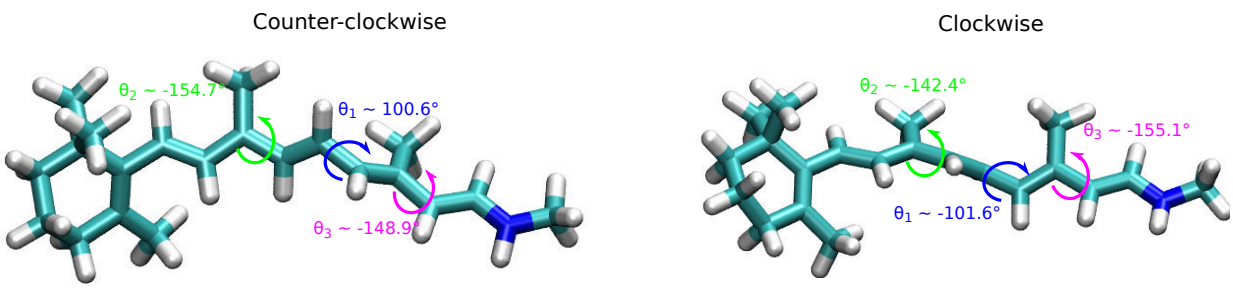


Figure 4.17: Optimized conical intersections of PSB.

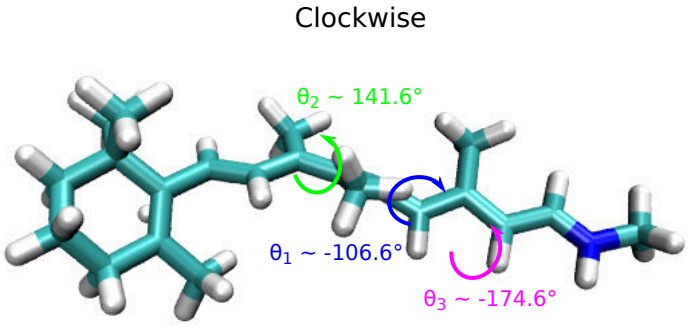


Figure 4.18: Optimized conical intersections of 10Me-PSB.

4.5 Semi-classical dynamics

The energies of the S_1 and S_0 states obtained out of a single semi-classical trajectory are shown on figure 4.19. The dynamics of 10Me-PSB clearly shows that a conical intersection is reached in about 0.6 ps, in remarkable agreement with the experimental value of around 0.7 ps, while for PSB, the energy difference is stable and stays around 50 kcal/mol along the 1.8 ps computed, here again in agreement with the experimental data. We have computed the evolution of the bond length alternation during the dynamics (see figure 4.20), defined as the difference between the average double bond lengths ($\approx$ 1.35 Å) and the average single bond length ($\approx$ 1.45 Å) — as defined in the ground state optimized structure. The BLA is expected to oscillate between negative (at FC geometry) and positive (for fully inverted structures) values, while a value close to zero indicates the full delocalization of the π-system, with all bond lengths having the same value.

In 10Me-PSB the BLA rapidly increases (figure 4.20), and then oscillates significantly until a stabilization obtained after 250 fs around 0.07 Å, indicating that an equilibrium is reached, corresponding to a structure with inverted bond lengths with respect to the FC geometry. The same rapid increase at the very beginning of the dynamics is also observed for PSB, but the equilibrated value is found around 0.03 Å, indicating a delocalized geometry. Thus, only the methylated system is able to efficiently reach an ABL geometry, which in turns allows for populating the torsion modes and lead to a twisted conical intersection. In contrast, the native system stays locked in a conformation that does not allow this population of the torsion mode. These findings are notably in agreement with the optimization carried out in previous sections 4.3.2, where an inverted ABL geometry is found for 10Me-PSB, while a delocalized EBL geometry is determined for PSB.

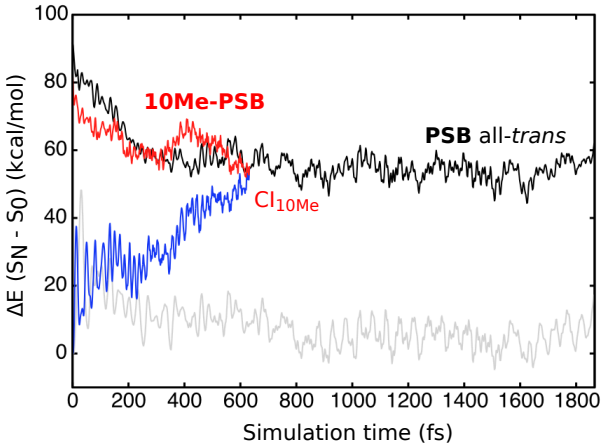


Figure 4.19: Evolution of the energies during a semi-classical trajectory for PSB (black and gray curves) and 10Me-PSB (blue and red curve).

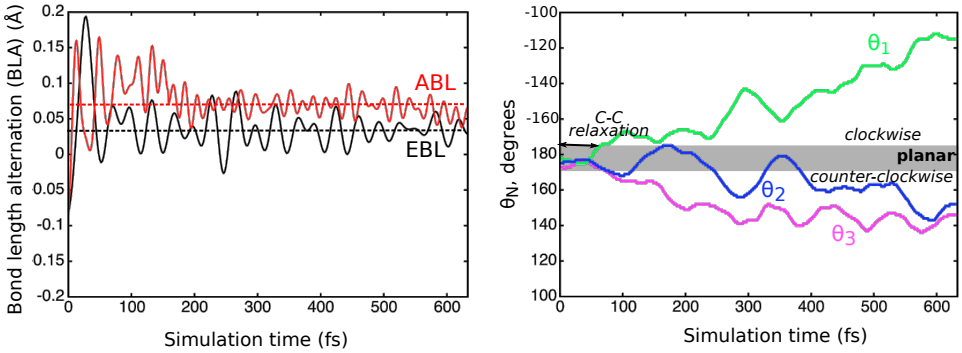


Figure 4.20: Left: Evolution of the bond length alternation (BLA). Right: Evolution of some torsional angles for 10Me-PSB.

Following the evolution of the torsional angles during the dynamics (figure 4.20), no significant evolution is observed for PSB, which is expected due to the locking of the system in a non-reactive geometry. In 10Me-PSB, a roughly planar structure is also observed in the first 80 fs of simulation, which correspond to the initial relaxation of the bond length. Then, a first and irreversible torsion is induced around C_9-C_{10}, which deviates by around $-40°$ while all other angles remain planar. The rotation around $C_{11}-C_{12}$ starts around 250 fs, after the BLA is equilibrated, with an oscillation around 30°, coupled with a reversed and reduced motion around $C_{13}-C_{14}$, between 250 and 350 fs. Finally, after 350 fs, a significant clockwise rotation around $C_{11}-C_{12}$ leads to a distorsion around 70°, while a concommitant counter-clockwise rotation is observed around $C_{13}-C_{14}$, leading to a distorsion of $-40°$. Thus, the asynchronous and opposite motions around C_9-C_{10} and $C_{11}-C_{12}$, accompanied by the movement around $C_{13}-C_{14}$,

is reminiscent of the double-bicycle pedal that has been postulated in bacteriorhodopsin [125]. This finding also suggests that the space-saving movement observed in protein [8, 125] also occur in solvent.

The results obtained out of the dynamics mostly confirm the findings obtained out of the minimum energy path. Indeed, a barrier due to an avoided crossing has been found for PSB, which traps the system in an EBL geometry, while the optimization of 10Me-PSB lead directly to the inverted reactive geometry. However, the conical intersection geometries for the photoisomerization of 10Me-PSB found in the two approaches are different. In the minimum energy path approach (see section 4.4.3), the optimized conical intersection displays a planar $C_{13}-C_{14}$ bond, while a distorted bond is obtained in the dynamics. Overall, these results indicate that the S_1/S_0 intersection space of 10Me-PSB involves both types of CIs, *i.e.* with planar or distorted $C_{13}-C_{14}$ dihedrals, and that either a single- or a double-bicycle rotation will lead to an effective S_1 decay to the ground state.

4.6 Conclusion and perspectives

In this chapter we have presented a modelling of the excited state of two retinal molecules in solution: the native PSB and its 10-methylated derivative, 10Me-PSB. After molecular dynamics excited state sampling, the ground and excited states of both systems have been characterized. The presence of a methyl group strongly affects the excited state manifold, with the removal of the wavefunction mixing between the S_1 and S_2 states observed in the native molecule. Excited state geometries were also found to be affected by the methylation. In the native PSB, a geometry where the bond lengths are equalized (EBL) is obtained when the $H \rightarrow L$ dominated root is optimized, while an alternated bond length (ABL) structure is found for 10Me-PSB. As the experiments show two emission signals, we assumed that the two minima can coexist in both molecules. To assess this assumption, an interpolation is carried out for both systems between the EBL and ABL geometries. Here again, methylating the retinal molecule has huge consequences. In native PSB, the S_1 and S_2 states are close in energy, and they share similar configurations in the wavefunction, resulting in an avoided crossing and a barrier of around 3 kcal/mol to go from EBL to ABL. This mixing is lifted in 10Me-PSB, and the transition is barrierless, resulting in an ABL than can be easily reached. This may arise from a destabilization of the EBL geometry in 10Me-PSB due to the particular repartition of the charges at this geometry, which is already the one that is expected for a fully inverted structure.

The photochemistry out of the ABL geometry has also been investigated. With constrained scans around the $C_{11}-C_{12}$ bond, we found that a conical intersection can easily be reached in 10Me-PSB, leading to a fast decay of the excited state once the ABL geometry is reached. For PSB, two directions were probed, and a slightly peaked CI has been found in both cases. This accounts for the higher excited state lifetimes observed experimentally in PSB than in 10Me-PSB. Moreover, the geometries found at the CI have shown that for both system, the dihedral around the $C_{11}-C_{12}$ bond is below the 90°barrier, thus showing that the CIs occur on the reactant side of the potential energy surface.

The picture that arises from these computations is the following:

- After excitation, both systems relax to an EBL-like minimum. This relates to the first emission signal observed in the expermiments.
- While the relaxation to ABL is very fast in 10Me-PSB, due to a barrierless transition, PSB has to overcome a low barrier (around 3 kcal/mol) to reach the active ABL geometry. Therefore, the emission signal arising from EBL has a longer lifetime in PSB.
- As soon as ABL is reached, a fast transition toward a CI point between S_1 and S_0 is expected in 10Me-PSB, leading to a fast decay of the excited state. In PSB, this transition is less favourable, due to the topology of its CI point. Here again, a longer lifetime is expected at the ABL geometry in PSB.

These findings have mostly been confirmed by semi-classical dynamics. In particular, the bond lentgth relaxation in 10Me-PSB is ultra-fast, and allows for an efficient population of the torsion mode, while in PSB the system is locked in an EBL geometry, and is thus unable to easily isomerize. Both molecules are expected to follow a space-saving mechanism for the isomerization, with either an asynchronous bicycle pedal movement, where the rotation around $C_{11}-C_{12}$ is accompanied by the reverse movement around C_9-C_{10}, or an asynchronous double-bicycle pedal movement, where a rotation around $C_{13}-C_{14}$ in the same direction as the one around C_9-C_{10} occurs. The conical intersections computed out of the MEP for PSB, as well as the trajectory obtained for 10Me-PSB, advocate for a flat region in the potential energy surface where the conical intersection seam is developed, and thus more than one mechanism can occur. A qualitative summary of what has been performed in this chapter is available on figure 4.21.

The global low quantum yield can be rationalized with the topology of the MEP computed on the isomerization mode, where the conical interesection are found for geometries that are quite far from the fully distorted one. Obtaining data for PSB is cumbersome due to the timescale of the photo-reaction, and

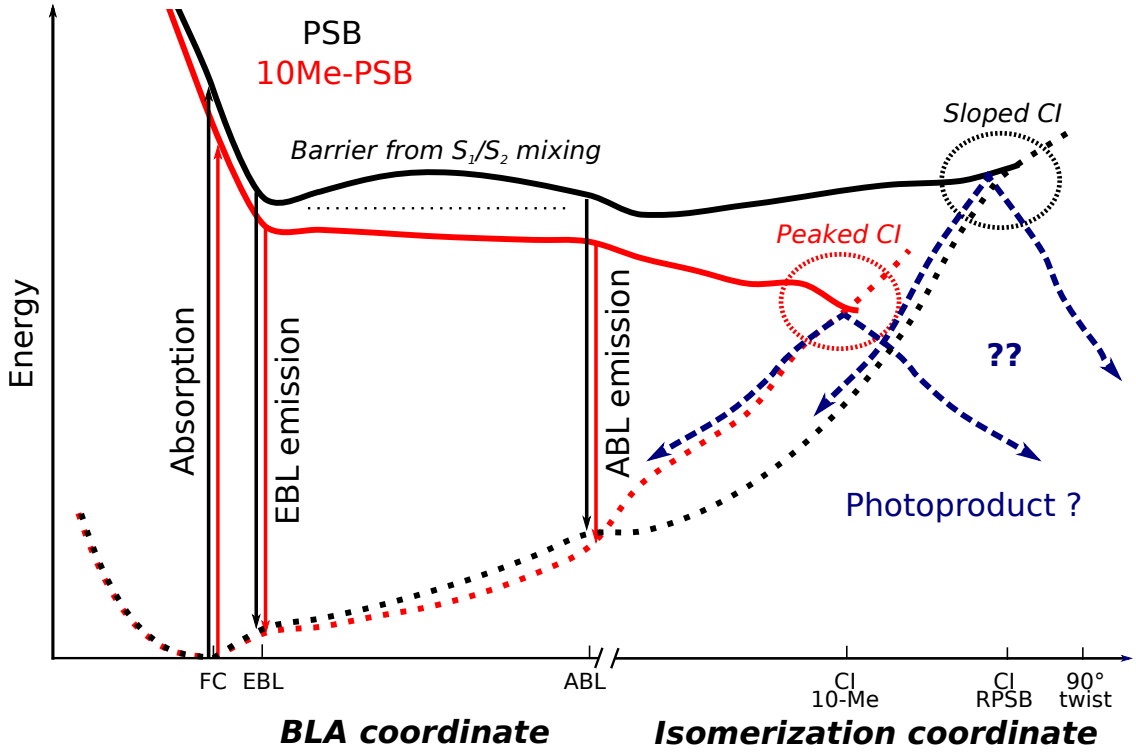


Figure 4.21: Qualitative summary of the conclusions of the study.

due to the excited state manifold obtained at the CASSCF level. For this case, scaled-CASSCF dynamics [8, 37], where CASSCF energies are corrected based on CASPT2 calibration, could be the method of choice. For the methylated molecule, several trajectories have been computed, starting from different starting points on the conical intersection seam, on the ground state. This also correlates well with the very low quantum yield obtained experimentally for 10Me-PSB. The difference with respect to the protein environment could then be an evolutionary fine tuning of the retinal cavity to bias the reaction in favour of the photo-product.

5 Understanding the tuning of the photophysics of rhodopsin mimics by a combined theoretical and experimental approach

Manuscript in preparation

Résumé Nous présentons dans ce chapitre notre modèle computationnel pour trois mimiques de rhodopsine obtenues récemment par bio-ingénierie, nommées M4, M8 et M10, qui absorbent à respectivement 507, 590 et 622 nm. Notre modéle est fondé sur un échantillonnage de l'hydratation des protéines, suivi de calculs QM/MM avec la théorie des perturbations multi-référentielles. Nous avons montré que l'hydratation des protéines est un paramètre crucial pour reproduire précisément l'absorption de chaque système. En combinaison avec des expériences de spectroscopie UV-Visible pompe-sonde ultra-rapide, nos calculs de chemin d'énergie minimale montrent que les états excités des trois systèmes évoluent très différemment. Nous avons interprété les différents signaux spectroscopiques par la présence de deux géométries peuplées séquentiellement pendant l'évolution temporelle de l'état excité, et nous avons montré que la forme de la protéine joue un rôle primordial dans le mécanisme de population. Ces deux géométries sont caractérisées par leur différentes alternations de longueur de liaisons, et nous avons obtenu des structures avec des liaisons égalisées (EBL) et d'autres avec des liaisons alternées (ABL). Chaque spectre expérimental a été reproduit quantitativement par nos modèles QM/MM, et nous avons rationalisé les différents temps de vie d'état excité obtenus par les différences dans la forme des surfaces d'énergie potentielle reliant la géométrie réactive ABL à la ligne d'intersection conique. Notre étude a montré que la poche qui encapsule le rétinal peut moduler non seulement l'absorption des protéines,

mais aussi les propriétés de leur différents états excités, et prépare des études plus approfondies pour la conception dirigée de nouveaux appareils photo-sensibles.

Abstract In this chapter we present our computational model for three recently bio-engineered rhodopsin mimics, namely M4, M8 and M10, with vertical absorptions of 507, 590 and 622 nm repsectively, based on extensive sampling of the hydration of the proteins followed by QM/MM computations with multi-reference perturbation theory. We show that the hydration of the proteins is a crucial parameter for the accurate reproduction of the vertical absorption of each system. In combination with ultra-fast pump-probe UV-Vis spectroscopy, minimum energy paths computations show that the three systems have very different excited-state evolution. We interpret the different experimental spectroscopic signals in terms of two geometries populated sequentially during the time evolution, and we show that the protein scaffold has a significant effect on these populations. These geometries are characterized by their different bond length alternation pattern, and either even bond lengths (EBL) and alternated bond lengths (ABL) were obtained. Each experimental spectrum is quantitatively reproduced by the QM/MM modeling, and the different excited-state lifetimes of the proteins are rationalized in term of conical intersection (CI) seam accessibility out of the reactive ABL geometry. This study shows that the retinal binding pocket is not only able to tune the vertical absorptions, but also the excited-state properties of the different proteins, and our model paves the way for the directed design of new tunable photo-sensitive devices.

5.1 Introduction

5.1.1 Bio-engineering of the hCRBPII protein

The natural variability of rhodopsin-like proteins, and the variety of their properties leading to different activities and functions, make them an interesting starting point for the design of tunable photo-activated bio-inspired devices, which could be used in opto-electronics, or opto-genetics [23, 126].

Artificial tuning of natural rhodopsins has already been successfully achieved, leading to fundamentally different biological functions. Single point mutations have been used to convert the inward proton pumping bacteriorhodopsin into a outward chloride pump [127], to allow interconversion between chloride, sodium and proton pumps in bacterio- and halorhodopsins [128], to create inward proton pumps [129], and to transform ion pumps into ion gated channels [130].

Directed design of new retinal-based devices require an accurate understanding of the interactions between the binding pocket and the retinal molecule. Indeed, the retinal chromophore is highly sensitive to modifications of the protein backbone, either *via* electrostatic interactions or steric constraints [19]. The presence of a counterion affects the relative energies of the ground and excited states, leading to significant changes in the vertical absorption spectra [27, 48, 50]. Moreover this is further tuned by the localization of the electrostatic potential along the retinal polyene chain, resulting in maximum absorptions ranging from 420 nm (for human short wave sensitive pigment, hSWS) to 587 nm (in sensory rhodopsin I). On the other hand, combined experimental and theoretical studies on gas-phase PSBs have demonstrated that an even distributed electrostatic field lead to significant red-shift of the absorption maximum [20].

Obtaining systems were the electrostatic potential around the retinal can be accurately tuned is of crucial importance for improving our current models and understanding of these interactions. The research group of J. Geiger recently proposed systems based on CRABPII (cellular retinoic acid binding protein type II) and hCRBPII (human cellular retinol binding protein type II), in which the local environment of the aldehyde site of an embedded all- *trans* retinol or retinoic acid molecule has been mutated in order to allow for the formation of a protonated Schiff base, thus mimicking what is observed in opsin proteins [131]. Mutations were carefully targeted to provide relative conformations between the embedded retinal molecule and the nearby lysine suitable for imine formation following a Bürgi-Dunitz trajectory, as presented on figure 5.1.

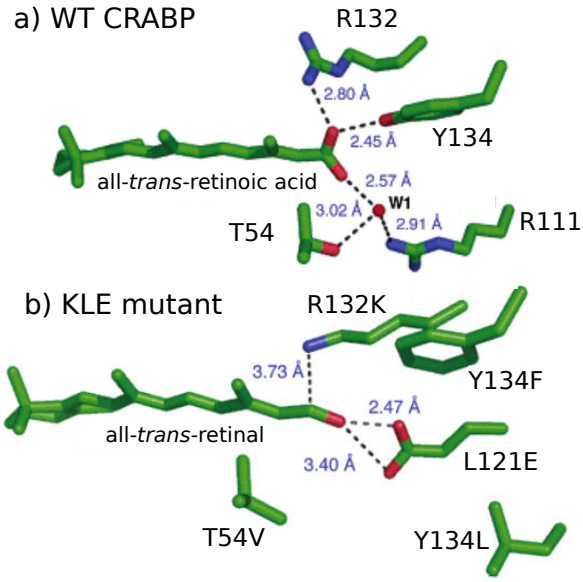


Figure 5.1: Comparison between wild type CRABPII protein (a) and mutated KLE-CRABPII (b). In particular, Arg132 has been mutated to lysine, and a water molecule has been removed with further mutations, to allow for a conformation of lysine and retinal favorable for Schiff base formation.

Successful results lead for further analysis, in which the binding pocket of the retinal has been further mutated, proving that effective tuning control required a complete embedding of the retinal inside the binding pocket [132]. Therefore, hCRBPII appears as one of the most promising systems, as the binding pocket is able to fully enclose a complete retinal molecule, contrary to what happens in CRABPII. Wang et al. recently proved that hCRBPII was indeed capable of covalently binding a retinal molecule through the Q108K and K40L mutations, and that a small number (6-7) of point mutations inside the binding pocket of hCRBPII can lead to systems displaying absorption maxima ranging from 425 to 644 nm [133] (see figure 5.2). It is worth to note that these rhodopsin mimics do not contain a negatively charged counterion associated with the retinal PSB.

This ground-breaking engineering paves the way for innovative retinal-base optical probes. Successful modeling of the opsin shift in some of these mutants has been achieved by Hayashi et al., by using the crystal structure of the minimally mutated structure (M4 from figure 5.2) as a basis, and manually including different mutations [134]. Then, extensive QM/MM free energy simulations were used to refine the models. This method has allowed to quantify the long-range effects of the point mutations,

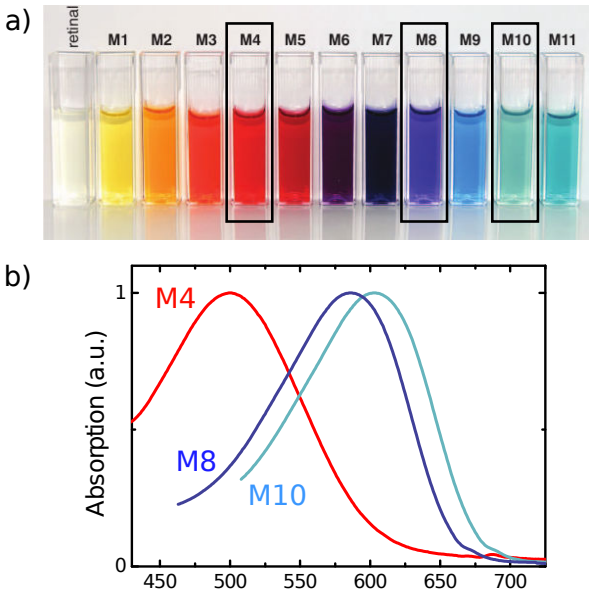


Figure 5.2: Absorption range obtained for hCRBPII mutants, with highlights on the mutants considered in the following (a), and corresponding absorption spectra measured in the group of Pr. Cerullo (b).

and to assess the determinants of the color tuning observed, and showed that the remote effect of some mutations could tune the hydration of the retinal, in turn tuning the absorption energies.

5.1.2 Systems considered and pump-probe spectroscopy

To the best of our knowledge, no data about photo-induced processes in these newly designed systems has been obtained so far. To investigate on the possible photochemistry and photophysics of these systems, a combined theory and experiments approach has been set up in collaboration with the group of Pr. Giulio Cerullo (Politecnico di Milano). Similar approaches based on state-of-the-art QM/MM methods, using multi-reference computations combined with perturbation theory, and pump-probe spectroscopy were successful in characterizing the primary events of the photochemistry happening in visual rhodopsins [8, 135].

The mimics labeled M4, M8 and M10 have been selected for this investigation, primarily due to the availability of crystal structures, but also because the comparison between them should lead to interesting conclusions. A representation of these proteins is displayed on figure 5.3.

M4 contains the minimal number of mutations to allow for the formation of the PSB. The retinal is surrounded mostly by polar residues, with a water molecule lying close to the iminium group, and

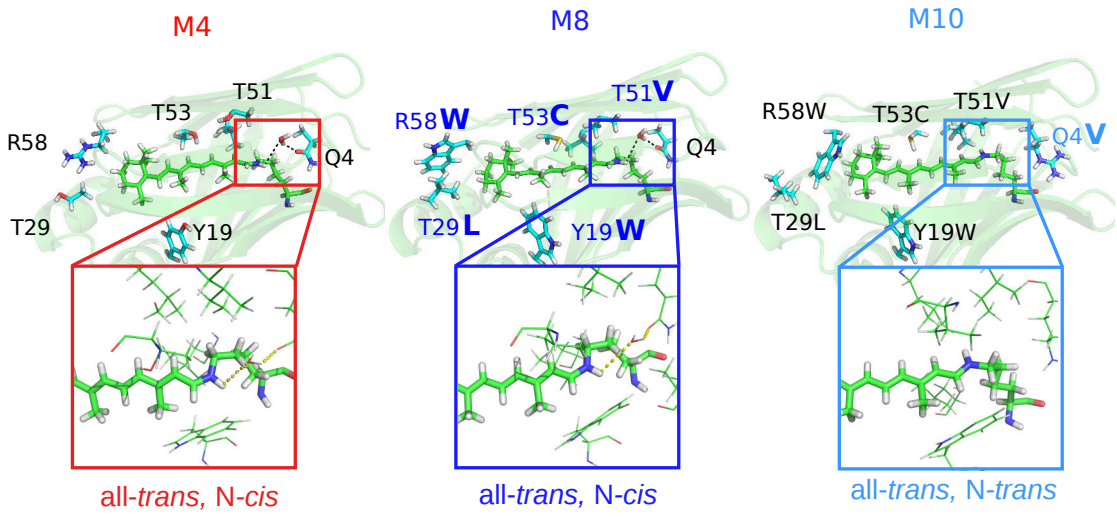


Figure 5.3: Systems considered in this study. Left: M4 (Q108K:K40L); Center: M8 (Q108K:K40L:T51V:T53C:Y19W:R58W:T29L); Right: M10 (Q108K:K40L:T51V:T53C:Y19W:R58W:T29L:Q4R).

bridging it with Gln4 [133]. Mutations leading to M8 aim at increasing the hydrophobicity of the binding pockets, by inserting apolar residues (tryptophans and leucines mostly). The last mutation, leading from M8 to M10, involves Gln4, which is mutated into an arginine, resulting in the loss of the ordering water molecule, and to the change of conformation of the iminium from *cis* to *trans*. This change of conformation happens in all mimics that feature a mutation of Gln4.

Pump-probe spectroscopy is one of the most useful experimental techniques to investigate on excited state evolution of organic molecules. A pump laser pulse, peaked around the targeted excitation, is used to populate the excited state. The spectrum obtained will serve as a reference. The wavepacket then evolves on the excited state potential energy surface (PES). This evolution is probed with ultra-fast laser pulses, with broad dispersion, for instance in the full visible range (400–700 nm). Spectra are recorded with varying probe delays, and the differences with respect to the reference spectrum are computed. It is useful to then represent the obtained differential map by plotting the wavelength signal with respect to the probe delay, as presented on figure 5.4. Positive signals are either due to ground state bleaching (GSB), arising from the depletion of the ground state, or to stimulated emission (SE), arising

from the interaction of the excited state with the incident light. Negative signals are due to excited state absorptions (ESA).

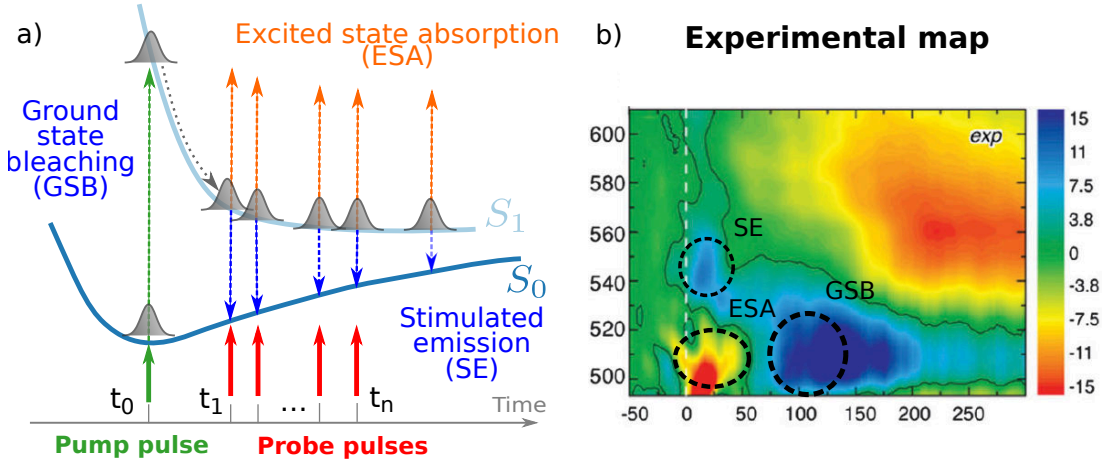


Figure 5.4: (a) Schematic pump-probe spectroscopy principle; (b) Example of differential map obtained in Rhodopsin [8, 135], with negative signals (ESA) in red and positive signals (SE and GSB) in blue

The pump-probe spectra have been recorded with a pump centered on the vertical absorption of each protein (508, 590 and 622 nm corresponding to 2.44, 2.10 and 1.99 eV for M4, M8 and M10 respectively). The resulting spectra are presented on figure 5.5.

Although the assignment of GSB signals is fairly easy (if the pump is set at the excitation wavelength of the sample, then the signal overlaps with the pump), SE and ESA are harder to interpret, because they arise from points on the little known excited state surface. Modeling is of paramount importance for understanding and interpreting these signals.

The spectrum of M4 is highly different from the others. Two possible SE are visible, peaking around 1.84 and 2.10 eV, and one ESA peaks around 2.60 eV. The SE signals have different properties, as the blue-shifted one is very short, while the second lasts longer. The photochemistry is also pretty fast, and the excited state lifetime can be estimated at around 3 ps. M8 and M10 are very different from this picture. Indeed, the GSB signals are broader, and the photochemistry of these two systems is far longer

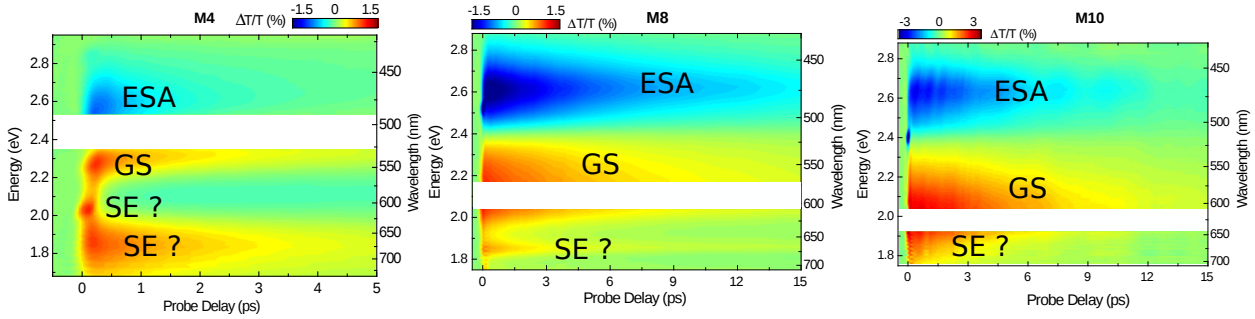


Figure 5.5: Experimental pump-probe spectroscopy results obtained for M4, M8 and M10. Stimulated emissions (SE) and ground state bleaching (GSB) appear in red (positive signals), while excited state absorptions appear in blue (negative signals).

than what is observed in M4 (around 12 and 10 ps for M8 and M10 respectively). In M8, a long lasting, low intensity SE signal is observed around 1.84 eV. The broad ESA signal peaks around 2.60 eV. M10 features an even broader GSB, but the positive signal around the pump wavelength is asymmetric, and a small SE bump can be guessed in the red-shifted region around 1.84 eV, at approximately the same energy as the SE of M8. The ESA is also very similar to what is observed in M8, and peaks around 2.61 eV.

Interpreting these different signals, as well as their relative intensities, and the different lifetime, requires a robust computational model. Obtaining such a model is not straightforward given the complexity of the environments, and the lack of extensive experimental characterization, contrary to the case of natural rhodopsins or bacteriorhodopsins. The next section tackles the set-up of such a model.

5.2 Building a reliable computational model

5.2.1 QM/MM setup

We started from the crystal structures available in the Protein Data Bank under the labels 4EXZ, 4EFG and 4EEJ for M4, M8 and M10 respectively. These proteins were crystallized in dimers. For this study, only one monomer was considered for each protein. The monomer displaying the most complete retinal chain has been kept. Hydrogen atoms have been added by using the standard Amber tool-chain [106]. No titrable residues are found in the close vicinity of the retinal. Therefore, all the amino-acids were considered in their standard protonation state. TIP3P model has been used for the water molecules, while the protein was treated with the ff99SB forcefield [108]. GAFF parameters [102] for the retinal were obtained with Antechamber [107], following the standard AMBER guidelines for parametrization

of new residues. Hydrogens in the protein were then relaxed at the MM level, while heavy atoms were constrained to their crystallographic positions, assuming that the crystal structure is a reasonable average representation of the real system.

All the QM/MM computations were handled by the COBRAM interface [59], with both High-Low and High-Medium-Low schemes, as shown in figure 5.6. The QM/MM frontier has been set at the C_ϵ of the lysine residue (Lys108) bonded to the retinal residue. The atom-link method has been used to complete the dangling bond. All retinal atoms were considered in the High Layer. No movable medium layer has been included for ground state computations. To reduce the computational effort, excited state optimizations were performed on a reduced system, similar to the one used in the previous chapter. An HML scheme has been considered, in which only the atoms in the polyenic chain were treated in the high (QM) layer, while the remaining atoms in the ionone ring, as well as all the atoms of the side chain of Lys108, were included in the movable medium (MM) layer. Figure 5.6 depicts the different schemes used in this study. This setup was used to account for possible changes in conformation in the lysine side chain upon excitation.

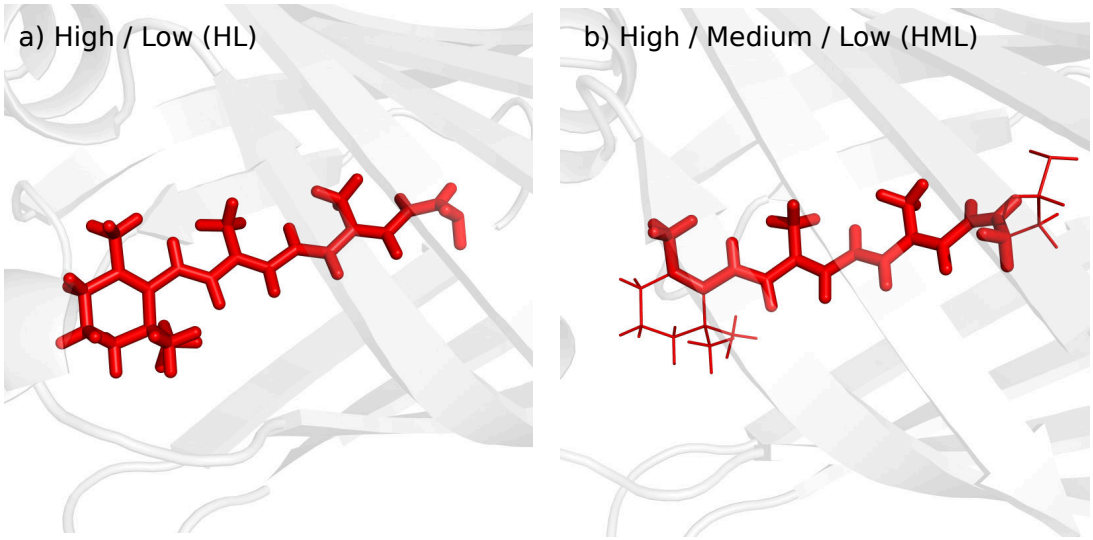


Figure 5.6: QM/MM scheme used in this chapter. (a) High / Low (HL) scheme used for GS optimizations; (b) High / Medium / Low (HML) scheme used for ES optimizations.

The complete active space self-consistent field (CASSCF) [40], with correction from second order perturbation theory (CASPT2) [41], were used throughout this study. The full π-system of the retinal has been included in the active space, resulting in 12 electrons in 12 orbitals (12,12). SA3-CASSCF(12,12)/6-31G* level has been used for all ground state geometry optimizations, while SA4-CASSCF(12,12)/6-31G*

level was used for excited state geometry optimizations to account for possible root flipping. All the energies have been computed at the SS- or MS-CASPT2 levels with ANO-S basis sets. Mulliken charges have been computed on top on CASSCF wavefunctions, and oscillator strengths were computed using CASPT2 energies. Molcas 8 [109] has been used as the QM software.

Minimum energy paths for bond lengths relaxation and photoisomerization were obtained with unrelaxed and relaxed scans, respectively. Conical intersections (CI) were optimized using the projected gradient method from Bearpark et al. [67] as implemented in COBRAM. Excited state absorptions (ESA) has been probed by increasing the number of states included in the SA-CASSCF procedure. 10 states were considered for ESA computations.

5.2.2 First attempt based on the crystal structures

After MM relaxation of the crystal structures, each system has been optimized following the method described in previous section. Energies have then been computed and are displayed in table 5.1. Although all the transitions computed at both SS- and MS-CASPT2 levels are strongly blue-shifted, the general trend is reproduced, with M4 absorbing at higher energies, followed by M8, and finally M10. The wavefunctions obtained show a high level of mixing between S_1 and S_2 in M4 and M8. Both states are dominated by a combination of the single $H \rightarrow L$ and the double $(H \rightarrow L)^2$ excitations. As a consequence, the oscillator strengths for both $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ transitions are very close. On the other side, the vertical absorption computed for M10 is surprisingly close to the experimental one, and the wavefunctions is cleaner than in the other cases, with S_1 being dominated by the $H \rightarrow L$ transition, and S_2 being dominated by the $(H \rightarrow L)^2$ transition. To assess the influence of the geometry optimization level, the retinal geometries were also obtained at the MP2 and DFT (B3LYP) levels of theory. CASSCF is indeed known to overestimate the bond length alternation (BLA) in retinal molecules, which leads to blue-shifted vertical absorption. However, even with MP2 and B3LYP geometries, strong blue-shifts were also obtained, ruling out that the optimization method is responsible for the discrepancy between computations and experimental data. Adding counterions also did not lead to significant improvements.

We concluded that something is missing in the model we used so far. The QM/MM setup used has been successfully applied in rhodopsin. In this protein, the retinal is embedded in a trans-membrane protein, which shields the retinal molecule from the external environment, as shown in figure 5.7. However, the mimics are smaller proteins, and the retinal may not be as well shielded in those as it is in rhodopsin. Moreover, the mimics are not trans-membrane proteins, and are therefore likely to be effectively solvated

System	CASPT2	MS-CASPT2	Exp.	ΔE CASPT2/Exp	ΔE MS-CASPT2/Exp
M4	3.56	3.30	2.53	1.03	0.77
M8	3.03	2.91	2.1	0.93	0.81
M10	2.36	2.50	1.99	0.37	0.51

Table 5.1: SS- and MS-CASPT2 energies (in eV) computed on top of CASSCF optimized geometries in M4, M8 and M10, and difference with respect to experimental data.

in vivo. Thus, we concluded that explicit solvation of these mimics could be a crucial step for accurately reproducing the experimental absorptions. Similar conclusions, and explicit solvation model, has been successfully used by Huntress et al. [136] in a study on similar mimics based on CRABPII. We then decided to follow a similar setup, presented in the next section.

5.2.3 Classical dynamics sampling

The principal goal of our study is to investigate on the primary events that occur upon photon absorption in the mimics. In particular, we have shown that simulating the hydration of these proteins is of paramount importance. Therefore, we are mainly interested in achieving a realistic sampling of an explicit solvent around the proteins, considering the crystal structure as reasonable averages of the protein conformations on the relevant timescale. As a first approximation, we want to stay as close as possible to the crystal structures, and a complete molecular dynamics (MD) study is out of the scope of this report. Cheng et al. [134] have recently presented a more in depth study of the long-range effects of the mutations involved, with QM/MM re-weighting free energy self-consistent field methods.

The relaxed crystal structures have been solvated in explicit octahedral boxes of TIP3P water molecules. The M4, M8 and M10 systems were neutralized with 2, 4 and 3 Na^+ ions respectively. Overall, each system contains around 16,000 atoms. A representative setup (for M4) is presented on figure 5.8.

After relaxation of the solvent molecules and hydrogen atoms, the systems were heated from 0 to 300 K at constant volume for 200 ps, followed by 500 ps of pre-equilibration steps at constant temperature (300 K) and constant pressure (1 atm), using the weak-coupling algorithm for temperature control, and Berendsen barostat. All the dynamics were performed with a 2 fs timestep, using the SHAKE algorithm to constrain all bonds containing an hydrogen atom. These pre-equilibration were carried out while constraining all protein atoms to their crystallographic positions with harmonic potentials (30 kcal/mol). The equilibration were kept running for 1.5 ns, followed by a production run of 100 ns trajectories finally

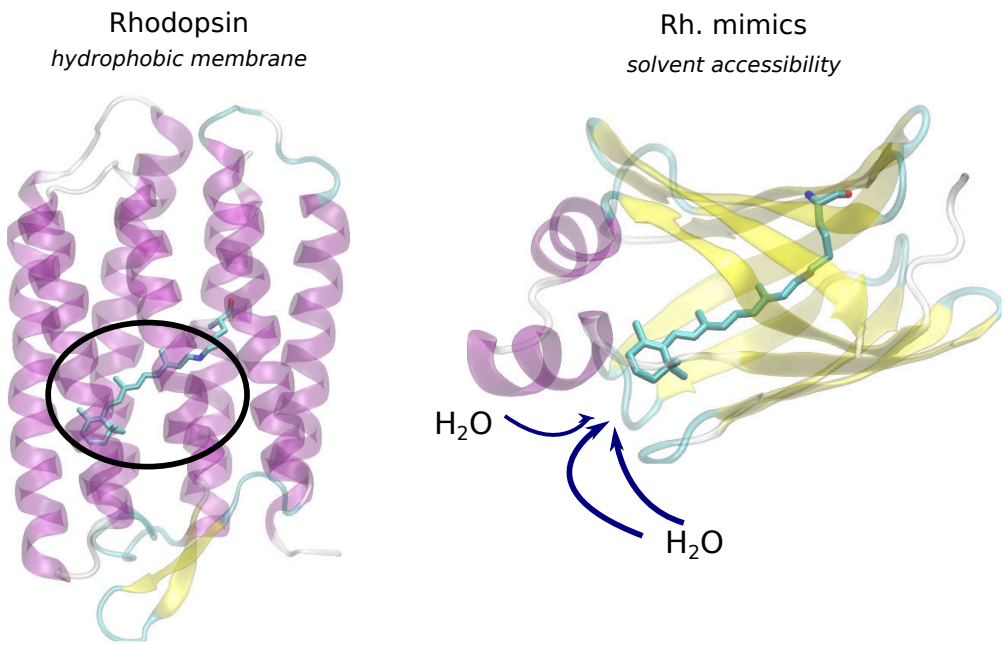


Figure 5.7: Comparison of the secondary structures of Rhodopsin (left) and Rhodopsin mimics (right). In Rh, the retinal is embedded in the hydrophobic membrane, and thus is shielded from the environment. On the contrary, rhodopsin mimics are smaller proteins that are surrounded by solvent.

used for analysis. The GPU implementation of PMEMD software [137], from AMBER 12, was used to compute each long trajectory. As a complete dynamical analysis was not required, frames were recorded every 200 ps. The harmonic restraints imposed on protein atoms have been kept for all molecular dynamics computations.

In particular, we are interested in the hydration of the retinal molecule inside the protein. This analysis has been performed with the MDAnalysis package [138, 139]. For each frame recorded, we computed the number of water molecules that are within 5 Å from any heavy atoms of the retinal.

We have found that the number of water molecules surrounding the retinal can vary widely along the trajectory, despite the protein atoms being frozen, especially in the cases of M4 and M8. The trend in hydration also follows our expectation. M4 is indeed the most hydrated, while M10 has the most hydrophobic environment. Inspection of the crystallographic water molecules inside the proteins show that these are very stable, and little to no exchange with the bulk is observed.

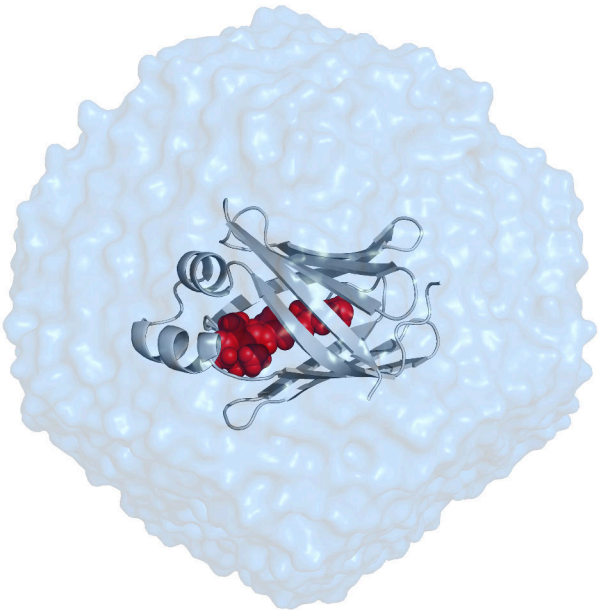


Figure 5.8: Representation of the octahedral box used for classical dynamics simulations. The protein (M4) is represented in white cartoon, and the retinal molecule in red spheres.

As was done in the study of Huntress et al., ten snapshots, equally spaced by 10 ns, were extracted out of each trajectory. In the following, we apply our QM/MM setup, described above, to all 30 samples (10 samples extracted for 3 different trajectories).

5.3 Ground state properties

In this section we will highlight the main properties of the ground state structures found after QM/MM optimization of all the selected snapshots for the three proteins. First, we will focus on the structural parameters of the optimized geometries, and on the vertical absorptions obtained. We will then highlight the electrostatic effect of the environment, by comparing the results obtained for the embedded retinal molecules to those obtained when the retinal is extracted and considered *in vacuo*. Eventually, the Mulliken charges, as well as the dipole moments computed for each system will be compared.

5.3.1 Geometries and vertical absorptions

Geometrical parameters

After QM/MM optimization of the ten snapshots taken out of the classical trajectory for each protein, we collected the values of the bond lengths of the retinal conjugated chain, as well as the corresponding values of the dihedral angles. In order to get more significant values, the deviation from planarity for a dihedral angle d, defined as

$$dev = (180 - |d|)\,\frac{|d|}{d}, \tag{5.1}$$

has been computed. Visual inspection of the optimized structures show that all of them have similar BLA, although M10 has more variations in the $C_{12}-C_{15}$ region, with values for the $C_{12}-C_{13}$ and $C_{14}-C_{15}$ bonds ranging from 1.433 Å to 1.448 Å and from 1.406 Å to 1.423 Å respectively. Dihedral angles are also fairly similar from one snapshot to another, although some variations are observed in certain region. In M4, the $C_{11}=C_{12}-C_{13}=C_{14}$ has deviations ranging from −1.41° to 10.58°, while all other angles have deviations varying by around 6°. In M8 and M10 the deviations observed only display variations by 8° at most. For easier comparison between the different proteins, these values were averaged over the ten snapshots, and the results are presented on figure 5.9, and on the corresponding table 5.2. Concerning the BLA, the three systems have a similar behaviour, and all of them display the alternation already observed for solvated retinals (see chapter 5). However, M4 has more pronounced alternation in the $C_{12}-C_{15}$ region, with elongated $C_{12}-C_{13}$ and $C_{14}-C_{15}$ bonds with respect to M8 and M10.

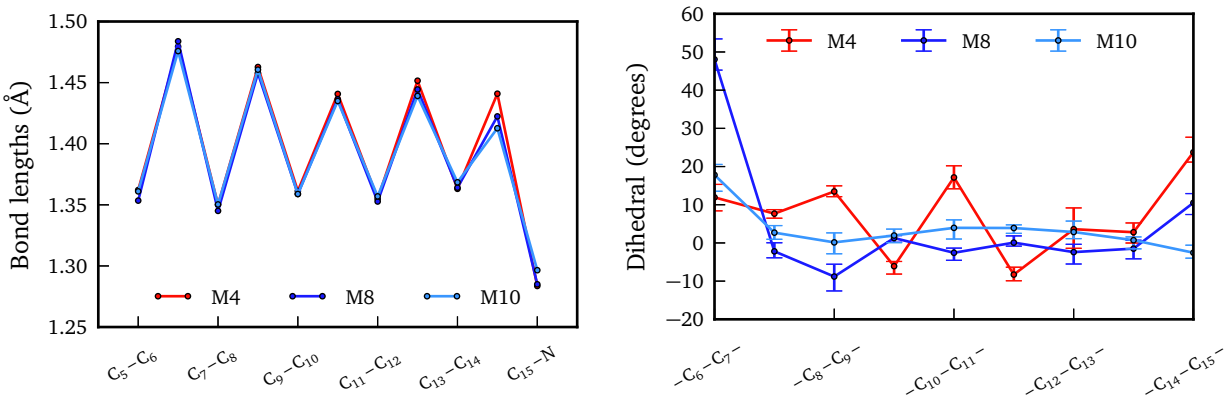


Figure 5.9: Average geometrical parameters for the three proteins. Left: bond length alternation (BLA); right: dihedral deviations from planarity, with error bars corresponding to lowest deviation and highest deviation.

	M4	M8	M10
C_5-C_6	1.362	1.353	1.361
C_6-C_7	1.479	1.484	1.476
C_7-C_8	1.350	1.345	1.350
C_8-C_9	1.463	1.458	1.461
C_9-C_{10}	1.361	1.359	1.359
$C_{10}-C_{11}$	1.441	1.436	1.435
$C_{11}-C_{12}$	1.355	1.353	1.357
$C_{12}-C_{13}$	1.452	1.444	1.439
$C_{13}-C_{14}$	1.363	1.364	1.368
$C_{14}-C_{15}$	1.441	1.422	1.413
$C_{15}-N$	1.284	1.285	1.297

	M4	M8	M10
$C_5-C_6-C_7-C_8$	11.93	48.05	17.81
$C_6-C_7-C_8-C_9$	7.67	-2.20	2.69
$C_7-C_8-C_9-C_{10}$	13.47	-8.80	0.15
$C_8-C_9-C_{10}-C_{11}$	-6.09	1.28	1.94
$C_9-C_{10}-C_{11}-C_{12}$	17.15	-2.58	3.95
$C_{10}-C_{11}-C_{12}-C_{13}$	-8.28	0.08	3.91
$C_{11}-C_{12}-C_{13}-C_{14}$	3.60	-2.40	2.86
$C_{12}-C_{13}-C_{14}-C_{15}$	2.81	-1.49	0.72
$C_{13}-C_{14}-C_{15}-N$	23.74	10.51	-2.56

Table 5.2: Average geometries for the three proteins: bond length alternation (left); distortions (right).

The retinal residue in M4 is also more distorted, with significant distortions around the $C_{10}-C_{11}$ and C_8-C_9 bonds, whereas the chain is nearly planar in M8 and M10. However, although M8 and M10 are highly similar along the conjugated chain, they differ significantly in the ionone ring region, with twistings around the C_6-C_7 bond of 48.05° and 17.81° for M8 and M10 respectively. This is quite unexpected, since the only amino-acid changing from M8 to M10 is located near the lysine side of the retinal molecule. Inspection of the crystal structure shows a different conformation of the W58 and L29 residues, located in the ionone ring region, as highlighted on figure 5.10. In M8, W58 points toward the bulk, while the binding pocket of the retinal is "closed" by the L29 amino-acid. The reverse situation appears in M10, where W58 is closer to the ionone ring. This difference in environment may explain the different distorsions observed for these two systems. We should also point out that this difference is inherent to the crystal structures published, and it is still unclear if this really reflects the situation in solution. Indeed, it is possible that the **Q4R** mutation, which triggers the change in conformation of the $C_{15}-N$ bond, induces an elongation of the retinal chain, and thus perturbs the initial arrangement of the amino-acids on the other side of the molecule. However, this could also be a flaw in the crystallization process, and it is also possible that the starting structure does not reflect the situation in solution. To investigate further on this point, more extensive classical dynamics would be required (see below).

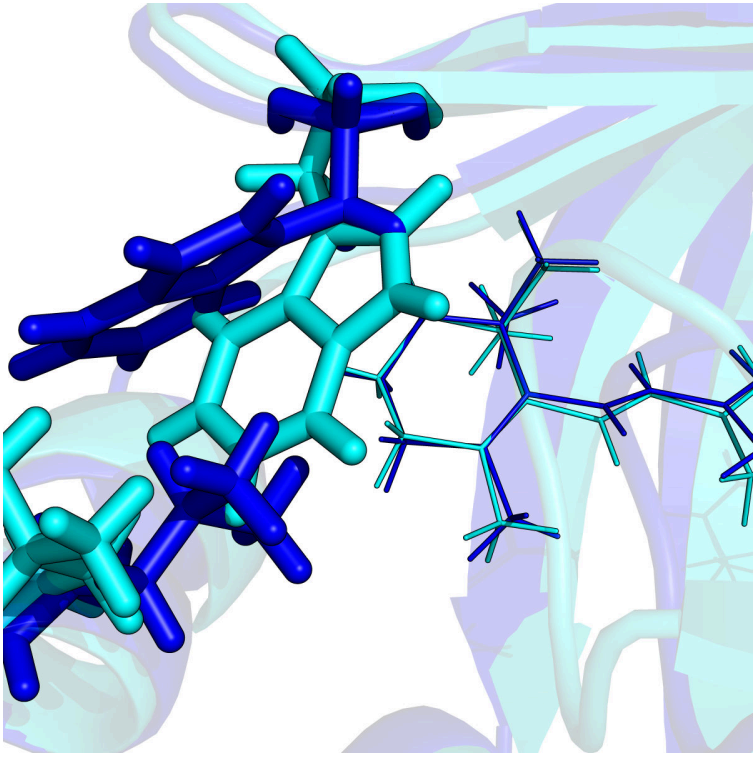


Figure 5.10: Overlap structures of M8 (blue) and M10 (cyan) in the ionone ring region. The W58 and L29 residues are highlighted in sticks representation.

Vertical absorptions

M4 The vertical absorptions values obtained for M4 at SS- and MS-CASPT2 levels are presented on figure 5.11, and the corresponding oscillator strength and wavefunctions are displayed on tables 5.3 (CASSCF wavefunctions) and 5.4 (perturbatively modified (PM)-CASSCF wavefunctions, obtained after MS-CASPT2 computations).

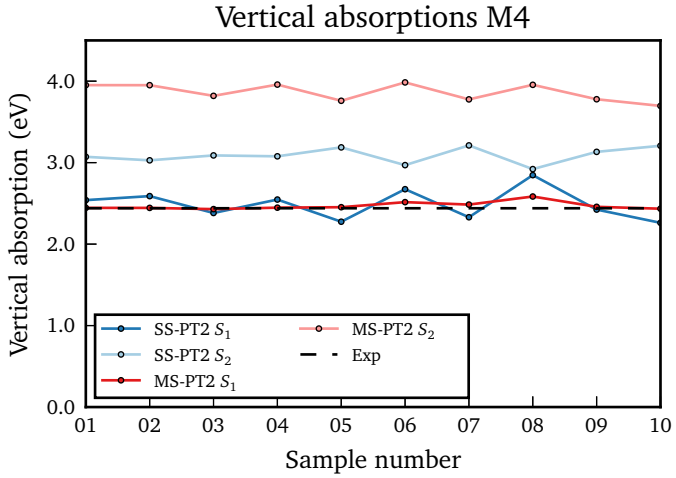


Figure 5.11: Vertical absorptions obtained at SS- and MS-CASPT2 levels for M4 protein. The labeling of the states for SS-CASPT2 refers to the corresponding CASSCF state.

These results clearly show the strong influence of the H-bonding network on the transition energies computed, especially when looking at the results obtained at the CASPT2 level. The values computed oscillate around the experimental absorption, and near-degeneracy occur for at least two points (samples 06 and 08). All along the trajectory, the CASSCF wavefunctions of the S_1 and S_2 states show a high level of mixing, and are both dominated by a combination of the $H \rightarrow L$ and $(H \rightarrow L)^2$ excitations in most samples. The $H \rightarrow L+1$ (for S_1) and $H-1 \rightarrow L$ (for S_2) excitations also have non-negligible weights. Oscillator strengths computed reflect this mixing, as in all the samples both $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ transitions are allowed, with a brighter state computed for the root that has the more significant $H \rightarrow L$ character. In most samples, this excitation is dominant in both excited states computed, resulting in similar oscillator strength for both transitions. Surprisingly, the sample which has the cleaner $H \rightarrow L$ character (sample 08) is also the one that gives the worst agreement with experimental values.

The varying high level of mixing along the trajectory highlights the importance of the initial sampling carried out, and makes the accurate assignment of the bright state for each sample a non-trivial task. Considering the transition with the highest oscillator strength for each sample, the average vertical absorption is computed at 2.67 eV, blue-shifted by 0.14 eV from the experimental maximum (2.44 eV). This is close to the standard 0.2 eV error accepted for the methodology used in this study. However, this average may not be accurate enough, given the mixing already highlighted. Taking into account MS-CASPT2 results help to largely reduce this error.

	CASSCF Root	CASSCF WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.75		
	2	$H \rightarrow L$ 0.34, $(H \rightarrow L)^2$ 0.13, $H \rightarrow L+1$ 0.13	2.54	0.45
	3	$H \rightarrow L$ 0.35, $(H \rightarrow L)^2$ 0.15	3.07	0.63
Sample 02	1	GS 0.75		
	2	$H \rightarrow L$ 0.31, $(H \rightarrow L)^2$ 0.14, $H \rightarrow L+1$ 0.14	2.59	0.41
	3	$H \rightarrow L$ 0.38, $(H \rightarrow L)^2$ 0.14	3.03	0.64
Sample 03	1	GS 0.75		
	2	$H \rightarrow L$ 0.42, $H \rightarrow L+1$ 0.12	2.38	0.58
	3	$H \rightarrow L$ 0.27, $(H \rightarrow L)^2$ 0.19	3.09	0.52
Sample 04	1	GS 0.75		
	2	$H \rightarrow L$ 0.34, $H \rightarrow L+1$ 0.13, $(H \rightarrow L)^2$ 0.13	2.55	0.45
	3	$H \rightarrow L$ 0.35, $(H \rightarrow L)^2$ 0.15	3.08	0.61
Sample 05	1	GS 0.75		
	2	$H \rightarrow L$ 0.51, $H \rightarrow L+1$ 0.11	2.27	0.67
	3	$(H \rightarrow L)^2$ 0.23, $H \rightarrow L$ 0.19, $H-1 \rightarrow L$ 0.11	3.19	0.39
Sample 06	1	GS 0.75		
	2	$H \rightarrow L$ 0.27, $(H \rightarrow L)^2$ 0.16, $H \rightarrow L+1$ 0.13	2.67	0.38
	3	$H \rightarrow L$ 0.43, $(H \rightarrow L)^2$ 0.12	2.97	0.75
Sample 07	1	GS 0.75		
	2	$H \rightarrow L$ 0.50, $H \rightarrow L+1$ 0.11	2.33	0.68
	3	$(H \rightarrow L)^2$ 0.22, $H \rightarrow L$ 0.19, $H-1 \rightarrow L$ 0.11	3.21	0.4
Sample 08	1	GS 0.75		
	3	$H \rightarrow L$ 0.56	2.85	0.86
	2	$(H \rightarrow L)^2$ 0.22, $H \rightarrow L$ 0.13, $H \rightarrow L+2$ 0.13	2.92	0.19
Sample 09	1	GS 0.75		
	2	$H \rightarrow L$ 0.44, $(H \rightarrow L)^2$ 0.12	2.42	0.61
	3	$H \rightarrow L$ 0.24, $(H \rightarrow L)^2$ 0.19	3.13	0.48
Sample 10	1	GS 0.75		
	2	$H \rightarrow L$ 0.52, $H \rightarrow L+1$ 0.11	2.26	0.66
	3	$(H \rightarrow L)^2$ 0.23, $H \rightarrow L$ 0.17, $H-1 \rightarrow L$ 0.12	3.21	0.36

Table 5.3: CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M4 protein.

Indeed, the multi-state method helps to lift near-degeneracies between states, and thus gives a clean perturbatively modified (PM) CASSCF wavefunction for all the roots computed, . This results in S_1 and S_2 states clearly dominated by the $H \rightarrow L$ and $(H \rightarrow L)^2$ excitations respectively, and with a clearly identifiable S_1 bright state for all the samples. The vertical absorptions computed at the MS-CASPT2 level agree well with the experimental value all along the trajectory. Our computations give an average

vertical absorption of 2.47 eV, which is perfectly in line with the experimental data. Therefore, the multi-state approach may be more appropriate than single-state in the case of M4, although the off-diagonal terms in the effective haimiltonian are found to be around 0.01 a.u., and are thus above the commonly accepted threshold of 0.002 a.u.

Overall, the retinal in M4 appears highly similar to what has previously been observed for solvated retinal molecules, where high degrees of mixing were found at the ground state geometry. This is not surprising, as the retinal in M4 is the most hydrated when compared to M8 and M10, as previously highlighted (see section 5.2.3). Therefore, similar features between the two systems are not unexpected and will be found all along this study.

	PM-CASSCF Root	PM-CASSCF WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.77		
	2	$H \rightarrow L$ 0.66	2.44	1.62
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.14	3.95	0.086
Sample 02	1	GS 0.77		
	2	$H \rightarrow L$ 0.65	2.44	1.55
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.14	3.95	0.093
Sample 03	1	GS 0.77		
	2	$H \rightarrow L$ 0.67	2.43	1.74
	3	$(H \rightarrow L)^2$ 0.29, $H-1 \rightarrow L$ 0.15	3.82	0.04
Sample 04	1	GS 0.77		
	2	$H \rightarrow L$ 0.65	2.45	1.61
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.14	3.96	0.077
Sample 05	1	GS 0.79		
	2	$H \rightarrow L$ 0.68	2.45	1.81
	3	$(H \rightarrow L)^2$ 0.30, $H-1 \rightarrow L$ 0.15	3.76	0.022
Sample 06	1	GS 0.77		
	2	$H \rightarrow L$ 0.64	2.52	1.65
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.14	3.98	0.13
Sample 07	1	GS 0.78		
	2	$H \rightarrow L$ 0.68	2.49	1.79
	3	$(H \rightarrow L)^2$ 0.29, $H-1 \rightarrow L$ 0.15	3.78	0.019
Sample 08	1	GS 0.76		
	2	$H \rightarrow L$ 0.61	2.58	1.48
	3	$(H \rightarrow L)^2$ 0.26, $H-1 \rightarrow L$ 0.12	3.96	0.27
Sample 09	1	GS 0.77		
	2	$H \rightarrow L$ 0.67	2.46	1.68
	3	$(H \rightarrow L)^2$ 0.29, $H-1 \rightarrow L$ 0.15	3.78	0.029
Sample 10	1	GS 0.78		
	2	$H \rightarrow L$ 0.67	2.43	1.75
	3	$(H \rightarrow L)^2$ 0.29, $H-1 \rightarrow L$ 0.15	3.70	0.026

Table 5.4: PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at the MS-CASPT2 level, obtained for the ten snapshots for M4 protein.

M8 The vertical absorptions along the trajectory, as well as the computed CASSCF and PM-CASSCF wavefunctions, and the corresponding oscillator strengths are displayed on figure 5.12 and tables 5.5 (CASSCF wavefunctions) and 5.6 (PM-CASSCF wavefunctions).

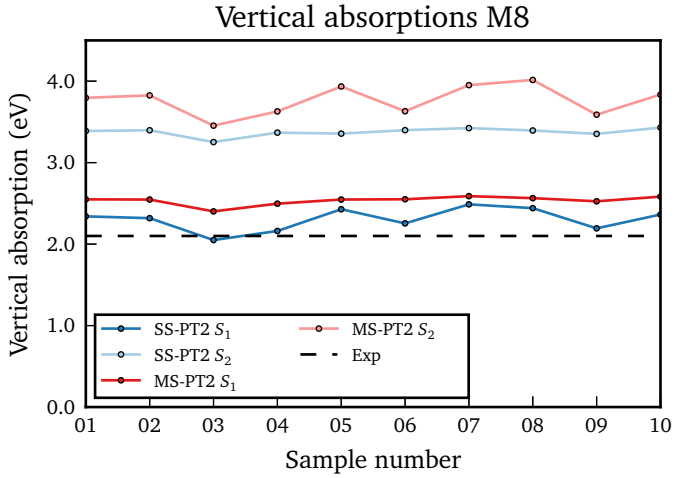


Figure 5.12: Vertical absorptions obtained at SS- and MS-CASPT2 levels for M8 protein. The labeling of the states for SS-CASPT2 refers to the corresponding CASSCF state.

Here again, the vertical absorption values are quite sensitive to the hydrogen bonding network. However, contrary to what is observed in M4, most values at the CASPT2 level are blue-shifted with respect to the experimental value, and only four point are in good agreement with the experiments: sample 03, 04, 06 and 09. Looking at the CASSCF wavefunctions for all the points, the main difference between samples 03, 04, 06 and 09 and the other points is that in these four samples, the S_2 state does not have significant contributions from the $H \rightarrow L$ excitation; this excitation also plays a greater role in the S_1 state. Therefore, the oscillator strengths for the two excitations are well separated in this set of points, whereas the remaining points feature the both transitions on a same level.

Two sets can thus be extracted from the initial sampling:

- "matching" set: samples that match the experimental absorption, *i.e.* samples 03, 04, 06 and 09;
- "non matching" set: all the remaining samples.

It should be noticed that for all the points in the "non matching" set, all S_1 states are clearly dominated by the $H \rightarrow L$ excitation, and thus have an higher oscillator strength. This is different from what happens in M4, where S_1 and S_2 had similar configurations composition. As a consequence, even if a multi-state treatment clears the S_2 state in M8, it does not make much sense to apply it in this system.

	CASSCF Root	CASSCF WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.76		
	2	$H \rightarrow L$ 0.57	2.34	0.81
	3	$(H \rightarrow L)^2$ 0.26, $H \rightarrow L$ 0.13, $H-1 \rightarrow L$ 0.13	3.39	0.32
Sample 02	1	GS 0.76		
	2	$H \rightarrow L$ 0.57	2.32	0.79
	3	$(H \rightarrow L)^2$ 0.26, $H \rightarrow L$ 0.13, $H-1 \rightarrow L$ 0.13	3.40	0.33
Sample 03	1	GS 0.76		
	2	$H \rightarrow L$ 0.66	2.05	1.03
	3	$(H \rightarrow L)^2$ 0.31, $H-1 \rightarrow L$ 0.19	3.25	0.15
Sample 04	1	GS 0.76		
	2	$H \rightarrow L$ 0.63	2.16	0.92
	3	$(H \rightarrow L)^2$ 0.30, $H-1 \rightarrow L$ 0.16	3.37	0.2
Sample 05	1	GS 0.76		
	2	$H \rightarrow L$ 0.51	2.43	0.72
	3	$(H \rightarrow L)^2$ 0.24, $H \rightarrow L$ 0.20, $H-1 \rightarrow L$ 0.11	3.36	0.43
Sample 06	1	GS 0.76		
	2	$H \rightarrow L$ 0.64	2.25	0.96
	3	$(H \rightarrow L)^2$ 0.29, $H-1 \rightarrow L$ 0.17	3.40	0.18
Sample 07	1	GS 0.76		
	2	$H \rightarrow L$ 0.52	2.49	0.74
	3	$(H \rightarrow L)^2$ 0.24, $H \rightarrow L$ 0.19, $H-1 \rightarrow L$ 0.12	3.42	0.41
Sample 08	1	GS 0.76		
	2	$H \rightarrow L$ 0.50	2.44	0.7
	3	$(H \rightarrow L)^2$ 0.24, $H \rightarrow L$ 0.20, $H-1 \rightarrow L$ 0.11	3.39	0.44
Sample 09	1	GS 0.76		
	2	$H \rightarrow L$ 0.64	2.19	0.98
	3	$(H \rightarrow L)^2$ 0.30, $H-1 \rightarrow L$ 0.17	3.35	0.17
Sample 10	1	GS 0.76		
	2	$H \rightarrow L$ 0.58	2.36	0.82
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.13, $H \rightarrow L$ 0.13	3.43	0.31

Table 5.5: CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M8 protein.

As seen on figure 5.12, all the MS-CASPT2 values are even more blue-shifted than their CASPT2 counterpart. For all these reasons, we can safely rule out the MS-CASPT2 treatment in this case. The vertical absorption averaged over all the samples at the CASPT2 level is computed at 2.38 eV, while the average over the matching samples only is computed at 2.13 eV, compared to an experimental value of 2.10 eV.

	PM-CASSCF Root	PM-CASSCF WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.55	1.83
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.17	3.80	0.029
Sample 02	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.55	1.87
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.17	3.83	0.03
Sample 03	1	GS 0.79		
	2	$H \rightarrow L$ 0.68	2.40	1.96
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.20	3.45	0.1
Sample 04	1	GS 0.79		
	2	$H \rightarrow L$ 0.69	2.50	1.94
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.19	3.63	0.061
Sample 05	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.55	1.79
	3	$(H \rightarrow L)^2$ 0.31, $H-1 \rightarrow L$ 0.16	3.93	0.03
Sample 06	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.55	1.85
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.19	3.63	0.05
Sample 07	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.59	1.73
	3	$(H \rightarrow L)^2$ 0.31, $H-1 \rightarrow L$ 0.17	3.95	0.03
Sample 08	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.56	1.82
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.17	4.01	0.03
Sample 09	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.53	1.92
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.19	3.59	0.05
Sample 10	1	GS 0.79		
	2	$H \rightarrow L$ 0.70	2.58	1.84
	3	$(H \rightarrow L)^2$ 0.32, $H-1 \rightarrow L$ 0.17	3.84	0.024

Table 5.6: PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at MS-CASPT2 level, obtained for the ten snapshots for M8 protein.

M10 The vertical absorptions along the trajectory, as well as the computed CASSCF and PM-CASSCF wavefunctions, and the corresponding oscillator strengths displayed on figure 5.13 and tables 5.7 (CASSCF wavefunctions) and 5.8 (PM-CASSCF wavefunctions).

In M10, the hydrogen bonding network has far less influence than in previous cases, which was expected since this system has the least hydrated retinal. Thus, water molecules are expected to have

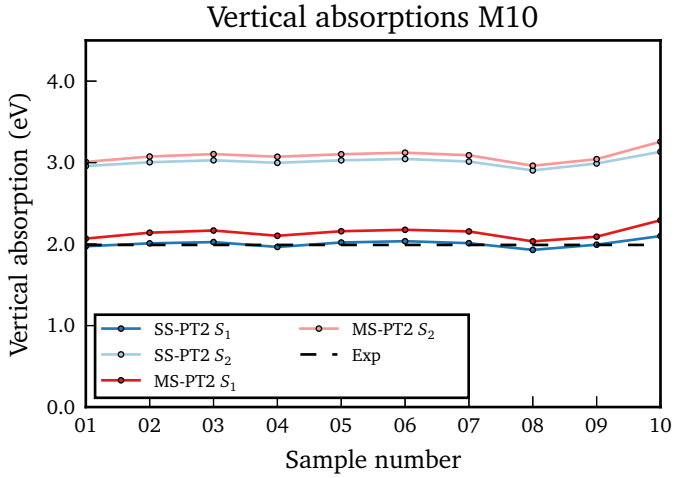


Figure 5.13: Vertical absorptions obtained at SS- and MS-CASPT2 levels for M10 protein. The labeling of the states for SS-CASPT2 refers to the corresponding CASSCF state.

little effect on the properties of this system. This is indeed what is observed, with a nearly constant vertical absorption all along the trajectory, at both SS- and MS-CASPT2, both in good agreement with experimental data. Inspection of the wavefunctions (both CASSCF and PM-CASSCF) shows that S_1 and S_2 have similar character along the trajectory, and are dominated by the $H \rightarrow L$ and a combination of $(H \rightarrow L)^2$ and $H-1 \rightarrow L$ excitations respectively, at both level of theory. The computed oscillator strengths show a clear S_1 bright state.

	CASSCF Root	CASSCF WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.79		
	2	$H \rightarrow L$ 0.65	1.97	1.28
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.27	2.96	0.2
Sample 02	1	Gs 0.78		
	2	$H \rightarrow L$ 0.63	2.01	1.2
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.25	3.00	0.21
Sample 03	1	GS 0.77		
	2	$H \rightarrow L$ 0.62	2.02	1.19
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.24	3.03	0.2
Sample 04	1	GS 0.78		
	2	$H \rightarrow L$ 0.63	1.97	1.17
	3	$(H \rightarrow L)^2$ 0.26, $H-1 \rightarrow L$ 0.26	3.00	0.22
Sample 05	1	GS 0.77		
	2	$H \rightarrow L$ 0.62	2.02	1.19
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.24	3.03	0.2
Sample 06	1	GS 0.77		
	2	$H \rightarrow L$ 0.62	2.04	1.15
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.25	3.05	0.22
Sample 07	1	GS 0.77		
	2	$H \rightarrow L$ 0.62	2.01	1.18
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.25	3.01	0.21
Sample 08	1	GS 0.79		
	2	$H \rightarrow L$ 0.65	1.93	1.3
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.27	2.90	0.2
Sample 09	1	GS 0.78		
	2	$H \rightarrow L$ 0.65	1.99	1.24
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.27	2.99	0.21
Sample 10	1	GS 0.75		
	2	$H \rightarrow L$ 0.58	2.10	1.02
	3	$(H \rightarrow L)^2$ 0.26, $H-1 \rightarrow L$ 0.22	3.13	0.24

Table 5.7: CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at CASPT2 level, obtained for the ten snapshots for M10 protein.

Averaging the vertical absorption at the SS-CASPT2 and MS-CASPT2 levels give values of 2.00 eV and 2.14 eV respectively, in very good agreement with the experimental value (1.99 eV).

	State	WF	$\Delta E(S_0 \rightarrow S_n)$	f
Sample 01	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.07	1.7
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.26	3.01	0.27
Sample 02	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.14	1.75
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.25	3.07	0.24
Sample 03	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.17	1.78
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.24	3.10	0.22
Sample 04	1	GS 0.80		
	2	$H \rightarrow L$ 0.63	2.10	1.73
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.26	3.07	0.28
Sample 05	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.16	1.75
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.24	3.10	0.22
Sample 06	1	GS 0.80		
	2	$H \rightarrow L$ 0.63	2.18	1.73
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.25	3.12	0.24
Sample 07	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.16	1.76
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.25	3.09	0.24
Sample 08	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.03	1.74
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.26	2.96	0.28
Sample 09	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.09	1.67
	3	$(H \rightarrow L)^2$ 0.27, $H-1 \rightarrow L$ 0.26	3.04	0.28
Sample 10	1	GS 0.80		
	2	$H \rightarrow L$ 0.64	2.29	1.82
	3	$(H \rightarrow L)^2$ 0.28, $H-1 \rightarrow L$ 0.22	3.26	0.17

Table 5.8: PM-CASSCF wavefunctions, vertical absorptions and oscillator strengths, with energies computed at MS-CASPT2 level, obtained for the ten snapshots for M10 protein.

Summary In this section we have considered the vertical absorptions computed at the SS- and MS-CASPT2 levels, as well as the wavefunctions (CASSCF or PM-CASSCF), to probe the validity of our computational model. Our findings are summarized on table 5.9. Overall, we obtained a good agreement between the averages computed along the trajectories and the experimental absorption maximum. However, the three systems display some very different behaviours. We showed that a multi-state treat-

ment was required for M4, where the S_1 and S_2 states are mixed. The case of M8 is more problematic, as two sets of samples have been extracted out of the trajectory; only one of these sets has a correct vertical absorption, and consistent wavefunctions and oscillator strengths. The other points are completely blue-shifted, but do not display the same level of mixing as observed in M4, thus ruling out a multi-state improvement. This separation in two sets will also be highlighted when the excited states of these proteins will be studied (see below). Eventually, M10 has the easiest treatment, since the hydrogen bonding network has little effects on its properties.

Structure	Level	Mean Value (eV)	Exp.	Diff.
M4	CASPT2	2.67	2.53	0.14
	MS-CASPT2	2.47		0.06
M8	CASPT2	2.13	2.1	0.03
	MS-CASPT2	2.47		0.37
M10	CASPT2	2.00	1.99	0.01
	MS-CASPT2	2.14		0.15

Table 5.9: Average vertical absorptions obtained for each protein, at the CASPT2 and MS-CASPT2 levels. The values indicated for M8 are those obtained out of the matching set only.

In the following sections, we will investigate deeper on the differences in the electronic structure of the retinal in the three proteins, in order to better account for the results highlighted above. In particular, we will quantify the electrostatic effect of the environment, and we will study the details of the charge transfer occurring upon excitation in the Franck-Condon region.

5.3.2 Dipole moments and charges variations

Dipole moments The permanent dipoles, extracted from the CASSCF wavefunctions, have been collected for all the ten samples for each protein, and the results are gathered on figure 5.14.

In M4, huge variations in direction are observed for the S_1 dipoles, which in some cases overlap with S_2 dipoles. The direction of the dipole moment in S_0 is consistent with the localization of the charge in the ground state on the nitrogen side of the retinal molecule. Upon excitation, a charge transfer along the retinal conjugated chain is expected, as seen in solvated retinals (see chapter 5) or in rhodopsin. However, the direction of the S_1 dipole moments indicates a quite ineffective charge transfer in this

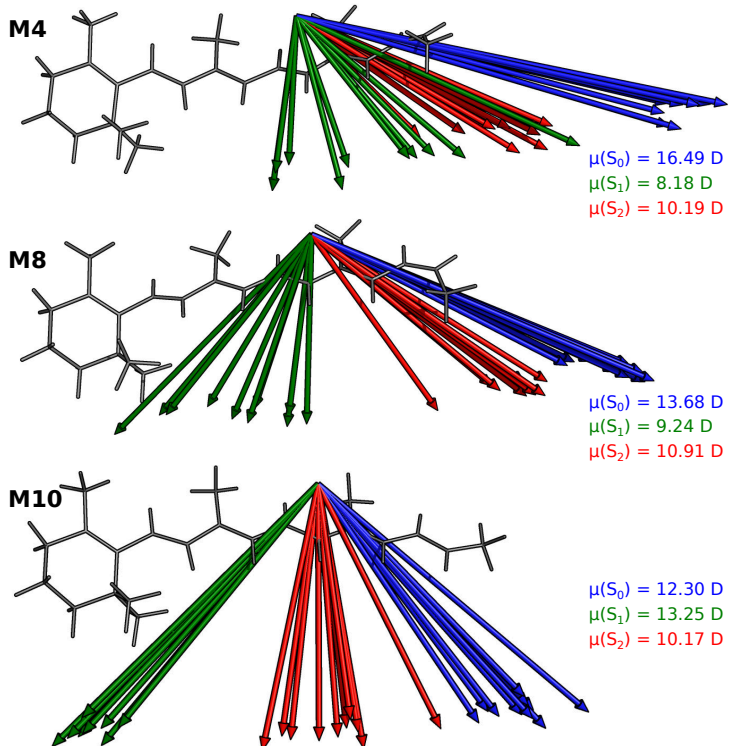


Figure 5.14: Permanent dipole moments collected along the trajectory for the three systems. Blue: S_0; Green: S_1; Red: S_2. The average values of the norm of the dipole moments for each state is also indicated.

protein. The intensity of the dipole moment also decreases sharply when going from S_0 to S_1, which indicates that the charge separation is low on the excited state. Finally, here again the huge influence of the HBN is clearly noticeable, given the wide variations in directions.

In M8, there is also a wide variation in the directions in S_1 dipole moments. However, the S_1 dipole moments are more oriented toward the ionone ring than in M4, indicating a more effective charge transfer in this protein. The increasing value of the intensity of this dipole is another proof of this more efficient charge transfer. Finally, we can notice that the S_0 dipole are less intense than in M4, which indicate a lower localization of charges in the ground state.

Finally, in M10, the dispersion of directions is reduced, consistently with the efficient shielding of the retinal from external environment highlighted above (see section 5.3.1). Also, the intensities of S_1 and S_0 dipoles are in the same range (13.25 D and 12.30 D respectively), indicating a very effective charge transfer.

More insight of the charge transfer efficiency can be obtained by computing the difference in electrostatic dipole moment as:

$$\Delta\vec{\mu}(S_0 \rightarrow S_n) = \vec{\mu}_{S_n} - \vec{\mu}_{S_0} \tag{5.2}$$

The results are gathered on figure 5.15. The evolution of the norm of $\Delta\mu(S_0 \rightarrow S_1)$ from M4 to M10 indicates that the charge transfer intensity is reduced in M4, and way more effective in M10. The influence of the HBN, here again, is more pronounced in M4 than in the two other proteins. In the three proteins, the direction of the difference dipole indicates a charge transfer from the nitrogen side to the ionone side of the retinal, consistently with what is expected for this system.

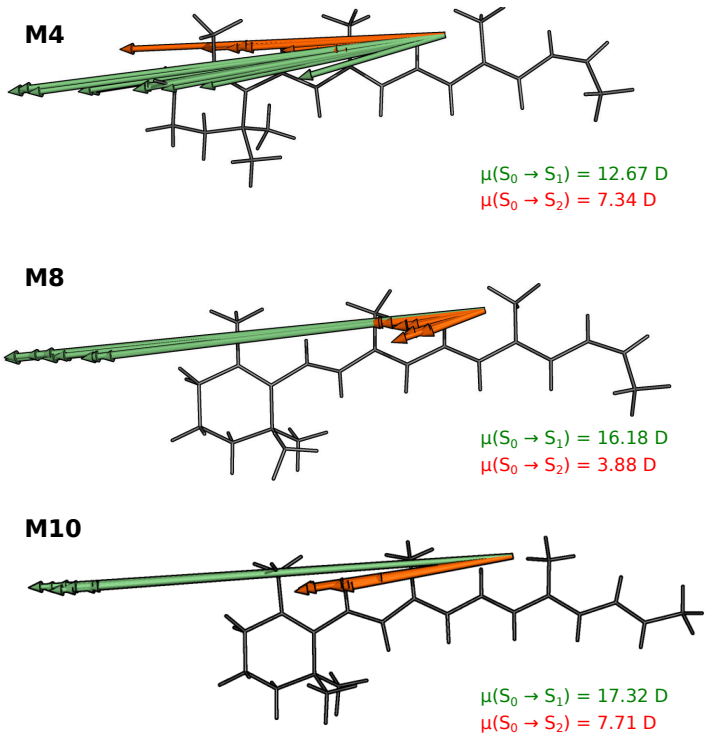


Figure 5.15: Difference in permanent dipole moment for $S_0 \rightarrow S_1$ (green) and $S_0 \rightarrow S_2$ (orange) transitions. The average values of the norm of the two transitions are also indicated.

Finally, we can mention that dipoles (both permanent and difference in permanent) coming from the "matching" set defined before have very close features with the ones coming from M10. A comparison is presented on table 5.10. Concerning the bright $S_0 \rightarrow S_1$ transition, the difference in dipole moments has the same intensity in both M8 matching and M10 systems, although the norm of the permanent

S_1 dipoles differ by around 3 Debye. This may be due to different character in internal charge transfer, which are to be investigated on in the next section.

	M8 – matching	M8 – non-matching	M10
S_0	12.247	14.643	12.299
S_1	10.512	8.387	13.247
S_2	10.143	11.418	10.172
$S_0 \rightarrow S_1$	17.356	15.389	17.320
$S_0 \rightarrow S_2$	3.471	4.155	7.707

Table 5.10: Dipole moment comparison between M8 "matching", M8 "non-matching", and M10 dipole moments, averaged over the ten snapshots.

Mulliken charges analysis The Mulliken charge analysis is carried out on the CASSCF wavefunctions. For each sample, these charges are computed for all atoms, and the hydrogen charges are summed up on heavy atoms. As was done in solvated retinals (see chapter 5), the retinal molecule is divided into two parts by cutting it at the $C_{11}-C_{12}$ bond. The two parts will be referred to as the lysine and ionone part, respectively. The charges of each atom on each part is summed up, and the average over the ten snapshots of each protein is computed. The results are presented on figure 5.16. This view completes the conclusions drawn out of the dipole moments analysis conducted above.

In particular, the stronger localization of the charge on the lysine side in M4 in the ground state (S_0) compared to M8 and M10 is clearly demonstrated. The lysine side indeed accounts for around 90% of the positive charge in M4, whereas this amounts to around 75% in M10. The charge transfer efficiency is computed as the amount of charge that is transferred from one side to the other upon excitation. We found that on average, 27%, 41% and 46% of the charge initially on the lysine side is transferred to the ionone side for M4, M8 and M10 respectively. This result confirms our conclusion that the charge transfer is significantly reduced in M4 with respect to what happens in the two other systems. In this domain again, the solution-like behaviour of M4 is still noticeable, as the computed internal charge transfer is comparable to what has previously been found for retinal solvated in methanol.

Here again, we should stress that when only taking into account the matching set in M8, the amount of charge transfered from the lysine side to the ionone side rises to 47%, and is therefore very close to the value obtained in M10.

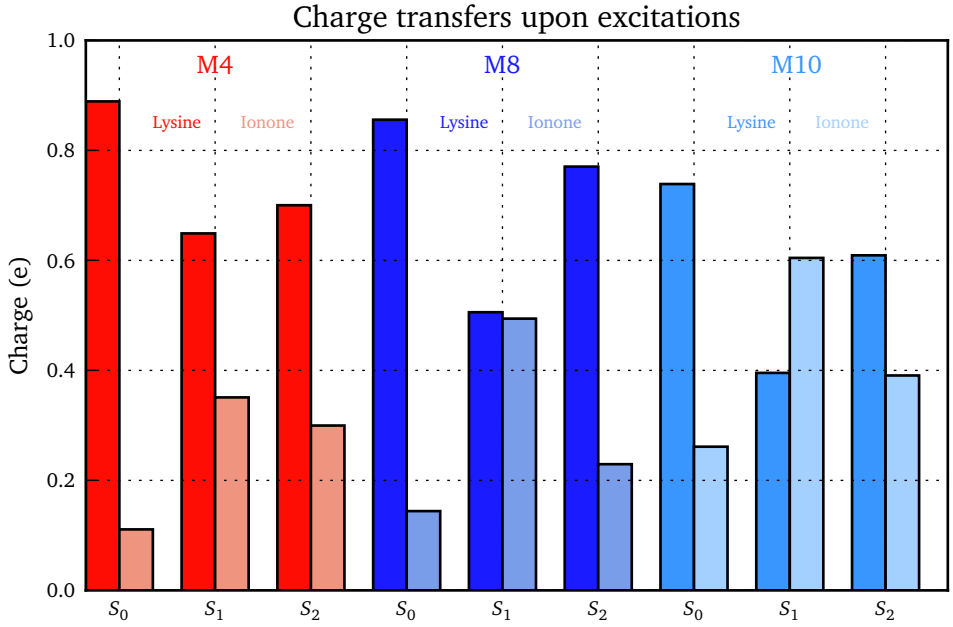


Figure 5.16: Mulliken charges repartition averaged on top of ten snapshots for each protein, for the two sides of the retinal molecule. The light color represents the charge on the ionone side, while the brighter one is for the lysine side of the retinal.

To investigate more in depth on the repartition of the charge in the different systems, a per-atom analysis was carried out, in which we consider only the atoms in the conjugated chain of the retinal. In this analysis, the charge carried by methyl groups are summed on the heavy atoms carrying the group in question. For instance, all the charges in the ionone ring are added to the one of C_5, and the charge carried by the CH_3 group capping the lysine side is added to the nitrogen of retinal. The charges obtained are then averaged for each state in each protein. The graph obtained is presented on figure 5.17. The bonds in M4 in the S_0 state appear more polarized than in the other proteins, with positive and negative charges alternating in the $C_{15}-C_9$ region; M8 presents the same pattern, although less pronounced. In M10, the positive charge is evenly spread across the $C_{13}-C_8$ region. The positive charge also appears strongly located on the $N-C_{15}$ bond in both M4 and M8. When exciting M4 to S_1, the charge of C_{10} and C_8 increases to positive values, while the polarization of bonds in the $C_{13}-C_{10}$ region is inverted. Overall, the polarization of bond is reduced in S_1. In M8, the same situation is observed, although with a far more positive charge transfered to $C_{10}-C_7$, and an even smaller polarization of bonds. In M10, a different picture appears. As expected from the bar diagrams presented above (figure 5.16), the positive

tail is transfered farther toward the ionone ring, while the relative values of the charge carried by atoms in the $C_{13}-C_{10}$ region is inverted.

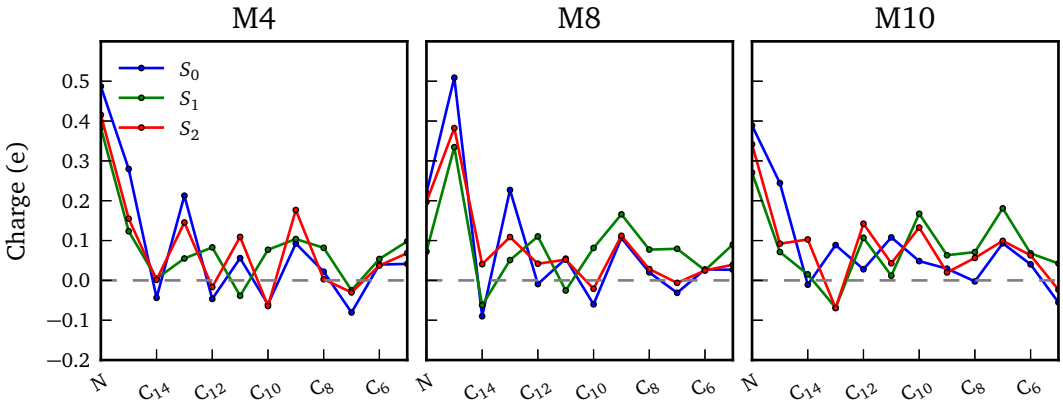


Figure 5.17: Average charges per atom for the S_0, S_1 and S_2 states for M4, M8 and M10.

5.3.3 Effects of the environment

Gas phase computations

To investigate more in depth on the effect of the protein scaffold, we extracted the retinal molecule as defined in the high layer from the proteins, after QM/MM geometry optimizations. Without any re-optimization, we re-computed the energies of the S_0, S_1 and S_2 states at the SS- and MS-CASPT2 levels. By calculating the difference between the energies obtained in protein and those obtained *in vacuo*, we expect to get insight on differential effects of the set of point charges used for describing the protein on the different states involved. All the results in this section were computed using the 6-31G* basis set.

Structure	Level	Average (eV)		Std. Dev. (eV)	
		Protein	*In vacuo*	*Protein*	*In vacuo*
M4	CASPT2	2.66	1.98	0.18	0.01
	MS-CASPT2	2.58	2.24	0.04	0.02
M8	CASPT2	2.47	2.17	0.14	0.02
	MS-CASPT2	2.63	2.40	0.07	0.02
M10	CASPT2	2.09	2.00	0.05	0.01
	MS-CASPT2	2.24	2.22	0.07	0.01

Table 5.11: Vertical absorptions obtained *in vacuo*, on top of QM/MM optimized geometries, averaged over the ten snapshots for each proteins. Results in protein are included for comparison (6-31G* basis set).

As can be seen on the table, the standard deviations obtained *in vacuo* are very low, and indicate that the vertical absorption is very stable along the trajectory. Moreover, the values obtained for the retinals extracted out of the three proteins are close to one another (1.98 eV, 2.17 eV and 2.00 eV for M4, M8 and M10 respectively at SS-CASPT2 level), indicating that the tuning of the vertical absorption is only due to the protein scaffold, and that the geometry of the retinal plays little role. These values obtained in good agreement with previous computational studies, and the blue-shift of M8 can be attributed to the twisted conformation of the ionone ring obtained [20] (see figure 5.9). Eventually, it is also noticeable that in M10, the results in protein and *in vacuo* are more or less the same, stressing the fact that in M10, the retinal is efficiently shielded from the external environment, and behaves as a molecule in gas phase.

Analysis of the effect of the protein scaffold

In this part, we define the electrostatic effect of the protein scaffold as the difference

$$E_{electro} = E_{protein} - E_{gas}, \tag{5.3}$$

where $E_{protein}$ is the QM energy obtained for the retinal embedded in the protein scaffold, and E_{gas} is the QM energy of the retinal in gas phase, considered in the exact same geometry. This difference is computed for the S_0, S_1 and S_2 states for all the snapshots considered before. A positive difference indicates that the protein destabilizes the state considered.

The results (computed with CASPT2 QM energies) obtained for the three protein are presented on figure 5.18. First, we can notice that the protein scaffold stabilizes all the states. However, all the states are not equally affected. In M4, the protein stabilizes more S_0 and S_2 than S_1, leading to a blue shift of the vertical absorption with respect to gas phase.

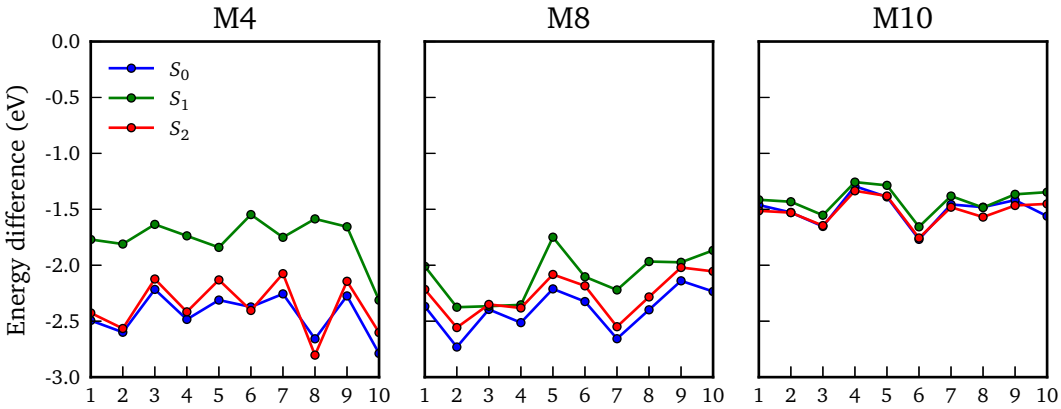


Figure 5.18: Electrostatic effect computed for the M8 protein.

As shown above (see section 5.3.2), the charge in M4 is strongly located on the nitrogen in M4. Locally, the charge is expected to be stabilized by the HBN involving the close water molecule and Gln4 residue, thus over-stabilizing S_0 and S_2.

In M10, all the states are similarly stabilized, and no differential effect can be observed. This is consistent with the shielding from external environment already highlighted. The stabilization with respect to gas phase may be due to the global counterion formed by the whole model (protein + water solvent molecules). Focusing on the matching sample of M8 (*i.e.* samples 03, 04, 06 and 09), a similar feature is observed, in which no differential effect is computed for the three states. However, the overall stabilization is higher than in M10. This is probably due to the mutation of Gln to Arg at position 4 in M10, leading to another positive charge at close proximity with the protonated Schiff base, thus less stabilizing the system.

5.3.4 Summary: Ground state properties

In this section we have presented the results obtained out of the QM/MM optimization of ten samples extracted from a classical MM trajectory, for each protein. First, we have shown that our model is able to reproduce fairly well the experimental vertical absorptions. Analyzing the wavefunction components,

we demonstrated that a strong level of mixing is observed at the CASSCF level in M4, similarly to what was previously observed in solvated retinal molecule. Concerning M8, two subsets of structures were considered. Indeed, only four out of ten samples reproduce the experimental absorption, and these are characterized by cleaner wavefunctions and oscillator strengths; they appear very close to what we computed for M10, where a low sensitivity to the initial sampling is observed.

Comparisons of the permanent dipole moments and Mulliken charges allowed us to further investigate on the differences between the three systems. Wide variations in the direction of the dipole moments are observed for M4. However, most dipoles are directed toward the lysine side of the retinal, indicating an ineffective charge transfer upon $S_0 \rightarrow S_1$ transition, contrary to what happens in M8 and M10. Comparison of the norm of the dipoles also shows that the charge separation is higher in M4 than in M8 and M10 in the ground state, while this separation is quenched in the bright S_1 excited state in M4. This view was confirmed by analysis of the Mulliken charges.

Eventually, the effect of the geometry of the retinal and of the electrostatic environment provided by the protein scaffold on the vertical transitions has been investigated. We have shown that the geometry of the retinal contributes little to the absorption by means of *in vacuo* computations. We also attributed the blue-shift observed for M4 to an over-stabilization of the covalent state S_0 with respect to the ionic S_1. M8 and M10 were found to be far less sensitive to the scaffold. Especially, the retinal in M10 can be considered as very efficiently shielded from the rest of the system, and behaves much like a gas phase molecule.

5.4 Excited state optimizations

In this section we present the results obtained after relaxation of the excited states of the retinal residue. The relaxation has been carried out for the bright $H \rightarrow L$ root for all the samples, in each protein. First, we will highlight the geometrical parameters obtained for the excited state minimum, and we show that two different geometries can be obtained. Next, we will present the results arising from constrained scans between the two geometries, and show that the energy surface connecting them is flat, and that the energy difference between them is quite small, making it possible for the two of them to coexist. Therefore, the excited state manifold will be computed for these two geometries, as well as for the ground state geometry, to get more insight about the signals that may arise from the different species.

5.4.1 Geometries and vertical emissions

Optimizing the excited state is not straightforward, mainly due to root flipping that can occur in the process. This has especially been the case for all the samples in M4, and for some samples in M8. At CASSCF level, the $H \rightarrow L$ root is found to be root number 3 for all the samples in M4. This was also the case for the samples belonging to the non-matching set samples of M8. In M10 and the matching samples of M8, the $H \rightarrow L$ root is always root 2.

M4 The case of M4, as was previously highlighted, is highly similar to the one of retinal in solution presented in the previous chapter. Indeed, here again the retinal has been optimized in an EBL-like geometry, characterized by even bond lengths in the conjugated chain of the retinal. However, more variability among the samples, compared to what happens for ground state optimizations, is observed, as can be seen on figure 5.19. Most samples have a strong elongation of the $C_{13}-C_{14}$ bond, from an average of 1.363 Å in the ground state to 1.447 Å in the excited state, although it varies from 1.437 Å to 1.459 Å.

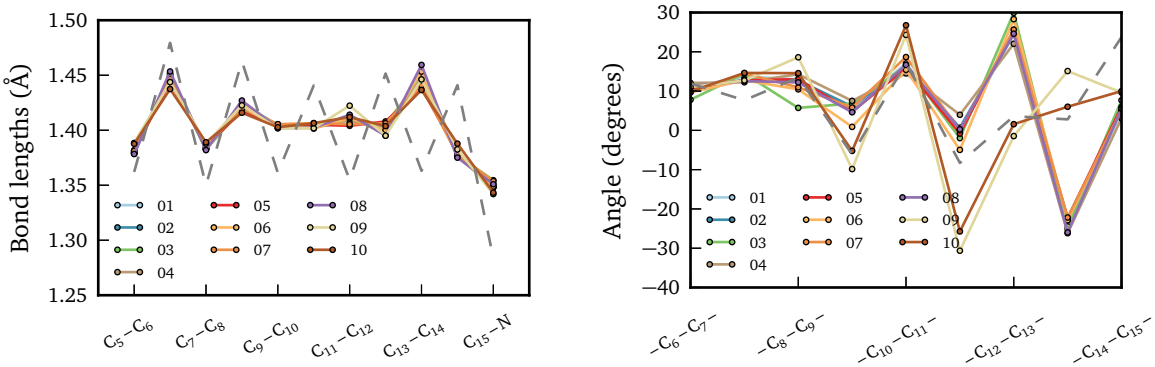


Figure 5.19: Geometries obtained after excited state optimizations in M4. Dashed gray lines indicate ground state average geometry. Left: bond lengths; Right: deviations from planarity of dihedral angles.

Inspection of the deviations from planarity of the bonds show that two samples stand out. Indeed, most structures show strong deviations around the $C_{12}-C_{13}$ and $C_{13}-C_{14}$ bonds, whereas not much change occurs in this region for samples 09 and 10. These two have strong distorsions around the $C_{11}-C_{12}$ bond instead. This change of behaviour stresses once again the importance of the initial sampling. Indeed, the HBN conformation around the retinal seems to strongly influence the excited state potential energy

surface, by favoring some distorsions over others. The pre-twist observed in both cases leads to different possible photochemical pathways. Further investigation on this point is presented below (see section 5.5.2).

M10 After optimization of the $H \rightarrow L$ root in the M10 protein, the geometries obtained feature inversion of bond lengths of the conjugated chain of the retinal molecule, much like what was obtained for 10Me-PSB in the precedent chapter which was labeled as ABL. The bond length obtained, as well as the deviation from planarity of the dihedrals are presented on figure 5.20. Contrary to what happens in M4, all the geometries obtained are highly similar when it comes to bond lengths. The highest elongation is obtained for the central $C_{11}-C_{12}$ bond, which goes from an average of 1.357 Å in the ground state to 1.466 Å in the excited state.

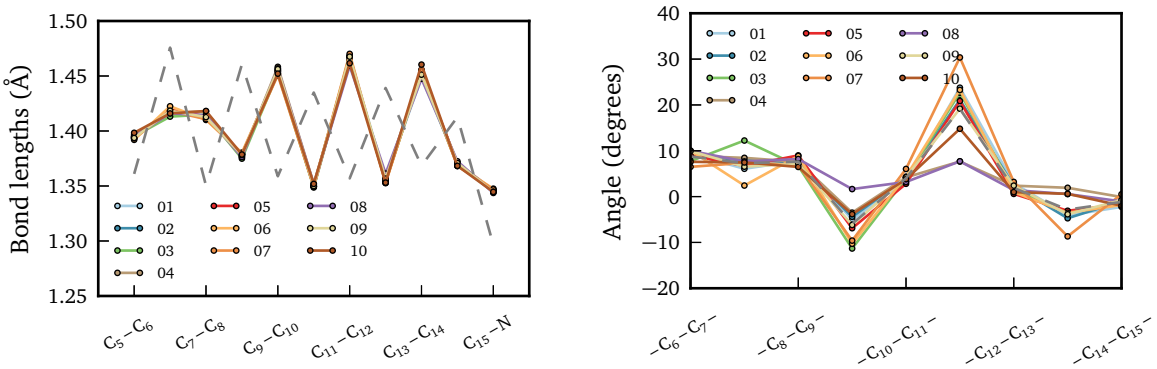


Figure 5.20: Geometries obtained after excited state optimizations in M10. Dashed gray lines indicate ground state average geometry. Left: bond lengths; Right: deviations from planarity of dihedral angles.

Concerning the dihedral angles, although some variety is observed, all the samples share the same global parameters. Overall, they are fairly close to the averaged ground state structure. However, the twisting around the $C_{11}-C_{12}$ bond is tuned along the trajectory, some sample (like sample 07) featuring a very strong distortion (up to 30°), while some others are nearly planar all along the chain (sample 08 and sample 04 for instance). This variety contrasts with our previous assumption that in M10, the protein scaffold shields the retinal from the external environment. However, some effect of the HBN is clearly noticeable in the excited state potential energy surface.

M8 What happens in M8 is less straightforward to interpret. Indeed, the separation in two sets highlighted before still hold in this case. The non-matching set, and sample 06, are give EBL-like optimized geometries, while the samples belonging to the matching set (except sample 06) are optimized to ABL-like geometry. It is worth to notice, however, that sample 06 lies in between, with a BLA pattern reminding the one labeled ABL* in previous chapter. Contrary to what happens in M10, though, the most elongated bond in the ABL optimized samples is the $C_{13}-C_{14}$ bond.

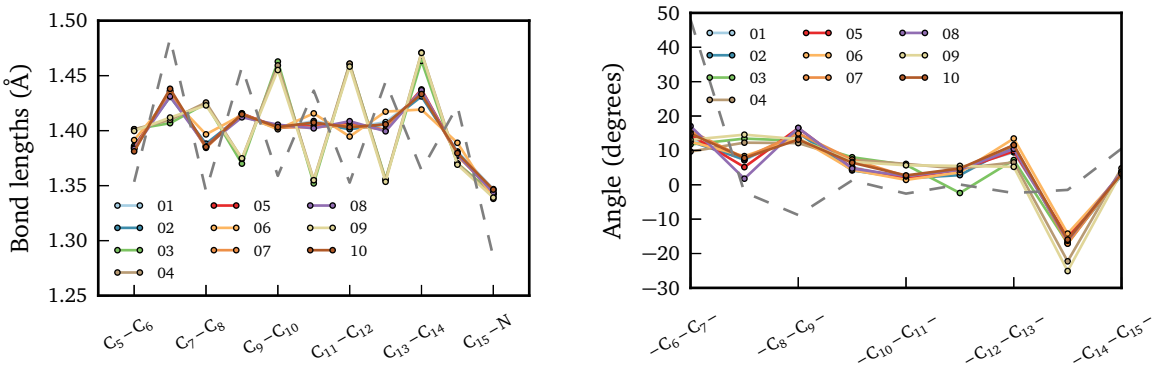


Figure 5.21: Geometries obtained after excited state optimizations in M8. Dashed gray lines indicate ground state average geometry. Left: bond lengths; Right: deviations from planarity of dihedral angles.

All samples show a twisting around the $C_{13}-C_{14}$ bond, and this does not depend on the BLA pattern, although the ABL samples have a stronger twisting. Another feature of the excited state geometries is that the ionone ring gets planarized, which makes sense given the double bond character acquired by C_6-C_7 bond upon excitation. Its bond length goes from 1.484 Å in the ground state to 1.425 Å in the excited state.

Discussion Two different BLA patterns have been found after excited state optimization. We have shown that the initial sampling have significant effects on the pattern obtained. Therefore, the excited state potential energy surface appears to be more sensitive to the HBN than the one of the ground state. Furthermore, these results have been obtained at the CASSCF level, and in our description the dynamical correlation is missed. We can wonder how this could affect the excited energy surface, and if in turn this effect could result different geometries. Moreover, the EBL geometry can be thought of as an intermediate geometry between the one in the ground state and the fully alternated one. Photochemistry

involving rotation around a formal double bond in the ground state (as is observed in bacteriorhodopsin and rhodopsin) requires the inversion of bond length, *i.e.* an ABL-like geometry. Therefore, reaching an ABL geometry has to be the key step for the photochemistry of these rhodopsin mimics. Investigating on the possibility of interchange between EBL and ABL will thus be the next step in this study.

5.4.2 Minimum energy paths connecting EBL and ABL geometries

Using the same methodology as in the case of solvated retinals, for each sample of each protein, starting from the excited state geometry, the other geometry has been obtained by linear interpolation, with ten points in between the two extreme geometries. The dihedral angles were not modified, and no geometry re-optimization was carried out. The energies of all points were computed at the CASPT2/6-31G* and MS-CASPT2/6-31G* levels. We chose the bond lengths of sample 03 of M8, which offers the best match with experimental vertical absorption, as the model for ABL geometry, which shows an elongation at $C_{13}-C_{14}$ elongation.

EBL-optimized samples In samples that are optimized into EBL geometries, the situation is highly similar to what has been found in solvated PSB in previous chapter. A representative map is represented on figure 5.22. This pictures what happens on most samples, although the position of the barrier is shifted for some samples. At the CASSCF level, the state dominated by the $H \rightarrow L$ is always S_2. This is also the state which has the highest oscillator strength all along the scan. From EBL to ABL, the $H \rightarrow L$ component of S_2 decreases steadily, and an increasing component $H \rightarrow L+1$ appears. For S_0, the closed shell GS configuration becomes more and more mixed with $H \rightarrow L$ when approaching ABL. As for S_1, starting from a major $(H \rightarrow L)^2$ component, a mix with $H \rightarrow L$ and GS appears along the scan.

As expected, the most stable geometry at the CASSCF level is the EBL geometry. However, when considering the CASPT2 energies, an avoided crossing between S_1 and S_2 appears, resulting in an energy barrier in the EBL to ABL transition. On average, the barrier is computed at even lower values than of solvated PSB (0.13 eV), at 1.65 kcal/mol (0.072 eV) on average. Given this very small value, and given that no transition state optimizations were carried out, it is likely that the two geometries are in "equilibrium", in the sense that the excited state wavepacket is expected to explore both regions of the potential energy surface, at least in a time scale shorter than the excited state lifetime. Two geometries in such "equilibrium" also fits nicely with the two signals observed in the pump-probe spectroscopy of M4.

At the MS-CASPT2 level, the picture is quite different, with a mixing between S_1 and S_2 disappearing. The potential energy surface for the S_1 $H \rightarrow L$ state is very flat along the scan, and the minimum is found around the middle of the scan. The EBL point is the least stable geometry.

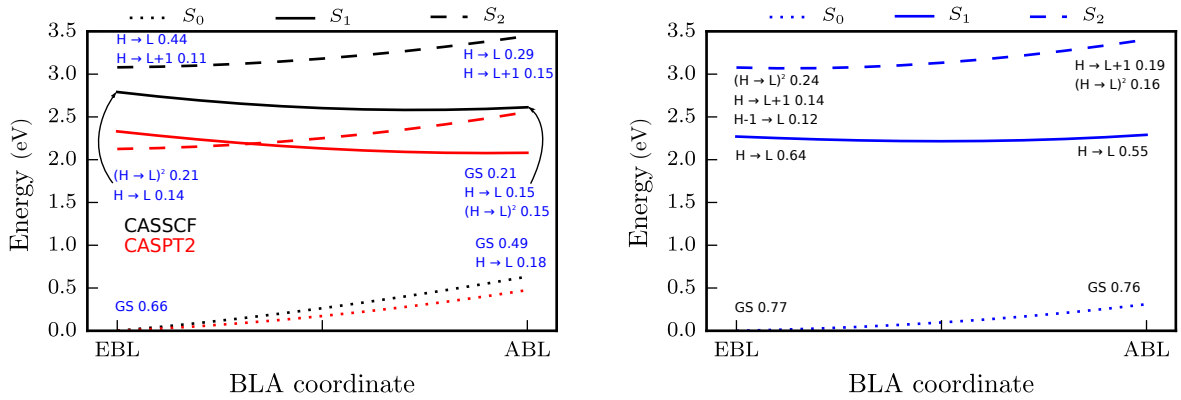


Figure 5.22: EBL optimized structure. Minimum energy path from EBL geometry to ABL geometry, with S_0 (dotted), S_1 (plain) and S_2 (dashed) states. Left: CASSCF (in black) and SS-CASPT2 (in red) energies, with the components of the CASSCF wavefunction at EBL and ABL points. Right: MS-CASPT2 energy, with the components of the PM-CASSCF wavefunction.

ABL-optimized samples This case is very similar to 10Me-PSB, and no mixing is observed along the EBL to ABL transition. At both CASSCF and CASPT2 levels, the $H \rightarrow L$ is always located as the S_1 state. When approaching the ABL geometry, the wavefunction displays higher *GS* component. A representative map, coming from one of M10 samples, is represented on figure 5.23. PM-CASSCF wavefunctions are very similar to the CASSCF ones along the MEP, but MS-CASPT2 energies appear blue-shifted with respect to CASPT2, and the qualitative picture is the same at both levels of theory. Overall, the S_1 potential energy surface is pretty flat, here again making likely the coexistence of the two minima, and no barrier is observed for the transition.

5.4.3 Relative stabilities and vertical emissions

The description of the two limit geometries has been refined by using the larger ANO-S basis set. The relative energies of EBL and ABL, as well as the vertical emissions obtained in the M4 protein, are

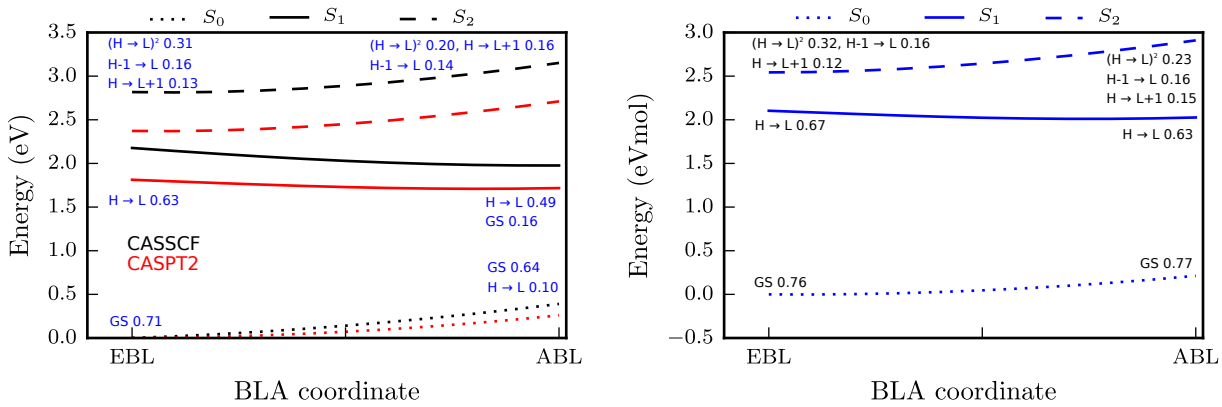


Figure 5.23: ABL optimized structure. Minimum energy path from EBL geometry to ABL geometry, with S_0 (dotted), S_1 (plain) and S_2 (dashed) states. Left: CASSCF (in black) and SS-CASPT2 (in red) energies, with the components of the CASSCF wavefunction at EBL and ABL points. Right: MS-CASPT2 energy, with the components of the PM-CASSCF wavefunction.

presented on table 5.13, where ΔE is computed as:

$$\Delta E = E(\text{EBL}) - E(\text{ABL}). \tag{5.4}$$

At the CASPT2 level, on average, the EBL geometry appears more stable than the ABL, contrary to what happens at the MS-CASPT2 level. However, these values are well below the error margin tolerated for our computational setup (around 0.2 eV or 4.6 kcal/mol). Here again, the influence of the HBN is clearly noticeable, with relative stabilities ranging from −1.18 kcal/mol to 2.70 kcal/mol at CASPT2 level, and from −0.23 kcal/mol to −2.99 kcal/mol at the MS-CASPT2 level. At both levels, it is thus perfectly possible that the two minima are in fact coexisting prior excited state decay to the ground state. Concerning the emission energies, the one selected for the CASPT2 level are the one coming from the lowest excited state, *i.e.* from S_1 state, although this is not the one with the highest oscillator strength, for the reason already stated in the previous chapter for solvated PSB: the system is considered in the lowest excited state. On average, the emission energy is found around 1.58 eV and 2.07 eV for ABL and EBL respectively. At the MS-CASPT2 level, the two signals are closer in energy, with an average of 2.03 eV and 2.26 eV for ABL and EBL respectively.

	CASPT2			MS-CASPT2		
	ΔE (kcal/mol)	ABL (eV)	EBL (eV)	ΔE (kcal/mol)	ABL (eV)	EBL (eV)
Sample 01	-0.05	1.58	2.03	-1.41	2.04	2.25
Sample 02	-0.40	1.62	2.07	-1.29	2.03	2.26
Sample 03	0.78	1.62	2.10	-2.64	2.01	2.21
Sample 04	0.71	1.57	2.06	-1.91	2.03	2.24
Sample 05	1.98	1.56	2.12	-1.68	2.06	2.31
Sample 06	-1.02	1.64	2.06	-0.93	2.05	2.29
Sample 07	1.96	1.58	2.13	-1.66	2.03	2.29
Sample 08	-1.18	1.67	2.08	-0.23	2.05	2.33
Sample 09	2.22	1.48	2.02	-2.78	1.99	2.16
Sample 10	2.70	1.51	2.08	-2.99	2.05	2.24
Average	0.77	1.58	2.07	-1.75	2.03	2.26

Table 5.12: Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M4 protein.

In M10, the qualitative picture is the same at both CASPT2 and MS-CASPT2 levels, with the ABL geometry stabilized by less than 1 kcal/mol. Here again, the relative stabilities computed for all the samples is well below the accuracy of the computational method. As expected, the influence of the HBN is less noticeable than in M4, with values separated by a maximum difference of around 2 kcal/mol at both CASPT2 and MS-CASPT2 levels. The vertical emissions computed are blue-shifted at the MS-CASPT2 level by around 0.3 eV, with values for ABL and EBL at 1.86 eV and 2.10 eV respectively, compared to the CASPT2 values at 1.50 eV and 1.81 eV. It is noticeable, however, that the difference between the two signals is comparable at both level (0.31 eV at CASPT2 and 0.24 eV at MS-CASPT2 levels).

	CASPT2			MS-CASPT2		
	ΔE (kcal/mol)	ABL (eV)	EBL (eV)	ΔE (kcal/mol)	ABL (eV)	EBL (eV)
Sample 01	1.05	1.49	1.78	0.46	1.79	2.01
Sample 02	1.09	1.50	1.82	0.64	1.86	2.10
Sample 03	1.37	1.46	1.78	1.01	1.83	2.09
Sample 04	0.10	1.56	1.84	-0.41	1.89	2.10
Sample 05	0.93	1.51	1.83	0.48	1.88	2.12
Sample 06	1.13	1.49	1.81	0.66	1.87	2.12
Sample 07	2.14	1.37	1.70	1.50	1.80	2.05
Sample 08	-0.11	1.59	1.83	-0.82	1.86	2.02
Sample 09	0.74	1.52	1.81	0.19	1.82	2.04
Sample 10	1.49	1.55	1.92	1.34	2.00	2.29
Average	0.99	1.50	1.81	0.51	1.86	2.10

Table 5.13: Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M10 protein.

As highlighted before, M8 gave both EBL and ABL geometries when optimizing its excited state. Surprisingly, the relative stabilization of one minimum compared to the other does not depend on the geometry obtained. For most samples, at the CASPT2 level, ABL is stabilized by less than 3 kcal/mol. However, sample 06 has the reverse behaviour, with an EBL geometry stabilized by 1.52 kcal/mol. On average, the ABL geometry gets stabilized by 2.02 kcal/mol.

The picture is quite different when it comes to the computation of the EBL and ABL vertical emission signals. Indeed, the ABL-optimized samples, *i.e.* samples 03, 04 and 09, have EBL emission strongly red-shifted, with respect to other samples (around 1.79 eV and 2.06 eV respectively). A similar red-shift, although less intense, is observed in the vertical emission from ABL geometry (1.42 eV and 1.54 eV on average, respectively). These differences disappear at MS-CASPT2 level, which give more or less the same picture as in M10.

	CASPT2			MS-CASPT2		
	ΔE (kcal/mol)	ABL (eV)	EBL (eV)	ΔE (kcal/mol)	ABL (eV)	EBL (eV)
Sample 01	1.81	1.54	2.07	-0.58	2.07	2.36
Sample 02	1.91	1.59	2.05	-0.83	2.16	2.35
Sample 03	**0.58**	**1.45**	**1.76**	**0.37**	**1.98**	**2.23**
Sample 04	**2.11**	**1.39**	**1.79**	**1.08**	**2.09**	**2.35**
Sample 05	3.29	1.52	2.12	-3.14	2.10	2.27
Sample 06	-1.52	1.53	1.91	0.56	2.06	2.39
Sample 07	3.01	1.53	2.10	-2.62	2.07	2.24
Sample 08	2.98	1.53	2.09	-3.41	2.12	2.27
Sample 09	**2.09**	**1.41**	**1.82**	**1.07**	**2.08**	**2.34**
Sample 10	2.68	1.53	2.11	-1.56	2.07	2.32
Average	2.02	1.54	2.06	-1.65	2.09	2.32
Average	**1.59**	**1.42**	**1.79**	**-1.21**	**2.05**	**2.31**

Table 5.14: Relative energies of EBL and ABL geometries (in kcal/mol), and vertical emission from EBL and ABL geometries (in eV) for M8 protein. Bold values indicate samples that were optimized in ABL geometry.

5.4.4 Conclusion: excited state geometries and the case of M8

In this part, we have shown that two different geometries can be obtained as excited state minimum. The first shows even bond lengths, and is reminiscent of what was obtained in solvated PSB. The second has alternated bond lengths, inverted with respect to the ground state geometry, and looks like what was obtained in 10-Me PSB.

Computing linear interpolations between the two limit geometries, we found that in EBL-optimized samples, the S_1 and S_2 energy surfaces are mixed, resulting in an avoided crossing along the map, with a small barrier (around 1.65 kcal/mol) at the CASPT2 level. At the MS-CASPT2 level, the S_1 surface is rather flat, and the two limit geometries are very close in energy. The relative stabilities of EBL and ABL are sample-dependent. For ABL-optimized samples, no mixing is observed along the scan, and ABL is always stabilized with respect to EBL. The S_1 surface, however, is here again rather flat, and the two minima are close in energy. Overall, it is likely that in all cases, the excited state wavepackets explore both limit geometries.

Given the similarity in crystal structures between M8 and M10, we expected and theoretically observed a very similar behaviour. This expectation is strengthened by the similar pump-probe spectra and the

close vertical absorptions. Thus, we would expect that M8 sampled configurations giving ABL geometries are more representative than those which are optimized into EBL. The ABL-optimized samples are the same as the one giving a matching vertical absorption, and dipoles and Mulliken charges in line with those coming from M10. As a consequence, we consider that these sampled configurations are the most representative, although they are a minority in our (limited)) sampling. Our findings suggest that the geometry obtained in the excited state is completely characterized by ground state properties. We have also highlighted before the structural discrepancy between M8 and M10 ionone ring region geometries in the ground-state. Following these ideas, two other samplings were performed:

- One arising from the same classical dynamics as before, but shifting the starting point to 5 ns, and then taking snapshots every 10 ns;
- One starting from a modified M8 structure, after rotations of W58 and L29 to have them in the same orientation as in M10.

The same QM/MM optimization steps as those that were presented and analyzed above have been performed. However, these new samplings barely helped to improve the matching set. Relaxing side chain residues when performing QM/MM optimizations does not help either. Our thinking is that the M8 crystal structure may present some flaws, and more careful initial sampling may have to be performed. These new computations are still carried out at the moment of the redaction.

In the next section, the data presented so far are gathered and compared to experimental pump-probe spectra. For the reasons stated above, we will consider only the samples 03, 04 and 09 for M8.

5.5 Interpretation of experimental spectra

So far, for each protein, we have three different geometries, obtained either *via* ground- or excited-state optimizations, or *via* constrained scans of linearly interpolated structures. The energies of these points were computed at the CASPT2 and MS-CASPT2 levels, with ANO-S basis set. For all these points, we also investigated on the excited state absorptions (ESA). For this purpose, 10 states were included in the state-average for the computation of the wavefunction.

For getting more insight, all the different signals (vertical absorptions, vertical emissions, and excited state absorptions from the different geometries) were plotted in figures 5.24, 5.25 and 5.26 for M4, M8

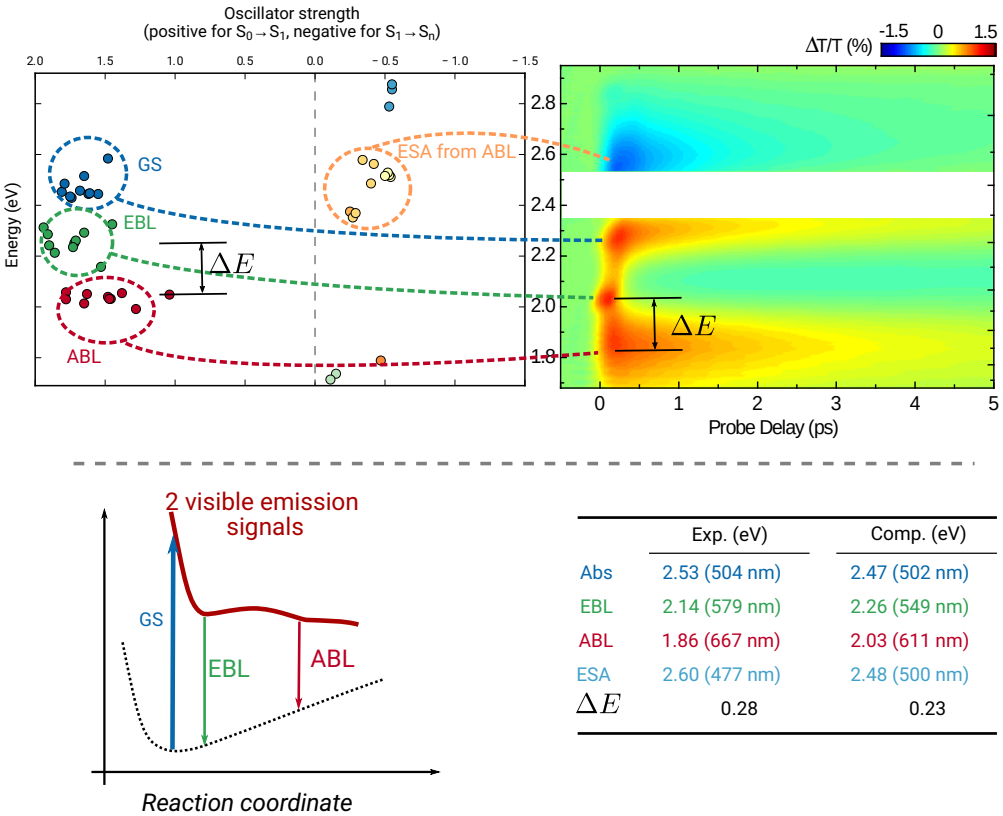
and M10 respectively. On these, we have chosen to represent the oscillator strengths of ESA as negative, to match the experimental representation.

5.5.1 Pump-probe spectra

M4 The experimental spectrum obtained for M4 shows three positive signals centered around 2.44, 2.14 and 1.86 eV, and one negative signal centered around 2.60 eV. The signals computed out of excited state absorptions at the MS-CASPT2 level agree well with the positive signals, with EBL and ABL signals computed around 2.26 eV and 2.03 eV respectively. Furthermore, the difference in energy between these emission bands is very well reproduced (0.23 eV in computations, and 0.28 eV in experiments). The results obtained at the CASPT2 level do not fit as nicely, as this was already highlighted in section 5.3.1. Indeed, the strong mixing between the first two excited states in all the geometries investigated indicate that the MS-CASPT2 results may be more reliable for computing the energies of this system.

Our computations agree nicely with the experimental map, and the sampling performed initially allows to reproduce the spread of each signal. The ABL signal does not fit as nicely as the others, but in this system ABL geometries were obtained by linear interpolation by arbitrarily choosing one of the M8 samples, and none were re-optimized. Optimizing the excited state is expected to lower the S_1/S_0 gap, and thus to red shift the signal, giving an even better agreement. The ESA is also nicely predicted as arising from ABL geometries, with a difference of only 0.12 eV between computed and experimental values. Here again, optimizing the ABL geometries is expected to lower the energy of S_1, and thus to increase the S_1/S_n gaps, so blue-shifting the ESA signal. At this point, the experimental map can be interpreted in terms of two geometries populated sequenially during the excited state dynamics. The EBL is reached first, and an equilibrium is rapidly established with the ABL geometry. Within around 300 fs, the EBL population is completely depleted in favor of the ABL one. The accurate mechanism for the interconversion, and the timescale on which it happens, cannot yet be fully explained by our static model. In the next section, we will try to get some insight on this photochemistry.

M8 The same methodology has been used for the M8 protein. The experimental spectrum has only two positive visible signals. The signal around the pump pulse (2.10 eV) corresponds to the ground state bleaching. The other one is centered around 1.84 eV. Our excited state computations at the CASPT2 level show an average vertical emission from EBL geometries at 1.79 eV, which correlates really well with the experimental peak. Given the good agreement, the peak at 1.84 eV can be assigned to a stimulated



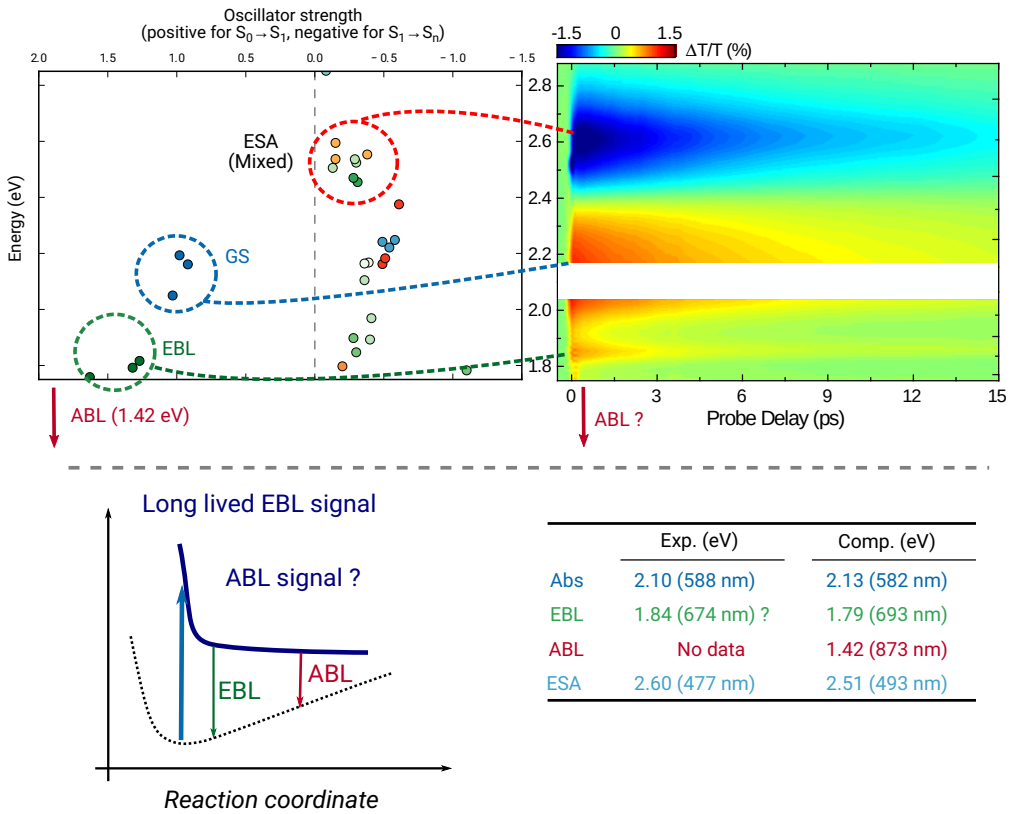
	Exp. (eV)	Comp. (eV)
Abs	2.53 (504 nm)	2.47 (502 nm)
EBL	2.14 (579 nm)	2.26 (549 nm)
ABL	1.86 (667 nm)	2.03 (611 nm)
ESA	2.60 (477 nm)	2.48 (500 nm)
ΔE	0.28	0.23

Figure 5.24: Top left: Pump-probe spectroscopy of M4, with the computed map displaying the excitation energy (in eV) with respect to oscillator strength. Positive oscillator strengths correspond to GSB and SE, negative ones correspond to ESA. Top right: correspondance with the experimental map. Bottom left: qualitative picture of the primary photo-event. Bottom right: comparison between the computed averages and the experimental maxima (in eV).

emission coming from an EBL population. However, we computed the average ABL signal at 1.42 eV, which is too far in the red for the experimental setup. The MS-CASPT2 computations do not correlate as well as the SS-CASPT2 ones, as was the case in ground state studies. Furthermore, no mixing is observed for any geometries in M8. For this reason, we consider that SS-CASPT2 as more accurate for the description of M8. The same mapping as in M4 has also been performed, and the results are presented on figure 5.25.

Although the sampling is far from perfect, with only three samples, the agreement between experiments and computations is fairly good. We especially predict an ESA signal around 2.51 eV, which correlates well with the experimental negative signal observed, peaking at 2.60 eV. This signal is found to be present for both ABL and EBL geometries. Experimental data indicate that the stimulated emission (SE) signal is

long lived, which can be interpreted with our computation as a long-lasting EBL population. Without any information about a possible ABL signal, it is hard to conclude on other features of the photochemistry occurring in the excited state.



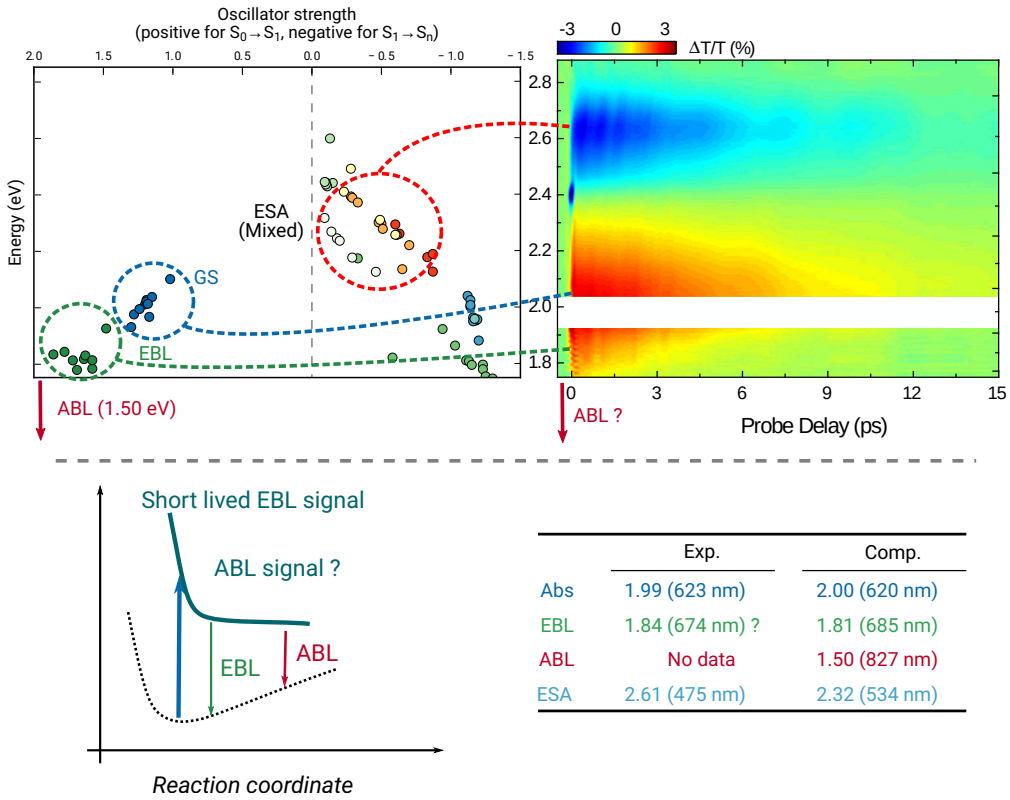
	Exp. (eV)	Comp. (eV)
Abs	2.10 (588 nm)	2.13 (582 nm)
EBL	1.84 (674 nm) ?	1.79 (693 nm)
ABL	No data	1.42 (873 nm)
ESA	2.60 (477 nm)	2.51 (493 nm)

Figure 5.25: Top left: Pump-probe spectroscopy of M8, with the computed map displaying the excitation energy (in eV) with respect to oscillator strength. Positive oscillator strengths correspond to GSB and SE, negative ones correspond to ESA. Top right: correspondence with the experimental map. Bottom left: qualitative picture of the primary photo-event. Bottom right: comparison between the computed averages and the experimental maxima (in eV).

M10 Although having a pump-probe spectrum similar to M8, some differences are observed in the case of M10, as stated previously. Most notably, the ground state bleaching is more widespread, and presents an asymmetry, with a tiny bump peaking around 1.84 eV. Our model shows that at SS-CASPT2 level, a signal from EBL population peaks at 1.81 eV, in very good agreement with the experimental bump. As was the case in M8, however, experimental setup only allows probing up to 1.77 eV, while the computations predict an ABL signal around 1.50 eV. For the same reason as in M8, MS-CASPT2 energies

are less relevant than SS-CASPT2 ones. The comparison between experimental spectrum and computed values is displayed on figure 5.26.

The proximity of ground state absorption and EBL emission is correctly predicted. The experimental spectrum shows intensity that are far weaker than those of M8. Our computations show that ESA signals from both ground state and EBL geometries that overlap with vertical absorptions and emissions respectively; thus, a weakening of the observed intensity is expected. Finally, like in M8, the ESA signal appears to be mixed between ABL and EBL geometries, and the maximum of all these signals is predicted around 2.32 eV, which is highly red-shifted from the experimental peak (2.61 eV). However, the higher number of states included in the state average, as explained before, may lead to inaccurate energy prediction, using a similar set of orbitals for all states computed.



	Exp.	Comp.
Abs	1.99 (623 nm)	2.00 (620 nm)
EBL	1.84 (674 nm) ?	1.81 (685 nm)
ABL	No data	1.50 (827 nm)
ESA	2.61 (475 nm)	2.32 (534 nm)

Figure 5.26: Top left: Pump-probe spectroscopy of M10, with the computed map displaying the excitation energy (in eV) with respect to oscillator strength. Positive oscillator strengths correspond to GSB and SE, negative ones correspond to ESA. Top right: correspondence with the experimental map. Bottom left: qualitative picture of the primary photo-event. Bottom right: comparison between the computed averages and the experimental maxima (in eV).

Summary In this section, we show that our model is able to interpret the experimental spectra of the three proteins by assigning the signals to some specific points in the potential energy surface of the embedded retinal molecule. The ground state bleaching, corresponding to the absorption maximum of the ground state species, was shown to be well reproduced at the beginning of this study (see section 5.3.1). Concerning the other positive signals, their presence has been interpreted as arising from two geometries in equilibrium, at least at the very beginning of the process. In M4, the two emitting population are observed, but have different lifetimes. The EBL population is observed at very early times, and depletes very fast, whereas the ABL one appears later, and is longer lived. In M8 and M10, the positive signals (or bump, in the case of M10) are assigned to EBL population. M8 seems to efficiently trap the population in the EBL geometry, given the long lived associated signal. In M10, the EBL population depletes faster. Furthermore, ESA signals are also interpreted as coming from excited state species, although the accuracy is lower for this case, due to the higher number of states that have to be included in the state average procedure.

At this point, the main signals in the spectra are better understood. However, the different behaviours from one system to another, especially concerning the lifetimes of the different populations, cannot be accounted for with our purely static point of view. An accurate model would require extensive semi-classical dynamics to be performed in each protein, to follow the evolution of the wavepacket on the excited state PES, and track it down to a conical intersection, which allows the system to decay either to the reactant, or to a hypothetical photo-product. To a lesser extent, information about the trajectory toward a possible CI can be obtained by computations of the MEP that connects the reactive ABL geometry to the CI geometry. This is the purpose of the next section.

5.5.2 Study of possible photoisomerizations by constrained scans

At the time of this study, no experimental evidence for a possible photo-product has been found yet. Therefore, multiple pathways have been tested. For each protein, we selected the sample that reproduces best the experimental vertical absorption (*i.e.* sample 03 for M4, sample 03 for M8 and sample 02 for M10). Starting from the ABL geometry, both $C_{13}-C_{14}$ (bacteriorhodopsin-like) and $C_{11}-C_{12}$ (rhodopsin-like) bonds have been rotated by 10° increments. Both clockwise (CW) and counter-clockwise (CCW) directions have been tested, apart when either $C_{13}-C_{14}$ (in M4 and M8, CCW) or $C_{11}-C_{12}$ (in M10, CCW) is pre-twisted (see section 5.4.1), resulting in 9 scans performed. Each geometry has been QM/MM optimized at the SA3-CASSCF/6-31G* level, and the energies have been corrected at the SS-CASPT2.

During the optimization, all protein residues were held fixed. This should not raise concerns for the rotations around $C_{13}-C_{14}$ as C_{14} is only linked to an hydrogen atom; however, C_{12} carries a methyl group, which give more steric interactions with the rest of the protein. The assumption is that the protein scaffold is rigid enough to avoid much structural change upon photoisomerization. In absence of relevant experimental data to support this assumption, the results presented here are far from complete and should be considered as a tentative modeling of the events following the accession to a reactive ABL geometry.

M4 Given the pre-twist in the CW direction for the $C_{13}-C_{14}$ bond, three scans have been performed for M4. We were unable to locate any crossing for scans around $C_{11}-C_{12}$, in each direction. However, a crossing was found for rotation around $C_{13}-C_{14}$. The corresponding graph can be found on figure 5.27.

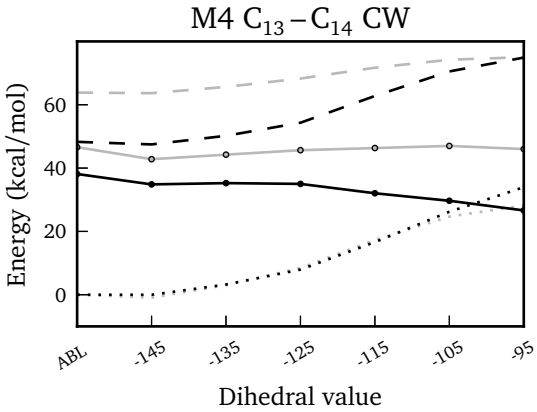


Figure 5.27: Constrained scan around $C_{13}-C_{14}$ in M4, CW direction (dotted: S_0; plain: S_1; dashed: S_2). Gray: CASSCF energies; Black: CASPT2.

From ABL geometry, the S_1 energy decreases steadily, and crosses the S_0 surface for a dihedral value around $-105°$. This map is strongly reminiscent of a peaked conical intersection topology, and indicates an efficient pathway for excited state decay to the ground state, consistent with the lower lifetime observed experimentally.

M8 Scanning around $C_{11}-C_{12}$ bond in M8, in both CW and CCW directions, gives S_1 energy surfaces that slightly goes up, leading to topologies reminiscent of sloped conical intersections. The computed energy barriers are around 2.80 and 7.24 kcal for CW and CCW directions respectively. The results are

presented on figure 5.28. The crossings are located for dihedral values of −107° and 102° for the CW and CCW directions respectively.

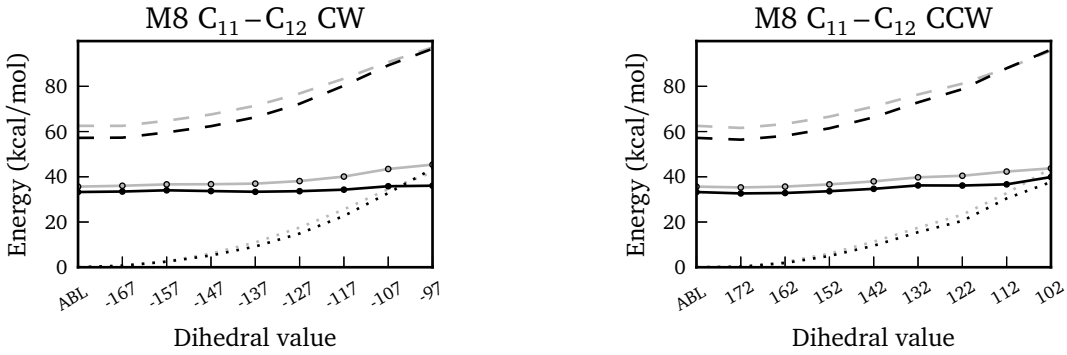


Figure 5.28: Constrained scans around $C_{11}-C_{12}$ and in M8 (dotted: S_0; plain: S_1; dashed: S_2). Left: CW; Right: CCW. Gray: CASSCF energies; Black: CASPT2.

On the contrary, the rotation around $C_{13}-C_{14}$, as can be seen on figure 5.29, leads to the same CI topology as in M4, with a crossing located for a dihedral value of −102°, with the difference that now there is a barrier of 2.45 kcal/mol along the S_1 energy surface. Therefore, reaching a CI is not favored in M8, with either a barrier, or sloped topologies. This finding is consistent with the long lifetime observed in M8.

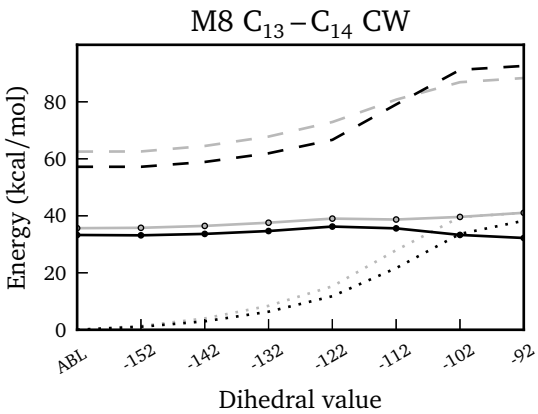


Figure 5.29: Constrained scan around $C_{13}-C_{14}$ in M8, CW direction (dotted: S_0; plain: S_1; dashed: S_2). Gray: CASSCF energies; Black: CASPT2.

M10 Contrary to what happens in M4 and M8, M10 is pre-twisted in the CCW direction around $C_{11}-C_{12}$. In this direction, the scan gives results highly similar to those obtained for rotations around $C_{13}-C_{14}$ in M4, with an S_1 energy that decreases steadily toward a peaked CI (see figure 5.30). The crossing has been located for a dihedral value of 115°. However, the surface is flatter than what is observed in M4.

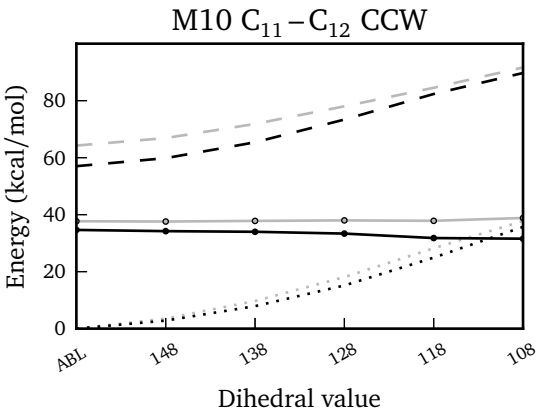


Figure 5.30: Constrained scan around $C_{11}-C_{12}$ in M10, CCW direction (dotted: S_0; plain: S_1; dashed: S_2). Gray: CASSCF energies; Black: CASPT2.

Rotating around $C_{11}-C_{12}$ gives high barriers, with S_1 energies that steadily goes up. Therefore, there is only one possibility for M10, which should be less favored than in M4. This is consistent with the longer lifetime of M10 compared to M4, and shorter lifetime compared to M8, where all possibilities are unlikely to occur.

5.5.3 First step towards the interpretation of the different excited state lifetimes

We have investigated on two possible photo-isomerizations in each protein: the first mimics what happens in vertebrate rhodopsins, with a rotation around $C_{11}-C_{12}$; the second relates to the rotation occurring around $C_{13}-C_{14}$ in bacteriorhodopsins. Out of the nine possibilities tested, CI were found in only five cases. In M4, the isomerization of $C_{13}-C_{14}$ bond was found leading to the topology of a peaked CI, with no barrier observed on the potential energy surface. This indicates that once an ABL geometry is reached, the system can easily reach the CI, thus leading to fast photochemistry (see figure 5.31). This also has the consequence of depleting the EBL geometry.

In M8, all the possibilities investigated on lead to possible CIs, although far less favorable than in M4. Around $C_{11}-C_{12}$ a sloped topology has been found, with significant energy barriers (around 3 and 7 kcal/mol). The rotation around $C_{13}-C_{14}$ lead to a peaked CI, but a 3 kcal/mol barrier has been found on the S_1 PES leading to the CI. Overall, all the possibilities should not allow a fast reaching of the CI, and thus explain the longer excited state lifetime in M8 than in M4.

Finally, M10 is found to be similar to M4. Only one direction, around $C_{11}-C_{12}$, leads to a CI, with a peaked topology, although with a surface that appears nearly flat. As a consequence, the decay to the ground state is expected to be less favorable than in M4, but faster than in M8.

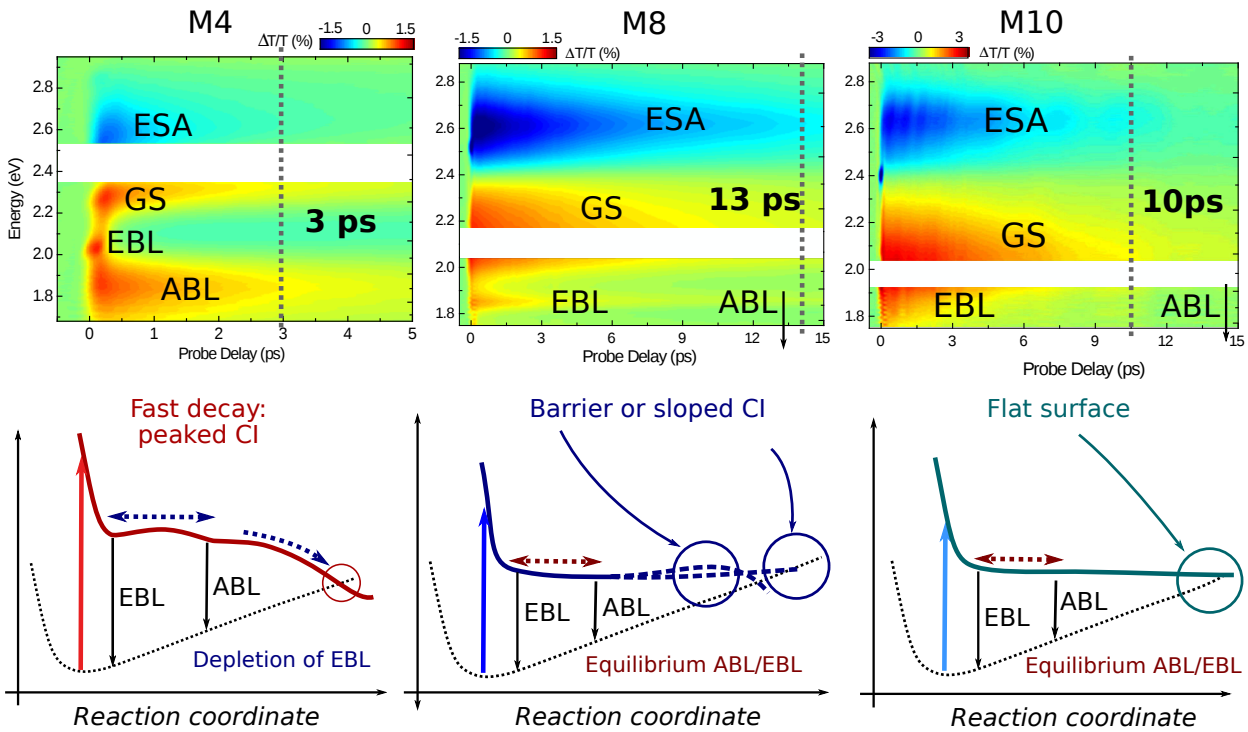


Figure 5.31: Tentative interpretation of the different lifetimes of each proteins.

Overall, our findings so far correlate well with the experimental lifetimes. The picture that arise is one where M4 decays back to the ground state *via* a rotation around $C_{13}-C_{14}$, while M10 does the same with a rotation around $C_{11}-C_{12}$. The case of M8 is more complicated, and may involve more than one rotation, with several energy barriers observed in the different pathways.

However, these results have been obtained by taking into account only one particular sample for each protein, and thus do not contain any influence of the sampling, although the sampling has been proven to have significant effects in the previous sections. Moreover, the selection of a sample of M8 as a model

for the ABL geometry in M4 can also bias to study given the elongation of $C_{13}-C_{14}$, which is more pronounced than for $C_{11}-C_{12}$, contrary to ABL geometries of M10. The sample selected for M4 was also strongly pre-twisted around $C_{13}-C_{14}$, whereas this is not the case in at least two other samples (see figure 5.19). The choices made can appear as somewhat arbitrary, but they arise from a relative ignorance when it comes to the photochemistry of these proteins. Indeed, as was highlighted at the beginning of this part, no experimental photo-product has been characterized yet. On the other side, our results lead to a qualitative understanding of the behaviour of the different systems. As such, this study paves the way for more thorough investigations, which may involve semi-classical dynamics to give a more accurate description of the experimental lifetimes.

5.6 Conclusion and perspectives

In this section we have presented a computational model that is able to interpret several properties of three bio-engineered rhodopsin mimics. This method relies on extensive solvent sampling, which has been shown to be crucial for the correct description of the vertical absorption tuning of these systems. The results presented here are of higher accuracy than those previously obtained by Cheng et al. [134] Following the correct description of ground state properties, where we have proven that the tuning arises from different shielding of the retinal by protein scaffold, we proved that relaxation from the FC region can lead to two quasi-degenerated geometries. One of them (ABL) features the classical bond length inversion with respect to the FC geometry, allowing for photo-isomerization reaction. The second (EBL) has even bond lengths in the polyenic chain, and appear as a transition structure between FC and ABL. Minimum energy paths interpolating between these two structures have shown that the potential energy surface is nearly flat in this region, proving that an equilibrium is likely to be reached. Combined with ESA computations, we were able to correlate the computed absorptions and emission to experimental pump-probe spectroscopy data, thus proving the validity of our approach. Finally, constrained scans of the minimum energy path for the photoisomerization coordinate showed that M4 is likely to easily reach a CI by rotating around $C_{13}-C_{14}$ bond (like in bacteriorhodopsin), and therefore to have a fast decay to the ground state. M10 also feature a downhill S_1 energy surface for the rotation around $C_{11}-C_{12}$ (like in rhodopsin), although less steep than the one in M4. On the contrary, in all the directions tested, M8 had either an uphill potential energy surface, or a barrier to overcome to reach a CI. These findings correlate well with the observed experimental lifetimes.

However, our current setup has large potential for improvement. In particular, the sampling of M8 was far less accurate than for the other systems, with high discrepancies from one sample to another. We assume that this result comes from discrepancies in the crystal structure. As pointed out by Cheng et al., the point mutations can have long-range effects, and just correcting the relative arrangements of W58 and L29 (see section 5.3.1) might not be sufficient, as highlighted before. Thus, allowing more relaxation in the protein side chains may help to correct the sampling. Following a setup similar to what has been recently presented by Melaccio et al. [140], in which side chains having atoms within 5 Å are allowed to move freely, we produced other trajectories for the three proteins, with similar parameters as those presented at the beginning of this chapter (see section 5.2.3). The first results are encouraging, with vertical absorptions in line with the experimental values for the three protein. Work is still ongoing in this direction, especially for the MEP computations between EBL and ABL geometries.

In this study we have presented photo-isomerization constrained scans for only one sample out of the ten available for each protein. Although the results obtained so far correlate well with the experimental data, we have no idea about the variability of the picture with respect to the change in the HBN around the retinal molecule. Finally, up to now our study also lack some dynamical characterization, which would be of paramount importance to better interpret the different signal lifetimes. In particular, we can hope to better understand the long lifetime of EBL in M8 compared to what happens in the other systems.

6 Evidence for a protonated Aspartate in the vincinity of the retinal in two microbial opsins

Résumé Nous nous intéressons dans ce chapitre à deux opsines bactériennes naturelles, qui servent de pompes à proton, activées par la lumière. La première est la bactériorhodopsine, où nous utilisons un modèle précédemment publié [33] et un modèle que nous avons créé à partir de la structure cristalline. La seconde est la xanthorhodopsine, où le rétinal est couplé à une molécule de carotène qui agit comme une antenne pour conduire l'excitation photo-induite au rétinal. Ces deux systèmes sont caractérisés par l'environnement quadrupolaire autour de l'iminium de la base de Schiff protonée, où deux aspartates et une arginine servent de contre-ion effectif. Nous avons montré que pour tous les modèles, nous surestimons l'énergie d'absorption d'au moins 1 eV. Cependant, les modèles dans lesquels un aspartate est protoné donnent un accord précis entre les calculs et les résultats expérimentaux. En nous concentrant sur la bactériorhodopsine, nous avons modélisé différentes conformations de l'aspartate protoné, et nous montrons que les structures obtenues reproduisent certaines données spectroscopiques connues — en termes d'absorption verticale, d'émission, et de temps de vie d'état excité.

Abstract In this chapter we focus on two natural opsins that can act as light-driven proton-pump. The first is the bacteriorhodopsin, where we use a model based on previously published results [33] and a model that we have built based on the crystal structure. The second is the xanthorhodopsin, where the retinal chromophore is coupled to a carotene that can act as antenna to drive the light excitation to the chromophore. These two systems are characterized by the quadrupolar environment around the iminium of the protonated Schiff base, where two aspartates and one arginine act as the effective counterion. We show that for all models, the computed vertical absorption is blue-shifted by around 1 eV with respect to experimental data. However, models where one of the aspartate is protonated give a very good agreement

with the experimental vertical absorption. Focusing on bacteriorhodopsin, we have modeled different conformation of the protonated aspartate, and we show that the resulting structures reproduce nicely the spectroscopic properties — in terms of vertical absorption, emission, and excited-state lifetimes.

6.1 Introduction: photo-induced proton-pumps

The bacteriorhodopsin (bR) of *Halobacterium salinarum* functions as a unidirectional light-driven proton-pump, and plays a crucial role in the synthesis of Adenosine triphosphate (ATP). Like other proteins from the opsin proteins family, this trans-membrane system is formed of seven α-helices, with an enclosed retinal molecule covalently bound to a nearby lysine amino-acid, forming a protonated Schiff base (PSB) [141]. The retinal is found in its all-*trans* conformation in the dark adapted state of the protein. Upon photon absorption, this retinal isomerizes around the $C_{13}-C_{14}$ bond, leading to its 13-*cis* conformer (see figure 6.1, panel b). This structural change induces steric and electrostatic perturbations that control the proton uptake on the cytoplasmic side and proton release on the extracellular side. This photocycle, completed on the millisecond timescale, involves six different different intermediates known as BR (the dark state), K, L, M, N and O, and involves several protonation states of selected amino-acids across the protein, namely Asp96 for the uptake of protons on the cytoplasmic side, and Glu204 and Glu194 for the proton release on the extracellular side [17], as presented on figure 6.1 (panel a).

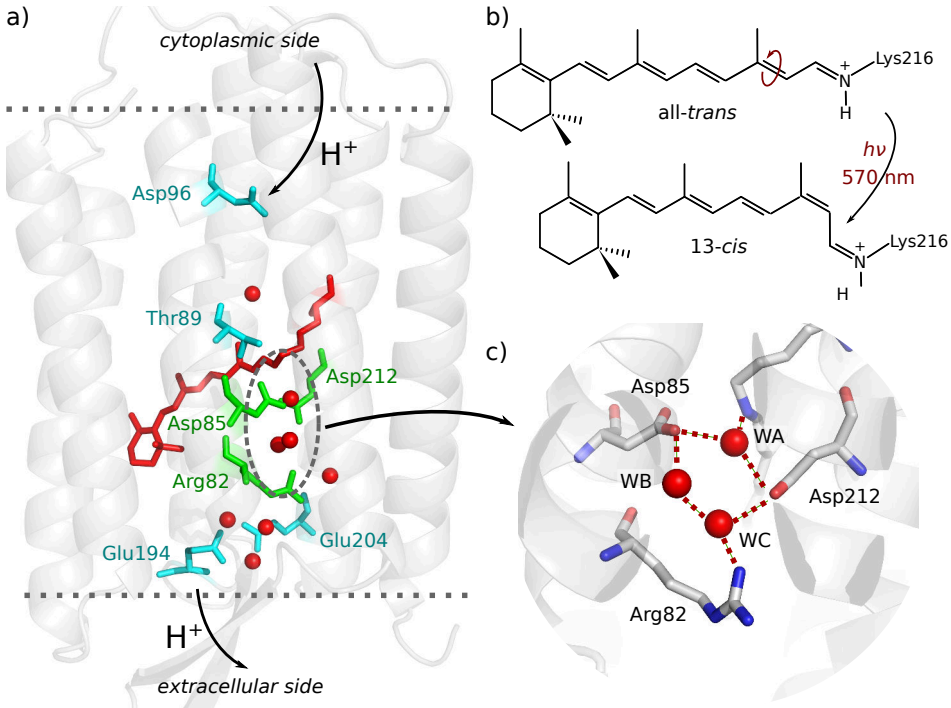


Figure 6.1: a) General structure of bacteriorhodopsin based on the crystal obtained by Luecke et al. PSB is highlighted in red, and amino-acids involved in the quadrupole are represented in green. Amino-acids that contribute to the pump activity are displayed in cyan, and relevant water molecules are displayed as red spheres. b) Photo-isomerization of the PSB. c) Quadrupole structure involving PSB, Asp85, Asp212 and Arg82. Oxygen atoms form a near planar pentagon.

As the bacteriorhodopsin is a prototype system for bacterial ion pumps, it has been extensively studied both experimentally and computationally, and the comprehension of its photocycle could lead to significant advances in nanotechnology, or for designing unique molecular devices. Several structures have been proposed for the local environment of the PSB, and at least four crystal structures are available in the protein data bank. The most recent one, obtained by Luecke et al. at a resolution of 1.55 Å [52], shows that the iminium of the PSB is involved in a quadrupole structure with two negatively charged aspartates and one positively charged arginine. Three water molecules are also found inside this quadrupole, forming a complex hydrogen bond network (HBN) linking the extracellular and cytoplasmic side (see figure 6.1). As the role of this protein is to convey protons through the hydrophobic environment formed by the membrane, this HBN is presumed to have an important role in the proton transport, and can undergo

significant changes during the photocycle [142–146]. The precise role of this HBN, and its structural changes upon excitation, is still widely debated.

Upon excitation of the retinal, its S_1 excited state is populated, and ultra-fast relaxation of the C–C bonds leads to the fluorescent I_{460} intermediate, named after its strong absorption band at 460 nm, and features an emission band centered around 760 nm [111, 112] extending far in the near IR [113]. This state has be shown to undergo photo-isomerization to 13-*cis* in around 500 fs, and decays to the ground state through a conical intersection (CI), reached in 200 fs [18]) to the vibrationnally hot ground state of the J intermediate. Vibrational cooling (in around 3 ps) leads to K intermediate. Accurate description of the evolution of the I_{460} intermediate has still to be achieved, although ultra-fast spectroscopy [113, 147, 148] and computational studies [30, 35, 125, 149] have lead to significant understandings. In particular, Altoè et al. [125] have shown, with minimum energy path computations (MEP) that the isomerization follows a space-saving double bicycle-pedal movement, with the $C_{13}-C_{14}$ bond rotating in the clockwise (CW) direction, while the adjacent $C_{11}-C_{12}$ and $C_{15}-N$ bonds rotate in the counter-clockwise (CCW) direction. Small barrier (under 1 kcal/mol) on the S_1 surface has also been proposed to explain the lifetime of the I intermediate.

Lanyi and Balashov [16, 150] have recently reported the discovery of Xanthorhodopsin (xR), which is a bacteriorhodopsin-like proton pump with enhanced efficiency triggered by the presence of a carotene molecule coupled to the retinal. This carotene acts as an antenna system to guide the photo-excitation to the retinal (see figure 6.2 panel a). Given the simplicity of this system compared to the photosynthetic reaction centers, that can involve dozens of coupled chromophores, this system could act as a archetypal setup for calibrating experimental and computational methods used for tracking excited-state energy migration. To the best of our knowledge, no computational model has yet been proposed for xR. However, a crystal structure has been published by Luecke et al. [150], and the antenna system has been investigated on by means of 2D UV-Vis spectroscopy [151] and circular dichroism [152].

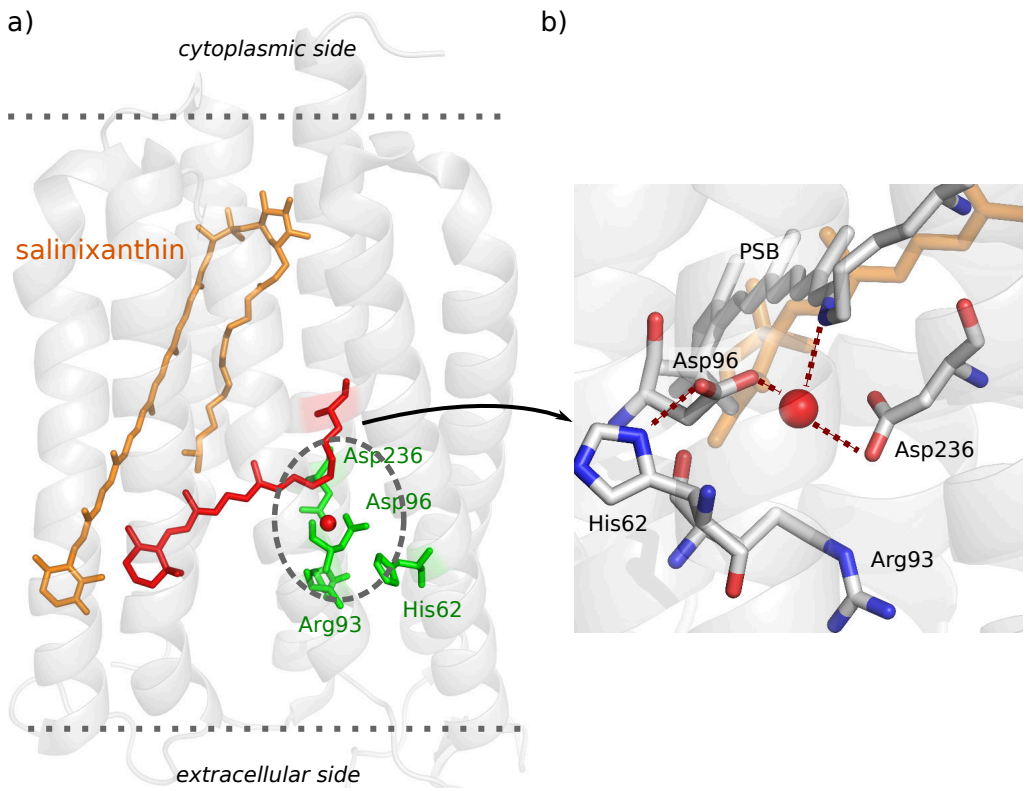


Figure 6.2: a) Xanthorhodopsin (xR) crystal structure from Luecke et al. [150] Retinal is displayed in red, and the amino-acids forming the quadrupole are highlighted in green. The carotene (salinixanthin) is displayed in orange. b) Structure of the quadrupole formed by the PSB, Arg93, Asp96 and Asp236.

In xR, the PSB is found in a quadrupolar environment very similar to the one observed in bacteriorhodopsin. This environment is displayed on figure 6.2 (panel b). However, only one water molecule is found inside, which could make the HBN more rigid. Another difference is the presence of a histidine residue, which has been proposed to make a zwitterionic like pair with the primary proton acceptor Asp96 (structurally corresponding to Asp85 in bR). Given the strong similarity of environments, obtaining a computational model for bR should lead to a model for xR with minimum modifications.

Our purpose in this chapter is to build a reliable model for both bR and xR in order to compute the excited state evolution of these two systems. In particular, we are interested in the topology of the S_1 surface in bR, and on the factors determining the high quantum yield observed for its photo-reaction. The model of xR will lead to further investigations on the coupling between the retinal chromophore

and the close lying carotene molecule, to interpret 2D UV-Vis spectroscopy data recently obtained in the group of Pr. S. Ruhman (unpublished data).

6.2 Computational details

6.2.1 Models considered

Our starting model was provided by Fujimoto et al., following their paper on computational study of color tuning in opsin proteins by SAC-CI methods [33]. This structure has been obtained by full QM/MM optimization of the whole protein obtained from 1C3W crystal from by Luecke et al. [52] Asp85 and WA were included in the QM region, and the optimization was carried out with B3LYP functional and Dunning's double-ζ plus polarization basis set D95(d). This structure will be referred to as **model1** in the following. Previously, similar refinements were carried out by Hayashi et al. [30, 153, 154] on an older crystal obtained by Luecke et al. [155] (1BRD, 1998) corrected with data for a missing loop from Grigorieff et al. [156] (2BRD, 1996), and for a water molecule in the quadrupole from Behrlali et al. [157] (1QHJ, 1999) and Luecke et al. (1C3W, 1999). In this work, the QM/MM segment was extended to the side chains of Asp85, Asp212 and Arg85, with the three water molecules WA, WB and WC (as labeled in figure 6.1), and was optimized at the HF/6-31G*. This structure will be labeled **model2**. These two refinements show similarities in the resulting structure of the quadrupole, and differ significantly from the 1C3W crystal structure when it comes to the participation of Asp212 in the HBN. Indeed, the QM/MM structures show that both oxygen of Asp212 can act as hydrogen acceptors, while in the crystal structure, Asp212 is slightly rotated, resulting in only one oxygen atom participating in the HBN. The O−H and N−H bond lengths are similar between the two QM/MM structures, but **model1** has significantly shorter $N_{ret}-O_{WA}$ length (2.54 Å and 2.63 Å for **model1** and **model2** respectively). As highlighted later in this document, an accurate description of the HBN in the quadrupole is of crucial importance for modeling the infra-red data, and thus a careful calibration has to be carried out.

For comparison purposes, we have rebuilt a model of bR from 1C3W, with the missing loop added with Scwlr4 software. All the residues were considered in their standard protonation, and Asp95, Asp115 and Glu204 are considered protonated [158, 159]. At first, the crystal has been taken as provided, and only MM relaxation has been carried out to correct for the possible discrepancies induced by the model ling of the loops and the addition of hydrogen atoms. This setup will be referred to as **bR1**.

A more realistic setup has also been produced. After neutralization with 3 Cl^- ions, the system has been relaxed at the MM level. Then, a short (1 ns) constrained classical dynamics has been performed at 300 K and 1 atm, where only the residues having at least an atom within 5 Å from the retinal were allowed to move freely. The retinal residue up to the C_ϵ of Lys216 was kept frozen during the dynamics. The last structure of the dynamics was minimized at the MM level, and the quadrupole (involving the retinal, Arg82, Asp85, Asp212, and the three water molecules WA, WB and WC as labeled in figure 6.1) has been further relaxed at the B3LYP/6-31G* level. This setup will be labeled as **bR2** in the following.

To the best of our knowledge, no computational model for xanthorhodopsin (xR) is currently available. Given the structural similarity between xR and bR, a successful modeling of bR should easily give good results on xR as well. For this reason, a very simple model of xR has been used in this study, where the crystal structure was taken as available in the Protein Data Bank from Luecke et al. [150] under the name 3DDL. In this protein, Asp96 (structurally equivalent to Asp85 in bR) is found to be closely associated to the nearby His62. Following the information provided by Luecke et al., this histidine is thought to be protonated toward Asp96 (see figure 6.2). Three models have been produced to test the validity of this assumption. **xR-His1** has His62 protonated toward Asp96, **xR-His2** has His62 protonated in the other way, and **xR-His+** has a charges His62. As in **bR1**, the resulting structures have only been relaxed at the MM level.

6.2.2 QM/MM setup

All the computations presented have been done within the QM/MM framework, and handled by the COBRAM interface [59]. Unless otherwise stated, the QM/MM frontier has been set at the C_ϵ of Lys216 for all computations involving the retinal, with the remaining atoms of the lysine side chains (up to C_β) put in the movable layer. For optimizations of the full quadrupole, QM/MM frontiers were set at the C_β of Arg82, Asp85 and Asp212. The MM part has been treated with the Amber ff99SB forcefield [108].

Retinal has been relaxed at the complete active space self-consistent field (CASSCF) [40] level of theory, with an active space comprising the full π electrons of the conjugated polyene chain, resulting in 12 electrons in 12 orbitals. Energies have been corrected by means on multi-reference second order perturbations, with either single state (SS) or multi-state (MS) CASPT2 [41]. Avoiding so-called "intruder states" has been done by using an imaginary shift of 0.2, and no IPEA has been set. Oscillator strengths have been evaluated using CASPT2 energies with the RASSI program as implemented in Molcas 8 [109]. The 6-31G* basis set has been used for all the computations.

Semi-classical dynamics have been performed with the Velocity Verlet algorithm as implemented in COBRAM, with a time-step of 1 fs. No surface hopping has been taken into account. These dynamics have been run with a movable layer that also includes the water molecule closest to the PSB, labeled WA in figure 6.1.

6.3 Refinements of the protonation model

In this section, we present the results obtained out of the QM/MM optimizations of the models presented above. We show that the strong blue shifts are obtained in all cases. However, neutralizing the charge of one of the aspartates in the quadrupole lead to better vertical absorptions. Therefore, a more in depth study of the possible protonation patterns, and the effect of the protonation on the HBN, is presented.

6.3.1 Standard protonation pattern

We start with an analysis of **model1**. After optimization of the ground state, a relatively planar structure is found, although with a clockwise twisted $C_{13}-C_{14}$ bond (the dihedral angle is found at $\approx 20°$), at MP2, B3LYP and CASSCF levels. This result is consistent with what has previously been reported [125]. As expected, CASSCF overestimates the BLA, while B3LYP and MP2 geometries are in good agreement in the C_5-C_{10} region. After C_{10}, B3LYP gives slightly less alternated bond lengths than MP2, which is the expected behaviour as proven by recent reviews [29]. The structural features of the retinal in **model1** are summarized on figure 6.3.

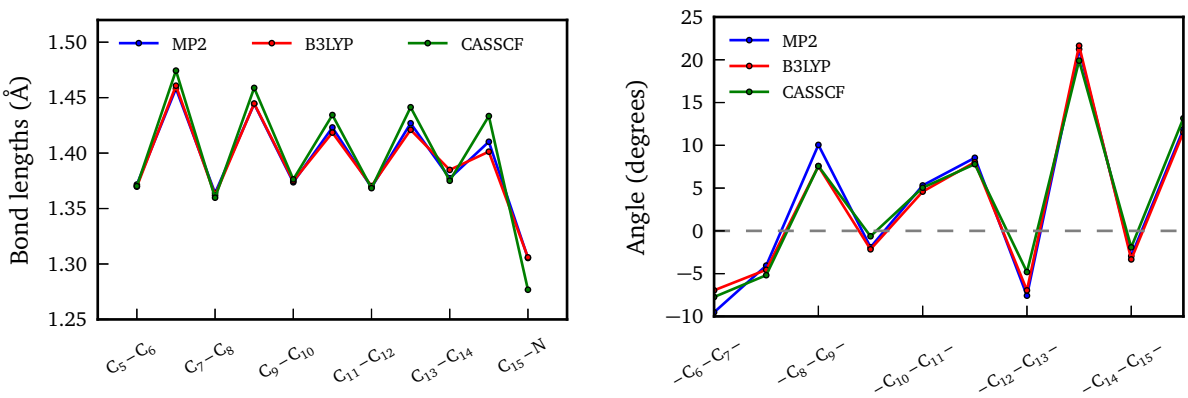


Figure 6.3: Structural parameters obtained for **model1** for different QM methods. Left: bond lengths; right: dihedrals angles along the chain.

Computation of the excited manifold on top of the optimized structures, with the CASPT2//CASSCF procedure described above, gives vertical absorptions with values of 2.67, 2.58 and 3.04 eV for MP2, B3LYP and CASSCF geometries respectively, largely blue shifted with respect to the experimental value (2.18 eV). At CASPT2 level, for all geometries, two closely lying excited states are found, although the computed oscillator strengths show that only S_1 is bright in the case of MP2 and B3LYP geometries. At the CASSCF geometry, S_2 appears as the bright state, but in this case, S_1 has a non-negligible oscillator strength, and leads to an absorption of 2.83 eV (see figure 6.4). MS-CASPT2 energies also show large blue shifts, although less pronounced than those obtained for SS-CASPT2. At this level, the energy differences are less sensitive to the geometry, as all the vertical absorptions are now in the same range, with values of 2.58, 2.62 and 2.57 eV for MP2, B3LYP and CASSCF geometries respectively.

CASSCF wavefunctions, associated with CASPT2 energies and oscillator strengths, are summarized in table 6.1. Analysis of the wavefunctions obtained again show that MP2 and B3LYP optimized structures have a lot in common. Both display an S_1 state dominated by the $H \rightarrow L$ excitation, with a little mixing with the closed shell (GS configuration), and an S_2 state dominated by the double $(H \rightarrow L)^2$ excitation, with contributions of $H \rightarrow L+1$ and $H-1 \rightarrow L$. This picture is somewhat inverted in the case of CASSCF optimized geometry, where S_2 appears clearly dominated by the single $H \rightarrow L$ excitation, and S_1 being a dominated by the double $(H \rightarrow L)^2$ excitation, with a little contribution of the single $H \rightarrow L$ excitation. This inversion in the character of the S_1 and S_2 states at the CASSCF geometry explains the fact that the $S_0 \rightarrow S_2$ has higher oscillator strength at this point, and the presence of a contribution of $H \rightarrow L$ in S_1 accounts for the non-negligible oscillator strength of the $S_0 \rightarrow S_1$ transition.

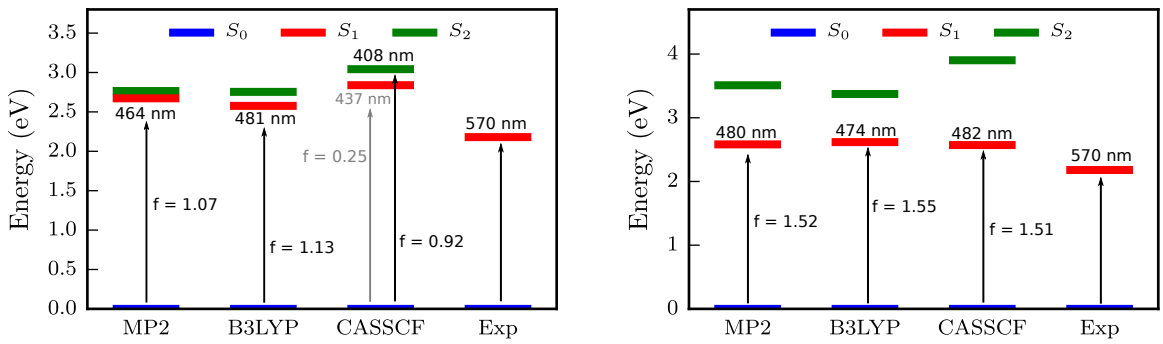


Figure 6.4: Excited state manifold on top of the optimized geometries of **model1**. Left: CASPT2; right: MS-CASPT2.

To investigate on possible influence of the refinement process used by Fujimoto et al. to produce **model1**, we produced our own model, **bR1** and **bR2** starting directly from the 1C3W crystal structure, with the method described above. The model **bR2** is the resulting structure of classical dynamics refinement, followed by a B3LYP optimization of the full quadrupole, including water molecules. A comparison between the **model1**, **bR1** and **bR2** models is represented on figure 6.5. The main difference between **model1** and the others is the position of Asp212, which is directly involved in the HBN in the quadrupole. This arrangement is not present in the crystal, nor is it found with our classical dynamics refinement. No differences in Asp85 and Arg82 is observed between the three model. However, **model1** and **bR2** have a similar relative arrangement of the water molecules WB and WC. These two models indeed have WC forming two H bonds (one with Asp212 and one with WC) and WB only forming one H bond with Asp85. The situation is reversed in **bR1**, where WB forms two H bonds, one with Asp85 and one with WC, and WC forms only one H bond with Asp212. However, the total number of H bonds is the same in all models, although their strengths are somewhat different, particularly concerning WA. The distances between hydrogens of WA and Asp212 and Asp85 (1.77 and 1.56 Å respectively) is way shorter than those obtained in **bR1** and **bR2** (2.12 and 2.10 Å for Asp212 and 1.66 Å for Asp85), indicating stronger H bonds in **model1**. The NH – C_{WA} bond is also shorter in **model1**, indicating a stronger coupling between the PSB and WA, which could influence the vertical absorption in this model.

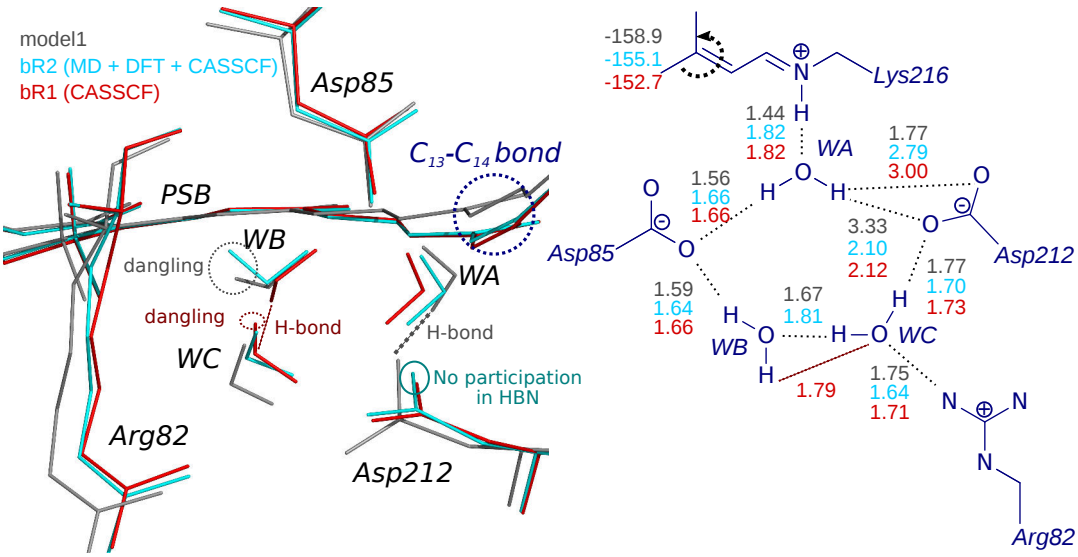


Figure 6.5: Overlay of the three bR models considered. Gray: **model1**; cyan: **bR2**; red: **bR1**. Left: 3D depiction of the quadrupole; right: schematic representation, indicating the lengths of H bonds and dihedral value around the $C_{13}-C_{14}$ bond of PSB.

However, these structural differences do not translate when computing the vertical absorptions on top of each CASSCF optimized geometries, as indicated in table 6.1. Indeed, **bR1** and **bR2** both have a bright S_2 state ($f = 0.72$ and $f = 0.78$ respectively), even more blue shifted with respect to experimental value (3.14 and 3.11 eV respectively). This S_2 state is dominated by the single $H \rightarrow L$ excitation. However, the $S_0 \rightarrow S_1$ transition, as in the case of **model1**, shows a non-negligible oscillator strength (0.47 and 0.41), higher than in **model1**. The S_1 state appears as a mix between the single $H \rightarrow L$ and $(H \rightarrow L)^2$ excitations, and the excitation energies computed for this transition are red-shifted with respect to that computed in **model1**, with values of 2.84, 2.56 and 2.62 eV for **model1**, **bR1** and **bR2** respectively. The computed oscillator strengths for the $S_0 \rightarrow S_1$ in **bR1** and **bR2** are in good agreement, although slightly blue-shifted, with what has been previously obtained by Altoè et al. [125] with a similar setup. At MS-CASPT2 level, a bright $S_0 \rightarrow S_1$ transition is obtained for both **bR1** and **bR2**, with corresponding energies of 2.48 and 2.49 eV, red shifted by around 0.1 eV from what has been computed for **model1**.

Model	Optimization	S_0 WF	S_1 WF	S_2 WF	$\Delta E(S_0 \rightarrow S_1)$	$\Delta E(S_0 \rightarrow S_2)$	$f(S_0 \rightarrow S_1)$	$f(S_0 \rightarrow S_2)$
model1	MP2	GS 0.66	$H \rightarrow L$ 0.55 GS 0.13	$(H \rightarrow L)^2$ 0.30 $H \rightarrow L+1$ 0.16 $H-1 \rightarrow L$ 0.13	2.67	2.76	1.08	0.11
	B3LYP	GS 0.64 $H \rightarrow L$ 0.11	$H \rightarrow L$ 0.55 GS 0.16	$(H \rightarrow L)^2$ 0.30 $H \rightarrow L+1$ 0.15 $H-1 \rightarrow L$ 0.14	2.58	2.75	1.13	0.08
	CASSCF	GS 0.70	$(H \rightarrow L)^2$ 0.25 $H \rightarrow L+1$ 0.16 $H \rightarrow L$ 0.12 $H-1 \rightarrow L$ 0.11	$H \rightarrow L$ 0.50	2.84	3.04	0.25	0.92
bR1	CASSCF	GS 0.68	$H \rightarrow L$ 0.23 $(H \rightarrow L)^2$ 0.21 $H \rightarrow L+1$ 0.13	$H \rightarrow L$ 0.37	2.56	3.14	0.47	0.72
bR2	CASSCF/DFT	GS 0.69	$(H \rightarrow L)^2$ 0.22 $H \rightarrow L$ 0.20 $H \rightarrow L+1$ 0.14	$H \rightarrow L$ 0.41	2.62	3.11	0.41	0.79

Table 6.1: CASSCF wavefunctions and CASPT2 energies and oscillator strengths obtained for the different models of bR considered. Energies are computed in eV.

The same results can be extended to the case of xR models. On these structures, given the strong similarities with the case of bR, we only performed CASSCF optimizations on the PSB on the three models considered. The results are presented in table 6.2. First, all three models have a strongly blue shifted absorptions compared to experiments (2.20 eV), with values of 3.12, 3.59 and 3.11 eV for **xR-His1**, **xR-His2** and **xR-His+** respectively, corresponding to a $S_0 \rightarrow S_2$ transition. Although **xR-His1** and

xR-His+ share similar characteristics, **xR-His2** stands out, with even more blue-shifted absorption and lower oscillator strengths, resulting from more mixed wavefunction. This result supports the fact that a careful protonation state has to be selected for His62. This amino-acid should be H bonded to Asp96. Moreover, **xR-His1** and **xR-His+** have more common features with the bR models investigated above, such as a non-negligible oscillator strength for $S_0 \rightarrow S_1$ transition, and an S_1 state having mixed character between $H \rightarrow L$ and $(H \rightarrow L)^2$ excitations.

Model	Optimization	S_0 WF	S_1 WF	S_2 WF	$\Delta E(S_0 \rightarrow S_1)$	$\Delta E(S_0 \rightarrow S_2)$	$f(S_0 \rightarrow S_1)$	$f(S_0 \rightarrow S_2)$
xR HisH1	CASSCF	GS 0.72	$(H \rightarrow L)^2$ 0.22 $H \rightarrow L+1$ 0.16 $H \rightarrow L$ 0.15 $H-1 \rightarrow L$ 0.10	$H \rightarrow L$ 0.47	2.93	3.12	0.19	0.77
xR HisH2	CASSCF	GS 0.67	$(H \rightarrow L)^2$ 0.18 $H \rightarrow L+1$ 0.17 $H \rightarrow L$ 0.13	$H \rightarrow L$ 0.25 $(H \rightarrow L$, L+1) 0.17	3.15	3.59	0.08	0.35
xR His+	CASSCF	GS 0.72	$(H \rightarrow L)^2$ 0.21 $H \rightarrow L$ 0.18 $H \rightarrow L+1$ 0.16	$H \rightarrow L$ 0.44	2.76	3.11	0.27	0.77

Table 6.2: CASSCF wavefunctions and CASPT2 energies and oscillator strengths obtained for the different models of xR considered. Energies are computed in eV.

Therefore, all setups investigated, either for bR or xR, and with different QM methods for geometry optimization, display highly blue shifted computed vertical absorptions. Correcting this error requires more than relaxing the quadrupole structure, and solving the problem requires to reconsider the point charges model we have used so far.

6.3.2 Analysis of the influence of particular amino-acids

Some insight on the electrostatic influence of each amino-acid of the quadrupole can be get by "switching off" successively each residue, by just setting the corresponding charge to zero in the electrostatic interaction term of the Hamiltonian. We have performed this analysis for **model1** of bR, switching off Asp85, Asp212 and Arg82, and **xR-His1** for xR switching off the corresponding Asp96, Asp236 and Arg93. The resulting CASSCF wavefunctions and CASPT2 energies are presented on table 6.3.

Model	Residue	S_0 WF	S_1 WF	S_2 WF	$\Delta E(S_0 \to S_1)$	$\Delta E(S_0 \to S_2)$	$f(S_0 \to S_1)$	$f(S_0 \to S_2)$
bR	Arg82	GS 0.72	$(H \to L)^2$ 0.23 $H \to L+1$ 0.16 $H \to L$ 0.15 $H-1 \to L$ 0.10	$H \to L$ 0.49	2.94	3.16	0.23	0.84
	Asp85	GS 0.74	$H \to L$ 0.54	$(H \to L)^2$ 0.23 $H-1 \to L$ 0.20 $H \to L$ 0.10	2.17	3.34	0.85	0.3
	Asp212	GS 0.77	$H \to L$ 0.56 $H \to L+1$ 0.10	$(H \to L)^2$ 0.27 $H \to L$ 0.13 $H-1 \to L$ 0.11	2.26	3.41	0.68	0.4
xR	Arg93	GS 0.66	$(H \to L)^2$ 0.18 $H \to L+1$ 0.17 $H \to L$ 0.13 $H-1 \to L$ 0.10	$H \to L$ 0.15 $(H \to L$, L+1) 0.13 $H \to L+2$ 0.10	3.24	3.87	0.07	0.2
	Asp96	GS 0.74	$H \to L$ 0.52 $H \to L+1$ 0.10	$(H \to L)^2$ 0.22 $H-1 \to L$ 0.20 $H \to L$ 0.12	2.11	3.26	0.76	0.35
	Asp236	GS 0.73	$H \to L$ 0.51	$(H \to L)^2$ 0.22 $H-1 \to L$ 0.19 $H \to L$ 0.12	2.01	3.26	0.73	0.33

Table 6.3: CASSCF wavefunctions, and CASPT2 energies, obtained by switching off selected amino-acids in the quadrupole.

Extinction of the Arg residue does not qualitatively change the results obtained in the previous sections. For both xR and bR, the bright transition is found as $S_0 \to S_2$. In bR, the computed vertical absorption are red-shifted by around 0.10 eV with respect to previous section, while the wavefunctions and oscillator strengths obtained are highly similar to what was obtained before. Switching off Arg93 in xR leads to results close to what has been obtained for **xR-His2**, and leads to even more mixed wavefunctions and blue-shifted energies.

However, aspartates have a much stronger influence on the compute properties. In bR, switching off Asp85 or Asp212 leads to bright $S_0 \to S_1$ transitions, with corresponding energies of 2.17 and 2.26 eV respectively, in very good agreement with experiment. The corresponding S_1 wavefunctions computed are also cleaner compared to previous computations, and are clearly dominated by the $H \to L$ excitation. The S_1 and S_2 states are also clearly separated in energy. The same conclusion holds for the extinction of Asp96 and Asp236 in xR, with computed vertical absorptions of 2.11 and 2.01 eV respectively. These

values are more red shifted with respect to the experimental value than those obtained for bR. Given the approximations made, this is still a very interesting path to follow.

These results lead us to assume than the standard model, with two deprotonated aspartates, may not be accurate. Indeed, removing the charges of the aspartates leads to significant improvements concerning the computation of vertical absorptions. These results hold when the aspartates are protonated, showing that a neutral residue is required in order to get energies comparable to experiments. Given the strong structural similarities observed between xR and bR, solving the problem for one protein should lead to conclusions that are straightforward to transfer to the other. As bR is one of the most studied opsin proteins, extensive experimental and theoretical results are available, for a wide number of properties, such as infra-red spectra [142, 145, 160], mutations studies [127, 161–164] or ultra-fast spectroscopy [113, 114]. With that in mind, we focus exclusively on bR in the following sections, and we hope that our conclusions will lead to structural models that are applicable to xR structure.

6.3.3 Analysis of protonated Asp212

Investigating more on possible protonations of aspartates in bR is not straightforward, as there are two possible residues to protonate (Asp85 or Asp212), and each of them can be protonated on two oxygen atoms. However, experimental evidence suggest that Asp85 acts as the primary proton acceptor during the photocycle, as the pumping activity breaks down if point mutations occur at this position [163], making it unlikely to be protonated in the dark state. Thus, we focused our efforts on Asp212. The problem of the protonation site is closely related to our approach, in which the protein is frozen in one possible conformation, while the aspartate can be expected to rotate around the $C_\alpha - C_\beta$ bond **in vivo**. To avoid extensive molecular dynamics sampling, and for testing purpose, we decided to simulate several arrangements and possibilities for the protonation of Asp212. First, the proton is added on one of the two terminal oxygen of the carboxylate group by tLeap, resulting in two different structures. Then, a few steps of NVE dynamics sampling are used to relax the systems, and two other arrangements are obtained, where the proton points toward the exterior of the carboxylic acid. For the four resulting structures, a B3LYP optimization of the full quadrupole is carried out (as was done in **bR2** above), and the retinal is further optimized at the CASSCF level. The resulting geometries of **O1-cry**, **O1-md**, **O2-cry** and **O2-md**, and the corresponding HBN structures are presented on figure 6.6. On this figure are also indicated the energies of the ground state of the retinal at the CASPT2 level, relatively to **O1-cry**. The QM/MM energy includes the MM contribution of the environment.

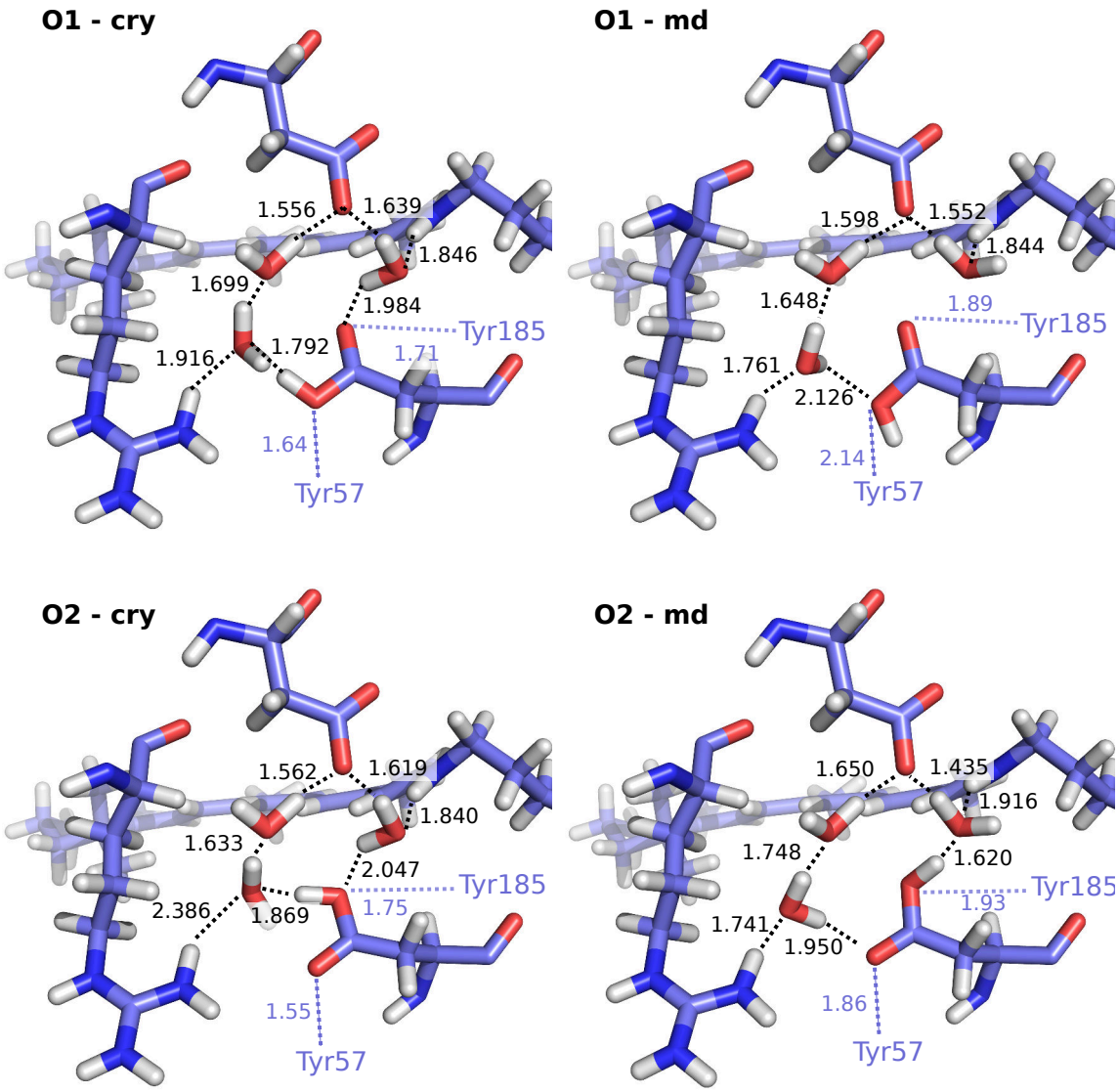


Figure 6.6: Protonations of **model1** obtained, with the structure of the HBN (bond lengths in Å). Energies indicated refer to the S_0 energy of the retinal, computed at the CASPT2 level; QM/MM energies include MM contributions from the environment.

First, it is noticeable that WA changes its conformation during the NVE dynamics, resulting in structures in which WA only forms an H bond with Asp85. This rotation of the water molecule is related to a shortening of the WA - Asp85 distance, from 1.639 Å to 1.552 Å in the case of O1 protonation, and from 1.619 Å to 1.435 Å in the case of O2 protonation, which may result from the neutralization of Asp212. This rotation also triggers the rotation of WC, which also gets closer to Arg82. As expected, the energy of the retinal is more affected by protonations on O2, which was originally directly H bonded

to WA. In particular, relaxing the system after O2 protonation leads to an overall stabilization of the full system by −3.49 kcal/mol, while the ground state of the retinal is destabilized by 9.61 kcal/mol. However, discussion of the obtained relative energy may not be relevant for this study, as the model proposed should be validated by molecular dynamics to investigate on the long term stability of this protonation pattern, and on the possible changes of conformation of Asp212. More interesting in the scope of this section is the computation of the excited manifold on top of each optimized structures, which is summarized on table 6.4.

Both **O1-cry** and **O1-md** share similar features, with a bright $S_0 \rightarrow S_1$ transition computed at 2.18 and 2.20 eV respectively, and an S_1 state dominated by the $H \rightarrow L$ excitation. However, some differences can be spot on the S_2 state, which does not have significant contributions from $(H \rightarrow L)^2$ in **O1-md**. **O2-cry** and **O2-md** have much differences. **O2-md** has a vertical absorption corresponding to a $S_0 \rightarrow S_1$ transition red shifted with respect to the one obtained in **O2-cry** by 0.11 eV. The wavefunctions are also cleaner in the case of **O2-md**, resulting in better separation between the oscillator strengths of both transitions.

Model	S_0 WF	S_1 WF	S_2 WF	$\Delta E(S_0 \rightarrow S_1)$	$\Delta E(S_0 \rightarrow S_2)$	$f(S_0 \rightarrow S_1)$	$f(S_0 \rightarrow S_2)$
O1 – cry	GS 0.69	$H \rightarrow L$ 0.43 GS 0.10	$H \rightarrow L$ 0.16 $H-1 \rightarrow L$ 0.15	2.18	3.20	0.83	0.33
O1 – MD	GS 0.69	$H \rightarrow L$ 0.43 $(H \rightarrow L)^2$ 0.10	$(H \rightarrow L)^2$ 0.21 $H \rightarrow L$ 0.16 $H-1 \rightarrow L$ 0.15	2.20	3.20	0.84	0.34
O2 – cry	GS 0.69	$H \rightarrow L$ 0.46	$(H \rightarrow L)^2$ 0.23 $H-1 \rightarrow L$ 0.16 $H \rightarrow L$ 0.13	2.14	3.20	0.88	0.28
O2 – MD	GS 0.72	$H \rightarrow L$ 0.55	$(H \rightarrow L)^2$ 0.28 $H-1 \rightarrow L$ 0.20	2.03	3.05	1.15	0.16

Table 6.4: CASSCF wavefunctions and CASPT2 energies computed for the protonated models of bR **model1**.

The geometrical parameters computed for the retinal in each model are represented on figure 6.7. The bond length alternation along the polyene chain of the retinal is slightly reduced compared to the original **model1**, with an even more pronounced reduction for **O2-md** in the $C_{10}-C_{15}$ region. The deformation along the polyene chain are also similar within the four setups, and show higher distortions around the C_8-C_9 and C_9-C_{10} bonds. Here again, **O2-md** behaves differently from the others in the $C_{12}-C_{15}$

region, with more distortions around the $C_{12}-C_{13}$ and $C_{14}-C_{15}$ bonds, and a reduced twisting around the $C_{13}-C_{14}$ bond. The three other structures are more twisted around this last bond, while they share similar characteristics with the reference **model1** around $C_{12}-C_{13}$ and $C_{14}-C_{15}$. Overall, **O2-md** shares a lot of similarities with **model1**, especially at the $C_{13}-C_{14}$ position.

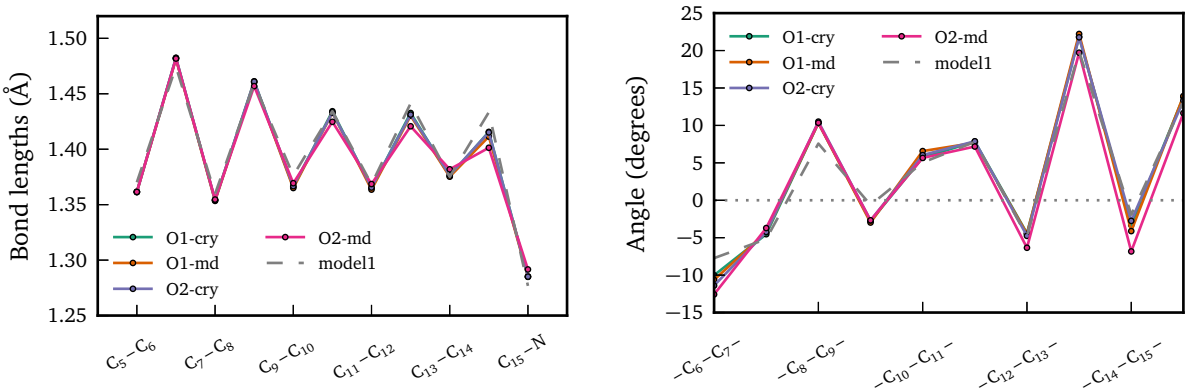


Figure 6.7: Geometrical parameters for the CASSCF optimized ground state of the protonated setups of **model1**. Left: bond lengths; right: dihedral angles.

The three different protonation pattern investigated share a lot of common features, from excited-state manifold structure to geometrical parameters. However, **O2-md** stands out, with a red-shifted vertical absorption, and reduced BLA along the polyene chain. The distortions in the iminium region are also different in this system, particularly with the lower pre-twisting around $C_{13}-C_{14}$. This last feature may induce differences in the photo-isomerization pathway.

In this section, we have shown that the standard protonation pattern of Asp85 and Asp212 leads to blue shifts in the computation of the vertical absorption of bR. However, neutralizing one of these residues give results that are in very good agreement with the experimental maximum absorption. In the next part, we will investigate on the influence of this new protonation pattern on the excited state of the retinal, and compare our results to experimental data.

6.4 Excited state characteristics

This section tackles the evolution of the excited state of the PSB out of the FC region. First, we will present the results obtained when relaxing the geometry on the S_1 energy surface in the four setups presented above. We also show the data obtained out of adiabatic semi-classical dynamics.

6.4.1 Excited state relaxation

On top of the CASSCF optimized structures of the four protonated setups, we performed an excited state relaxation of the bright $H \rightarrow L$. Although all the systems had similar behaviours in the ground state, two different types of geometries are obtained after excited state optimization. The structural parameters obtained are presented on figure 6.8. **O1-cry** and **O2-cry** both lead to fully distorted geometry around the $C_{13}-C_{14}$ bond in the clockwise (CW) direction, characteristics of a structure obtained at a conical intersection (CI). This distorsion is associated with corresponding counter-clockwise (CCW) rotations of the $C_{11}-C_{12}$ and $C_{14}-C_{15}$, as was previously pointed out by Altoè et al. On the other hand, **O1-md** and **O2-md** lead to alternated bond lengths (ABL) geometries, characterized by inverted bond lengths in the polyene chain with respect to the ground state geometry. Remarkably, **O1-md** features an important elongation of the $C_{13}-C_{14}$ bond (from 1.375 Å in the ground state to 1.472 Å in the excited state), while this elongation leads to much shorter bond in **O2-md** (from 1.382 Å in the ground state to 1.450 Å in the excited state). It is interesting to note that the length of $C_{13}-C_{14}$ bonds in **O1-md** is similar to the one obtained in the CI-like structures of **O1-cry** and **O2-cry**, indicating that this structure can be considered as an intermediate between the relaxed excited state structure (*i.e.* **O2-md**) and the CI). **O1-md** is also characterized by a stronger dis torsion around $C_{13}-C_{14}$, with the corresponding dihedral going from around 22° to 35°, while no increase in this position is observed for **O2-md**, which is consistent with the relaxed excited state structure that was obtained by Altoè et al.

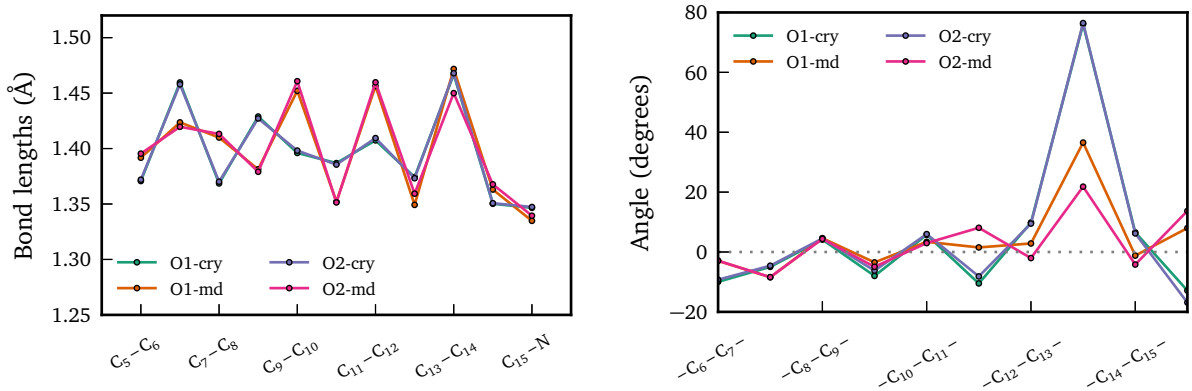


Figure 6.8: Geometrical parameters for the protonated setups of **model1** in the excited state minimum. Left: bond lengths; right: dihedral angles.

Computations of the excited state manifolds on top of the optimized geometries confirmed that the geometries obtained for **O1-cry** and **O2-cry** correspond to a point where S_0 and S_1 are degenerated. The vertical emission has been computed at 1.39 and 1.53 eV for **O1-md** and **O2-md** respectively. The value obtained for **O2-md** is close to the maximum of the fluorescence spectra obtained by Schenkl et al. at 1.63 eV (761 nm), while **O1-md** appears much red-shifted, but correlates well with the stimulated emission peak obtained by Ruhman et al. around 1.31 eV (950 nm). Therefore, these two limit structure could correspond to the same I intermediate, at different evolution state toward the isomerized J intermediate.

To investigate on the influence of the geometry, and in particular of the torsion around $C_{13}-C_{14}$, on the excited state manifold, we re-optimized the excited states of **O1-cry**, **O2-cry** and **O2-md** by constraining the dihedral angle around $C_{13}-C_{14}$ to the one obtained in **O1-md**, and **O1-cry**, **O2-cry** and **01-md** with this angle constrained to the value obtained in **O2-md**.

Constrained to O1-md The geometrical parameters obtained are presented on figure 6.9. This new optimization leads to geometries that are very similar for all the systems investigated, although **O2-md** still has some differences, notably a shorter $C_{13}-C_{14}$ bond and different torsions along the chain. The vertical emissions out of these new geometries are computed at 1.34, 1.33 and 1.47 eV for **O1-cry**, **O2-cry** and **O2-md** respectively. Thus, not much change is observed for **O2-md**, while the two others now have features similar to the one observed before in **O1-md**.

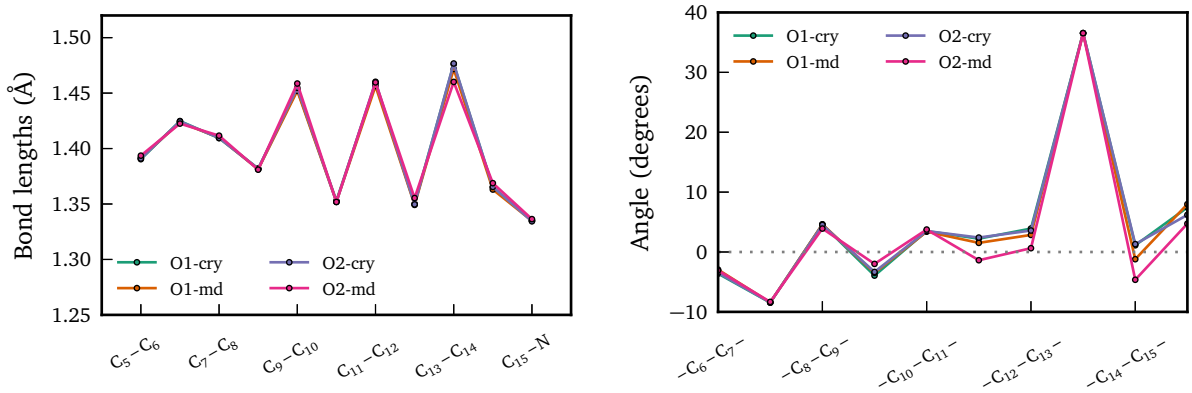


Figure 6.9: Geometrical parameters for the protonated setups of **model1** in the excited state minimum, with constrained dihedral angle around $C_{13}-C_{14}$ to the value obtained for **O1-md**. Left: bond lengths; right: dihedral angles.

Constrained to O2-md Here again, the bond lengths obtained in all cases are similar, with a noticeable difference **O2-md** at the $C_{13}-C_{14}$, which appears much shorter than for the three other setups (see figure 6.10). The dihedrals obtained also differ significantly between **O2-md** and the other structures, with a $C_{11}-C_{12}$ bond significantly less distorted in **O2-md**. The distorsion observed for **O1-md**, **O1-cry** and **O2-cry** are around 5° higher than what is found in their ground-state, which could indicate that oscillations of this particular angle are likely to appear during the excited-state relaxation. The vertical emissions found are very similar, contrary to what happens in the other constraining setup, with emissions found at 1.45, 1.48 and 1.44 eV for **O1-cry**, **O1-md** and **O2-cry** respectively, and are all consistent with the experimental fluorescence of bacteriorhodopsin around 1.63 eV, which further proves the fact that **O2-md** is a good candidate geometry for the fluorescent state.

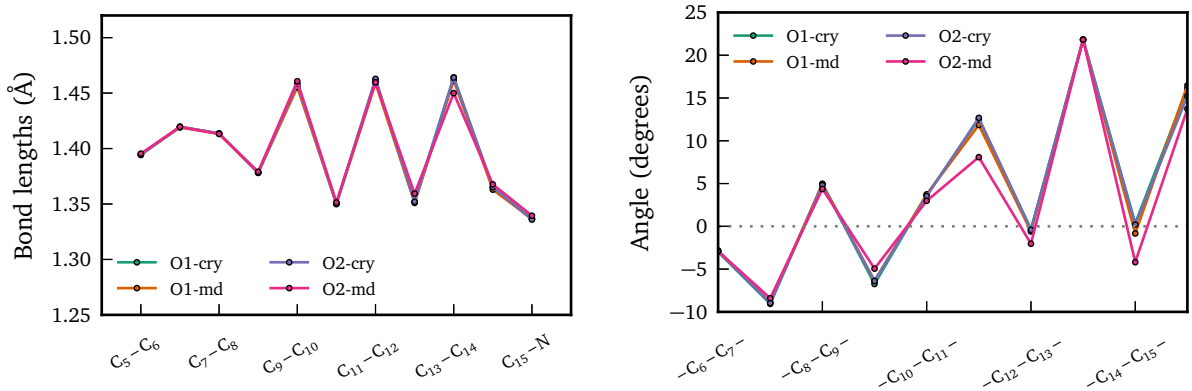


Figure 6.10: Geometrical parameters for the protonated setups of **model1** in the excited state minimum, with constrained dihedral angle around $C_{13}-C_{14}$ to the value obtained for **O2-md**. Left: bond lengths; right: dihedral angles.

In this section, the minimum energy path connecting the FC region and the excited state minimum has been computed for the three models protonated systems. Two of those models lead directly to the CI, in contradiction with knowledge on the photocycle of bR. Indeed, an intermediate has been characterized during the BR to K transition. However, the fluorescence spectra is broad, which can be attributed to the evolution of the fluorescent state. Indeed, we have proven that the geometries, with a constrained angle around $C_{13}-C_{14}$, reproduce the signal observed in fluorescence spectroscopy and the maximum of the stimulated emission spectra. Investigations on the mechanisms controlling the evolution of the systems from FC to the CI will be carried out in the next section.

6.4.2 Semi-classical trajectories

We have performed 0 K semi-classical dynamics on the four systems, with the computational methods described before. We have computed a 400 fs for each system. Out of each trajectory, the bond length alternation at each step has been computed by taking the difference between the average length of the double bonds (considered in the ground state) and the average length of the single bonds. The angles controlling the evolution toward the CI have also been monitored. As seen in the previous section, three angles are distorted in the transition to the CI:

- θ_1: $-C_{13}-C_{14}-$;
- θ_2: $-C_{11}-C_{12}-$;

- θ_3: $-C_{15}-N-$.

The evolution of the energies, the BLA, and the dihedrals are presented on figure 6.11 (for O1 protonation) and figure 6.12 (for O2 protonation). Both O1 protonated systems reach a CI in around 200 and 220 fs, for **O1-cry** and **O1-md** respectively. The surface energy computed for S_1 in **O1-cry** decreases slowly after a stabilization during the first 100 fs, leading to a CI that is around 2.25 eV with respect to the FC S_0 energy. The picture is different for **O1-md**, where the S_1 energy surface is flatter and leads to a CI found 2.60 eV higher than the ground state. This could be the reason why optimizing S_1 in **O1-cry** lead directly to a CI twisted structure, while the same optimization carried out on **O1-md** stabilizes an ABL intermediate. The photo-isomerization is lead by the clockwise rotation around $C_{13}-C_{14}$ bond (angle θ_1), which stabilizes around 80° for both **O1-cry** and **O1-md**, in very good agreement with the CI obtained by optimizing the S_1 state in **O1-cry**. This movement is accompanied by a CCW rotation of the θ_2 and θ_3 angles, with a little difference between the two protonation patterns. Indeed, both angles stabilize around −15° in **O1-cry** (here again, this agrees with what has been found in the optimizations in the previous section), while the rotation of θ_2 is very small in **O1-md** (however, a further rotation seems to start at the very end of the trajectory).

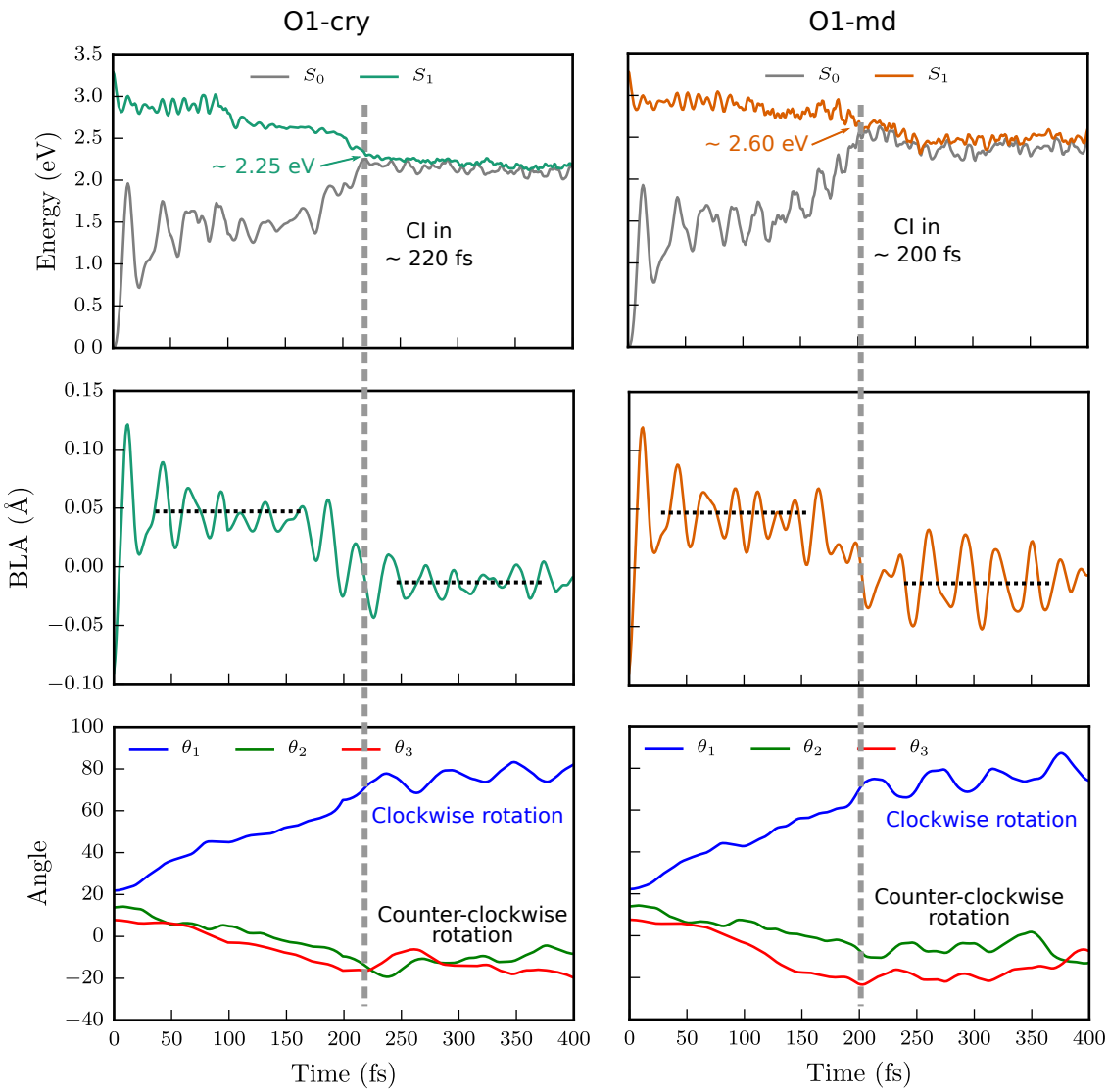


Figure 6.11: Energy, BLA and dihedral angle evolution along the 400 fs trajectory for **O1-cry** (left) and **O1-md** (right).

The dynamics obtained out of the O2 protonated setups are very different. While the trajectory obtained for **O2-cry** is similar to the ones obtained in O1 protonated setups, with a downhill potential energy surface obtained for the S_1 state, **O2-md** is locked in one conformation during the whole trajectory. The structure of this geometry is again in very good agreement with what has been obtained above, with angles θ_1, θ_2 and θ_3 oscillating around their initial values. The dynamics obtained for **O2-cry** is faster than the ones obtained above, and the system reaches the CI in around 160 fs. The CI structure is again in agreement with the one obtained out of excited state optimization.

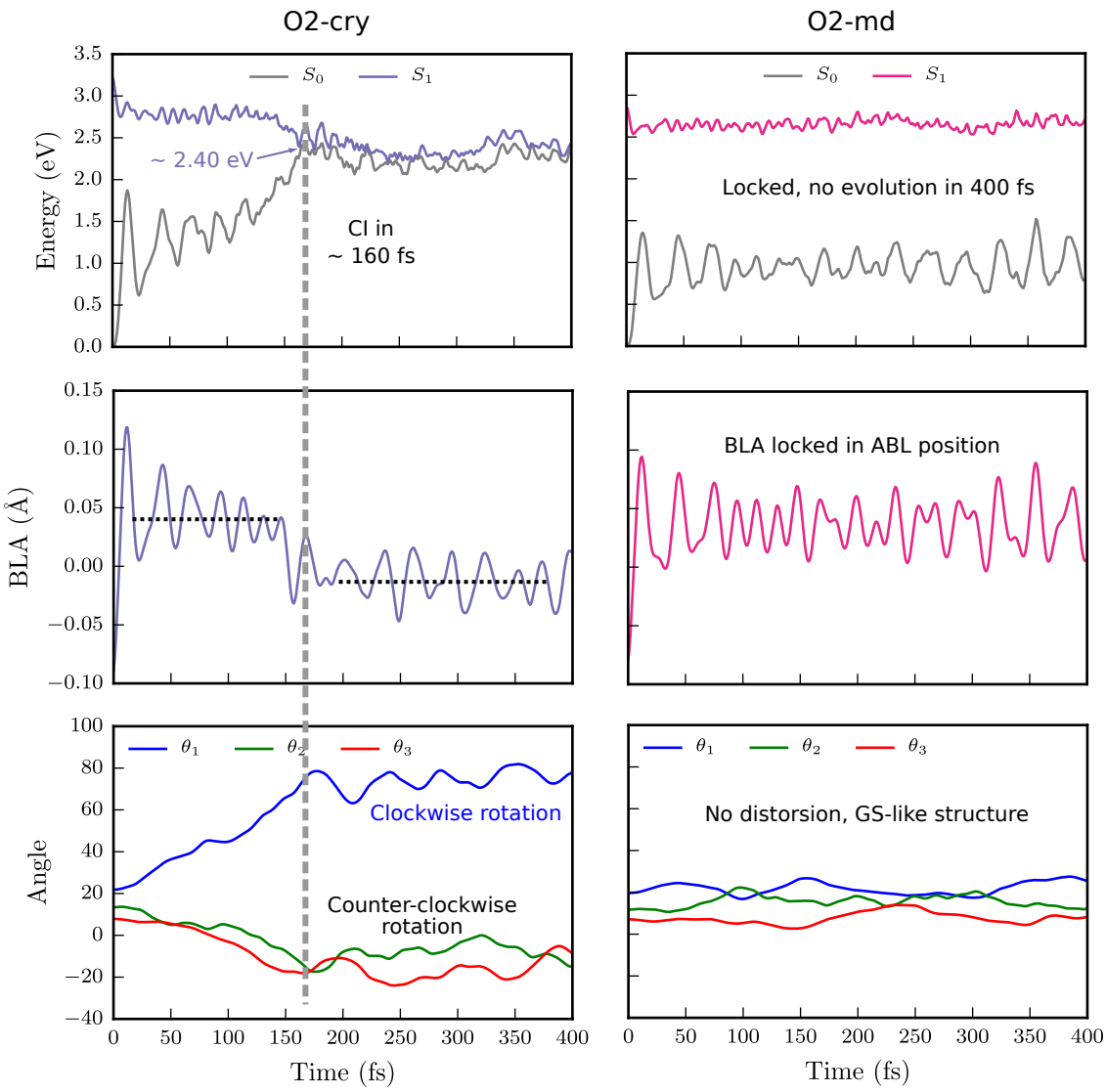


Figure 6.12: Energy, BLA and dihedral angle evolution along the 400 fs trajectory for **O2-cry** (left) and **O2-md** (right).

For all the dynamics that show an evolution (*i.e.*, for **O1-cry**, **O1-md** and **O2-cry**), the evolution of the BLA shows that two patterns can be separated. Before the CI, the BLA oscillates at high positive values (around 0.05 Å). This high BLA is consistent with the ABL geometries isolated for **O1-md**, and for the constrained computations in **O1-cry** and **O2-cry**. This state is reached extremely rapidly (below 50 fs), and its lifetime is found around 150-200 fs, much lower than the experimental lifetime observed for the I intermediate. However, **O2-md** does not start the isomerization as far as 400 fs, which correlates better with experimental data. We are currently working on increasing the time of the simulation to see

if the system stays locked forever, or if an isomerization is triggered. A limit of these simulations is that they are only 0 K trajectories, and a proper thermalization of the system could lead to different pictures. Allowing surface hopping is also a needed improvement of the results presented here.

6.4.3 Conclusion

After showing that protonation of Asp212 lead to correct computations of the vertical absorption of bR, we investigated on the different protonation patterns that could be obtained. This has lead us to four model systems. All these systems have the expected behaviour in the ground state. However, different behaviour are observed during the optimization of their excited state, with two setups leading directly to a CI, while the others lead to ABL geometries. Computation of semi-classical trajectories allows to account for these differences. Indeed, a steep downhill PES was found for the systems that lead directly to a CI, while a flat surface is obtained for the two others. One of the systems investigated, **O2-md**, does not reach a CI during the trajectory computed, contrary to what happens for the three other setups.

Nonetheless, the properties computed out of the three other systems correlate well in term of excited state lifetime, vertical absorption or vertical emission. However, these results just show that the protonation of Asp212 allows the reproduction of the known properties of the photo-physics of the PSB in bR. We have also proven that the structure of the HBN in the quadrupole can largely influence these properties. Validating this protonation pattern requires a more in depth investigation on the structure of the HBN around the chromophore. In particular, infra-red spectroscopy has been shown as a reliable probe for resolving this structure.

6.5 Conclusion and perspectives

In this chapter we have presented possible improvement in the commonly accepted protonation pattern inside the quadrupole structure formed by the PSB and Asp85, Asp212 and Arg82 in bR. We have shown that protonation of Asp212 allows the accurate computation of the vertical absorption, and that it could also capture the main properties of the I excited state intermediate. However, this change of protonation leads to question the hydration switch mechanism proposed by Kandori et al. [144], which require a negatively charged Asp212, at least during the L to M intermediate transition. So far, we have only presented a new model, but much validations are still required.

First, the results presented in this chapter relates to protonation of the model provided by Fujimoto et al., where both oxygen of Asp212 are involved in the HBN. However, as explained in section 6.3, this is not the case in the models directly derived from the 1C3W crystal structure. It is possible that protonating the oxygen not involved in the HBN lead to minor structural changes. The setups involving protonations of Asp212 in 1C3W are under investigation at the time of the writing. Another development would be a proper sampling of the different conformations obtained out of a protonated Asp212 by classical molecular dynamics. This should give representative structures that can directly be compared to experimental data on the structure of the HBN.

A first mandatory step for validating our model is to reproduce the vibrational spectrum of the O−H groups of water molecules. Some computations have been done in this direction, but few technical problems have hampered a conclusive study.

Finally, some data have also been obtained for mutated residues with the help of Léa Rochet, a student from CPE Lyon, who performed (under my co-supervision) modeling and QM/MM optimizations on some previously investigated mutants. The comparison of the results presented in this chapter, results on mutants, and vibrational analysis is expected to give much insight on the new protonation pattern presented.

The results obtained for bR can also lead to accurate modelings of xR, for more in depth studies on the coupling between the carotene molecule and the retinal. Computations of transition dipole moments at different level of theory (TD-DFT, CASSCF and RASSCF) are currently done in collaboration with F. Segatta (University of Bologna), and will be used in excitonic model Hamiltonian to better understand the 2D UV-Vis spectroscopy of this system.

7 Ultraviolet vision: photophysical properties of the unprotonated retinyl Schiff base in the Siberian hamster cone pigment

Published in Theoretical Chemistry Accounts

Link: https://link.springer.com/article/10.1007/s00214-016-1869-x

DOI: 10.1007/s00214-016-1869-x

Résumé Le pigment visuel ultraviolet du hamster sibérien (SHUV) possède, dans l'état adapté à l'obscurité, un chromophore rétinal sous forme de base de Schiff non protonée (USB), qui devient protonée (PSB) après excitation lumineuse dans les premières étapes du photocycle. Alors que le processus de relaxation photochimique de SHUV n'est pas encore très bien compris, on s'attend à des différences sensibles par rapport à la base de Schiff protonée des rhodopsines visuelles. Nous reportons ici une étude des propriétés photophysiques de la base de Schiff non protonée de SHUV (SHUV-USB), fondée sur des méthodes multi-configurationnelles et multi-référentielles dans une approche hybride QM/MM. En comparant les résultats multi-réferentiels et TD-DFT, nous montrons que les deux méthodes prédisent un état excité ionique (S_1), similaire à celui de la rhodopsine, avec un minimum possèdant des longeurs de liaisons égalisées dans la région centrale du polyène. L'analyse des divers états excités dans la région de Franck-Condon et dans celle du minimum de l'état S_1 indique que la relaxation du squelette initiée sur la surface S_1 devrait impliquer un croisement des surfaces S_1 et S_2. Ces résultats forment une base pour des études futures sur les mécanismes de photoisomérisation de SHUV-USB.

Abstract The Siberian hamster ultraviolet (SHUV) visual pigment has an unprotonated Schiff-base (SB) retinyl chromophore in the dark state, which becomes protonated after photoexcitation during the early stages of the photobleaching cycle. While the photochemical relaxation processes of the SHUV

remain poorly understood, they are expected to show significant differences when compared to those of the protonated SB (PSB) chromophore in visual rhodopsin. Here, we report a study of the photophysical properties of the SHUV unprotonated SB (SHUV-USB), based on multiconfigurational and multireference perturbative methods within a hybrid quantum mechanics/molecular mechanics scheme. Comparisons of multireference and time-dependent density functional theory results indicate that both methodologies predict an ionic excited state (S_1), similar to the PSB of rhodopsin, although its minimum has even bond-lengths in the central region of the retinyl polyene chain. The analysis of excited-state manifolds at the Franck–Condon region and S_1 minimum configuration indicates that the skeletal relaxation initiated in the S_1 surface is likely to involve S_1/S_2 surface crossing. These results provide valuable insights for future studies of the SHUV-USB photoisomerization mechanism.

8 Conclusion and perspectives

In this manuscript we have presented a QM/MM modeling of several systems, either artificially obtained — solvated chromophores or bio-engineered proteins — or found in living organisms — bacterial opsins and UV-pigment. In all cases, a combination of classical dynamics sampling, or homology modeling, and of our QM/MM scheme, where the electrostatic coupling is taken into account at the QM level, has been shown to be highly effective for reproducing and interpreting the different spectroscopic properties of the retinal chromophore. This approach is required to accurately model the external and complex environment, either for accounting for its inherent flexibility (in the case of solvents), or for cases where no atomic description is experimentally available.

First, this setup has been applied to the interpretation of the observed difference in photochemistry between an all-*trans* protonated Schiff base (PSB) and its 10-methylated derivative (10Me-PSB) in methanol solution. Indeed, the apparently minimal (and minor) modification of the backbone turns a slow process into an ultra-fast, protein-like excited state decay, as has been experimentally proven by Kukura et al. [121] by means of pump-probe spectroscopy. We have first proven that the dual-peaked stimulated emission observed can be attributed to the sequential population of two species that differ by their bond length alternation in the polyene chain. The first populated has even bond lengths (EBL), while the second has alternated bond lengths (ABL), inverted with respect to the ground state geometry. We have proven that the exchange between these two populations is controlled by the interplay between the S_1 and S_2 excited states in the path that connects the two geometries. Indeed, these two states remain close in energy for PSB, leading to an avoided crossing that translates into a small barrier for the transition. The methylation at position 10 lifts the quasi-degeneracy between the two excited states, and leads to a barrierless decay toward the photoactive ABL geometry. This explains the shorter lifetime observed for the first signal in 10Me-PSB than in PSB. We have also shown that the energy path along the photo-isomerization coordinate is downhill steep in 10Me-PSB, characteristic of a peaked conical intersection, indicating a very effective decay to the ground state and consistent with the very low excited

state lifetime observed experimentally. On the contrary, in PSB, the energy path out the ABL geometry is flatter, slightly uphill, and is characterized by a sloped conical intersection, indicating a less efficient photo-chemistry. All these observations have been confirmed and extended by means of semi-classical dynamics computations, where we have shown that the skeletal relaxation in 10Me-PSB is ultra-fast and leads to a mostly planar ABL geometry, while in PSB the system stays locked in the EBL geometry. We have also shown in 10Me-PSB that the excited state decay follows an space-saving asynchronous double-bicycle pedal motion, as it is the case in bacteriorhodopsin.

We have then studied three rhodopsin mimics (M4, M8 and M10), recently bio-engineered to produce a range of systems that are able to absorb in the wide visible range. We have first shown that a modeling of the hydration of the protein is mandatory for accurately reproducing the vertical absorptions of the different systems. We have also shown that the point mutations of a few amino-acids in the binding pockets have significant differential effects on the different states at the Franck-Condon geometry, that can explain the variability observed in the vertical absorptions, and that affect the photo-induced charge transfer. Combining our modeling with experimental pump-probe spectroscopy, we have then proceeded to investigate on the excited state properties of these three systems. Here again, we have shown that in each system, the same excited state geometries as those found in solvated chromophores (EBL and ABL), are sequentially populated. The path connecting the two geometries is highly affected by the protein scaffold. We have obtained similar paths to the case of solvated chromophores, with the most hydrated systems (M4) showing an avoided crossing during the EBL to ABL transition, while the two others (M8 and M10) have barrierless transitions. In most case, though, the ABL geometry is more stable than EBL. The signals computed out of the three geometries, for all samples in each protein, have provided computed spectra in quantitative agreement with the experimental pump-probe experiments, further validating our approach. Information about the photo-isomerization path has also provided evidence for a tuning of the photochemistry of the systems, with the isomerization being favoured at different locations, as well as a tuning of the speed of the process, with the tuning of the topology of the energy surfaces.

The transition from a ground state geometry (FC) to a reactive geometry (ABL), where the bond lengths are inverted, has to cross a region where the bond lengths are equalized (EBL). In these two parts of the thesis, we have shown that in specific environments, this equalized bond lengths geometry can be stabilized, which gives another method for finely tuning the photo-physical properties of the

chromophore. In particular, this geometry seems to be stabilized when the ionic S_1 and the covalent S_2 states are close to one another, in a way that the transition from EBL to ABL leads to an avoided crossing between the two surfaces. We have shown that such a situation arises in the case of solvated retinal chromophore, or in the case of embedded retinals in proteins that have a large number of water molecule in the vicinity of the chromophore. Thus, we could imagine a new method for tuning the photophysics of the retinal, that involves a control of the solvation shell of the chromophore.

Another part of this thesis has been carried out on natural proteins. In particular, we have modeled two bacterial opsins, the bacteriorhodopsin and the xanthorhodopsin, which both act as light-driven proton pumps. In these systems, the charged iminium of the retinal is involved in a quadrupolar structure with two aspartates and one arginine. In both cases, our modeling resulted in highly blue-shifted vertical absorptions (more than 1 eV) when these three amino-acids are considered in their standard protonation state. However, a much realistic result is obtained when one of the aspartate is protonated. As Asp85 is the primary proton acceptor in the photocycle, and as such is unlikely to be protonated, we decided to study more in depth the protonation of Asp212. Four models obtained by protonating Asp212 have spectroscopic features that are compatible with spectroscopic data, for absorption, fluorescence and stimulated emission. Furthermore, we have shown by means of semi-classical excited state dynamics that the excited state lifetime for three models out of the initial four are consistent with experiments. These promising results are currently further validated by means of mutants modeling, and by comparison with available infra-red data.

If validated, this protonation pattern will be readily applied to the case of xanthorhodopsin, where a carotene molecule is coupled to the photoactive center. The simplicity of this photochemical system — as only two chromophores are involved — makes it a prototype for the study of electronic coupling between chromophores. Therefore, an accurate modeling of xanthorhodopsin could lead to tremendous progress in the description of 2D electronic spectroscopy.

Finally, we have investigated a model of the UV-pigment from the Siberian hamster, which has been obtained by homology modeling. In this protein, the retinal is bound to a nearby lysine *via* an unprotonated Schiff base (USB). We have shown that the S_1 spectrocopic state, corresponding to an ionic $H \rightarrow L$ excitation, relaxes to an EBL minimum. However, the skeletal relaxation upon photo-excitation is likely to involve a crossing with the covalent S_2 state, corresponding to a double $(H \rightarrow L)^2$ excitation. The corresponding minimum has an ABL-like geometry. It is furthermore known experimentally that a proton

transfer from a nearby glutamine occur in the primary steps of the photocycle, to form a protonated Schiff base. Our QM/MM modeling has shown that this proton transfer is unlikely to happen on the S_1 energy surface.

This first modeling has been used as a basis for more advanced studies. We are now investigating on a possible singlet-triplet coupling that could occur on the S_2 covalent surface, where S_2 is found to be near degenerated with T_2. We are also investigating the proton transfer, starting on different models, and with different mechanisms.

These four different projects have lead to accurate descriptions for four different kinds of systems, that will serve as a basis for more in depth study of their photophysics and photochemistry. These projects have also proven the reliability of the COBRAM QM/MM interface for handling a wide range of retinal-containing systems, for characterizing the potential energy surfaces topologies as well as the dynamical behaviour of the chromophore, and for simulating its ultrafast electronic spectroscopy.

9 Conclusion et perspectives (version française)

Nous avons présenté dans ce manuscrit une modélisation QM/MM de plusieurs systèmes, qui sont ou bien obtenus artificiellement — des chromophores en solution ou des protéines obtenues par bio-ingénierie — ou bien trouvés dans des organismes vivants — des ospines bactériennes et un pigment UV. Dans tous les cas, la combinaison entre un échantillonnage en dynamique classique, ou un modèle obtenu par homologie, et notre potentiel QM/MM, où le couplage électrostatique est traité au niveau quantique, s'est montrée hautement efficace pour reproduire et interpréter les différentes propriétés spectroscopiques du chromophore rétinal. Cette approche est nécessaire pour modéliser précisément l'environnement externe et complexe, soit pour prendre en compte sa flexibilité intrinsèque, soit pour des cas où aucune description à l'échelle atomique n'est disponible expérimentalement.

Notre modèle a d'abord été appliqué pour interpéter la différence de comportement photochimique entre une base de Schiff protonée en conformation all-*trans* (PSB) et son dérivé méthylé en position 10 (10Me-PSB) dans une solution de méthanol. En effet cette modification du squelette carboné, en apparence minimale (et mineure), transforme un processus lent en une décomposition de l'état excité ultra-rapide, comme dans une protéine, comme démontré expérimentalement par Kukura et al. par spectroscopie pompe-sonde. Nous avons commencé par montré que l'émission stimulée, composée de deux pics dans le spectre expérimental, peut être attibuée à deux espèces, peuplées séquentiellement lors de l'évolution de l'état excité, qui diffèrent dans l'alternation des longueurs de liaison dans la chaîne polyénique. La première population présente des longueurs de liaison égalisées (*even bond lengths*, EBL), alors que dans la seconde, les longueurs de liaison sont alternées (*alternated bond lengths*, ABL) et inversées par rapport à la géométrie de l'état fondamental. Nous avons prouvé que l'échange entre ces deux populations est contrôlé par l'interaction entre les états excités S_1 et S_2 dans le chemin qui connecte les deux géométries. En effet, ces deux états conservent des énergies proches dans PSB, menant à un

croisement évité qui se traduit par une petite barrière énergétique dans la transition EBL vers ABL. La méthylation en position 10 lève cette quasi-dégénérescence entre les deux états excités, et mène à une transition sans barrière vers la géométrie photoactive ABL. Cela explique le temps de vie plus court observé pour le premier signal dans 10Me-PSB par rapport à PSB. Nous avons aussi montré que le chemin énergétique le long de la coordonnée d'isomérisation est en descente abrupte pour 10Me-PSB, ce qui caractérise une intersection conique "pentue", et indique une décomposition efficace vers l'état fondamental, cohérente avec le temps de vie très court observé expérimentalement. Au contraire, dans PSB, le chemin énergétique partant de la géométrie ABL est plus plat, remontant légèrement, ce qui caractérise une intersection conique "pointue", et indiquant une photochimie bien moins efficace. Toutes ces observations ont été confirmées et étendues au moyen de calculs de dynamiques semi-classiques, où nous avons démontré que la relaxation du squelette de 10Me-PSB est ultra-rapide et mène majoritairement à une géométrie ABL plane, alors que dans PSB, le système reste bloqué dans la géométrie EBL. Enfin, nous avons prouvé que la décomposition de l'état excité dans 10Me-PSB se développe à travers un mécanisme, limité spatialement, de double pédale asynchrone, comme dans la bactériorhodopsine.

Nous avons ensuite étudié trois mimiques de rhodopsine (M4, M8 et M10), modifiées récemment par bio-ingénierie pour produire une gamme de systèmes qui peuvent absorber dans tout le spectre visible. Nous avons d'abord démontré qu'il est indispensable de modéliser explicitement la solvatation de chaque protéine pour reproduire précisément les absorptions verticales expérimentales. Nous avons aussi montré que les mutations ponctuelles de quelques acides aminés dans la poche de liaison du rétinal ont des effets différentiels significatifs sur les différents états (fondamental et excités) à la géométrie de Franck-Condon, qui peuvent expliquer les variations de maximum d'absorption, et qui influent sur le transfert de charge photo-induit. En combinant notre modèle avec des expériences de spectroscopie pompe-sonde, nous avons ensuite étudié les propriétés des états excités de ces trois systèmes. Là encore, nous avons pu démontrer que dans chaque protéine, les mêmes géométries d'état excité que celles trouvées dans le cas des chromophores solvatés (ABL et EBL) sont peuplées séquentiellement. La structure de la protéine influe sensiblement sur le chemin qui connecte ces deux géométries. Nous avons obtenus des chemins avec des caractéristiques similaires à ceux obtenus pour les chromophores solvatés. Les systèmes les plus hydratés (M4) présentent un croisement évité durant la transition de EBL vers ABL, alors que les autres (M8 et M10) ont une transition sans barrière. Dans la plupart des cas, la géométrie ABL est cependant plus stable que EBL. Les signaux calculés à partir des trois géométries (FC, EBL et ABL) pour tous les

échantillons dans chaque protéine, donnent des spectres calculés en accord quantitatif avec les spectres pompe-sonde expérimentaux, ce qui valide encore notre approche. Des informations obtenues sur le chemin de photo-isomérisation montrent enfin une modulation de la photochimie de ces systèmes, avec une isomérisation favorisée à différents endroits du chromophore, ainsi que de la vitesse du processus, à travers une modulation de la topologie des surfaces d'énergie potentielle.

La transition entre une géométrie à l'état fondamental (FC) et une géométrie réactive à l'état excité (ABL), où les longueurs de liaison sont inversées, doit passer par une région dans laquelle les liaisons sont toutes équivalentes (EBL). Dans ces deux parties de la thèse, nous avons démontré que dans des environnements spécifiques, cette géométrie avec des longueurs de liaison égalisées, peut être stabilisée, ce qui donne un autre moyen de moduler finement les propriétés photophysiques du chromophore. En particulier, cette géométrie semble être stabilisée quand l'état ionique S_1 et l'état covalent S_2 sont proches l'un de l'autre, de façon à ce que la transition entre EBL et ABL donne lieu à un croisement évité entre les deux surfaces d'énergie. Nous avons montré que cela se produit dans le cas de chromophore rétinal solvaté, ou bien dans le cas de protéines ayant un grand nombre de molécules d'eau à proximité de ce chromophore. Ainsi, on peut imaginer une nouvelle méthode pour contrôler la photo-physique du rétinal, en jouant sur la couche de solvatation du rétinal.

Une autre partie de cette thèse concerne des protéines naturelles. En particulier, nous avons modélisé deux ospines bactériennes, la bactériorhodopsine et la xanthorhodopsine, qui servent toutes deux de pompes à protons activées par la lumière. Dans ces deux systèmes, l'iminium (chargé) du rétinal est intégré dans une structure quadrupolaire avec deux aspartates et une arginine. Dans les deux cas, notre modèle donne des absorptions verticales très surestimées (plus de 1 eV) quand ces trois acides aminés sont considérés dans leur état de protonation standard. Cependant, des résultats beaucoup plus réalistes sont obtenus lorsqu'un des deux aspartates est protoné. Comme Asp85 est le premier accepteur de proton dans le photocycle, il est peu susceptible d'être protoné. Nous avons donc décidé d'étudier principalement la protonation d'Asp212. Quatre modèles obtenus en protonant Asp212 présentent des caractéristiques compatibles avec les données spectroscopiques pour l'absorption, la fluorescence et l'émission stimulée. De plus, nous avons montré avec des dynamiques semi-classiques à l'état excité que le temps de vie de l'état excité pour trois modèles sur les quatre sont en accord avec les valeurs expérimentales. Ces résultats prometteurs sont actuellement validés par la modélisation de mutants et de propriétés de spectroscopie infra-rouge.

S'il est validée, ce schéma de protonation sera directement appliqué à la xanthorhodopsine, où une molécule de carotène est couplée au centre photoactif. La simplicité de ce système photochimique — seulement deux chromophores sont impliqués — en fait un prototype pour l'étude des couplages électroniques entre chromophores. Par conséquent, une modélisation précise de la xanthorhodopsine devrait mener à des progrès formidables dans la description de la spectroscopie électronique 2D.

Enfin, nous avons étudié un modèle du pigment UV du hamster sibérien, qui a été obtenu par modèlisation par homologie. Dans cette protéine, le rétinal est lié à une lysine proche *via* une base de Schiff non protonée (*unprotonated Schiff base*, USB). Nous avons montré que l'état spectroscopique S_1, qui correspond à une excitation ionique $H \rightarrow L$, relaxe en une géométrie de type EBL. Cependant, la relaxation squelettale suivant la photo-excitation implique certainement un croisement avec l'état covalent S_2, qui correspond à une double excitation $(H \rightarrow L)^2$. La géométrie optimisée correspondante est de type ABL. De plus, les données expérimentales montrent qu'un proton est transféré depuis une glutamine proche dans les premières étapes du photocycle, pour former une base de Schiff protonée. Notre modèle QM/MM montre que ce transfert de proton ne se produit probablement pas sur la surface d'énergie S_1.

Ce premier modèle a été utilisé pour des études plus avancées. Nous étudions à présent la possibilité d'un couplage singulet-triplet qui pourrait se produire sur la surface de l'état covalent S_2, où S_2 est presque dégénéré avec un état triplet T_2. Nous étudions aussi le transfert de proton, à partir de différents modèles, et par différents mécanismes.

Ces quatre différents projets ont donné des modélisations précises de quatre systèmes différents, qui pourront être utilisées pour étudier plus en profondeur leur photophysique et leur photochimie. Ces projets ont aussi prouvé la robustesse de notre interface de QM/MM, COBRAM, pour traiter une large gamme de systèmes qui contiennent du rétinal, pour caractériser la topologie des surfaces d'énergie potentielle ainsi que le comportement dynamique du chromophore, et enfin pour simuler sa spectroscopie électronique ultra-rapide.

Bibliography

1. Blankenship, R. E. *Molecular Mechanisms of Photosynthesis* 2nd ed. 312 pp. (Wiley-Blackwell, 2014) (see p. 19).
2. Palczewski, K. Chemistry and biology of vision. *The Journal of Biological Chemistry* **287,** 1612–1619 (2012) (see pp. 19, 21).
3. Garavelli, M., Bernardi, F., Robb, M. A. & Olivucci, M. Computer simulation of photoinduced molecular motion and reactivity. *International Journal of Photoenergy* **4,** 57–68 (2002) (see pp. 20, 24, 36, 38).
4. Yarkony, D. R. Conical Intersections: The New Conventional Wisdom. *The Journal of Physical Chemistry A* **105,** 6277–6293 (2001) (see pp. 20, 37).
5. Hu, X. & Schulten, K. How nature harvests sunlight. *Physics Today* **50,** 28–34 (1997) (see p. 20).
6. Paterson, M. J., Robb, M. A., Blancafort, L. & DeBellis, A. D. Mechanism of an Exceptional Class of Photostabilizers: A Seam of Conical Intersection Parallel to Excited State Intramolecular Proton Transfer (ESIPT) in o-Hydroxyphenyl-(1,3,5)-triazine. *The Journal of Physical Chemistry A* **109,** 7527–7537 (2005) (see p. 20).
7. Hammes-Schiffer, S. & Soudackov, A. V. Proton-Coupled Electron Transfer in Solution, Proteins, and Electrochemistry. *The Journal of Physical Chemistry B* **112,** 14108–14123 (2008) (see p. 20).
8. Polli, D., Altoè, P., Weingart, O., Spillane, K. M., Manzoni, C., Brida, D., Tomasello, G., Orlandi, G., Kukura, P., Mathies, R. A., Garavelli, M. & Cerullo, G. Conical intersection dynamics of the primary photoisomerization event in vision. *Nature* **467,** 440–443 (2010) (see pp. 21, 23, 29, 30, 41, 86, 90, 92, 97, 99).
9. Yoshizawa, T. & Wald, G. Pre-Lumirhodopsin and the Bleaching of Visual Pigments. *Nature* **197,** 1279–1286 (1963) (see p. 21).

10. Chosrowjan, H., Mataga, N., Shibata, Y., Tachibanaki, S., Kandori, H., Shichida, Y., Okada, T. & Kouyama, T. Rhodopsin Emission in Real Time: A New Aspect of the Primary Event in Vision. *Journal of the American Chemical Society* **120,** 9706–9707 (1998) (see p. 21).

11. Schoenlein, R. W., Peteanu, L. A., Mathies, R. A. & Shank, C. V. The first step in vision: femtosecond isomerization of rhodopsin. *Science* **254,** 412–415 (1991) (see p. 21).

12. Kim, J. E., Tauber, M. J. & Mathies, R. A. Wavelength Dependent Cis-Trans Isomerization in Vision. *Biochemistry* **40,** 13774–13778 (2001) (see p. 21).

13. Johnson, P. J. M., Halpin, A., Morizumi, T., Prokhorenko, V. I., Ernst, O. P. & Miller, R. J. D. Local vibrational coherences drive the primary photochemistry of vision. *Nature Chemistry* **7,** 980–986 (2015) (see p. 21).

14. Johnson, P. J. M., Farag, M. H., Halpin, A., Morizumi, T., Prokhorenko, V. I., Knoester, J., Jansen, T. L. C., Ernst, O. P. & Miller, R. J. D. The Primary Photochemistry of Vision Occurs at the Molecular Speed Limit. *The Journal of Physical Chemistry B* **121,** 4040–4047 (2017) (see p. 21).

15. Schulten, K. & Hayashi, S. in (eds Mohseni, M., Omar, Y., Engel, G. S. & Plenio, M. B.) 237–263 (Cambridge University Press, 2014) (see p. 21).

16. Lanyi, J. K. & Balashov, S. P. Xanthorhodopsin: A bacteriorhodopsin-like proton pump with a carotenoid antenna. *Biochimica et Biophysica Acta* **1777,** 684–688 (2008) (see pp. 21, 153).

17. Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757,** 1012–1018 (2006) (see pp. 21, 151).

18. Mathies, R. A., Cruz, C. H. B., Pollard, W. T. & Shank, C. V. Direct Observation of the Femtosecond Excited-State *cis-trans* Isomerization in Bacteriorhodopsin. *Science* **240,** 777–779 (1988) (see pp. 22, 153).

19. Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S. & Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **114,** 126–163 (2014) (see pp. 22, 95).

20. Rajput, J., Rahbek, D. B., Andersen, L. H., Hirshfeld, A., Sheves, M., Altoè, P., Orlandi, G. & Garavelli, M. Probing and modeling the absorption of retinal protein chromophores in vacuo. *Angewandte Chemie. International Edition in English* **49,** 1790–1793 (2010) (see pp. 22, 41, 95, 125).

21. Nielsen, M. B. Model systems for understanding absorption tuning by opsin proteins. *Chemical Society Reviews* **38,** 913–924 (2008) (see pp. 22, 28).

22. Freedman, K. A. & Becker, R. S. Comparative Investigation of the Photoisomerization of the Protonated and Unprotonated n-Butylamine Schiff Bases of 9-cis-, 11-cis-, 13-cis-, and all-trans Retinals. *Journal of the American Chemical Society* **108,** 1245–1251 (1986) (see p. 22).

23. Hampp, N. A. Bacteriorhodopsin: mutating a biomaterial into an optoelectronic material. *Applied Microbiology and Biotechnology* **53,** 633–639 (2000) (see pp. 23, 95).

24. Kato, H. E., Kamiya, M., Sugo, S., Ito, J., Taniguchi, R., Orito, A., Hirata, K., Inutsuka, A., Yamanaka, A., Maturana, A. D., Ishitani, R., Sudo, Y., Hayashi, S. & Nureki, O. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nature Communications* **6.** doi:10.1038/ncomms8177 (2015) (see p. 23).

25. Sudo, Y., Okazaki, A., Ono, H., Yagasaki, J., Sugo, S., Kamiya, M., Reissig, L., Inoue, K., Ihara, K., Kandori, H., Takagi, S. & Hayashi, S. A Blue-shifted Light-driven Proton Pump for Neural Silencing. *Journal of Biological Chemistry* **288,** 20624–20632 (2013) (see p. 23).

26. Garavelli, M., Celani, P., Bernardi, F., Robb, M. A. & Olivucci, M. The $C_5H_6NH_2^+$ Protonated Shiff Base: An ab Initio Minimal Model for Retinal Photoisomerization. *Journal of the American Chemical Society* **119,** 6891–6901 (1997) (see pp. 23, 41, 72).

27. Cembran, A., González-Luque, R., Altoè, P., Merchán, M., Bernardi, F., Olivucci, M. & Garavelli, M. Structure, Spectroscopy, and Spectral Tuning of the Gas-Phase Retinal Chromophore: The β-Ionone "Handle" and Alkyl Group Effect. *The Journal of Physical Chemistry A* **109,** 6597–6605 (2005) (see pp. 23, 24, 95).

28. Cembran, A., González-Luque, R., Serrano-Andrés, L., Merchán, M. & Garavelli, M. About the intrinsic photochemical properties of the 11-cis retinal chromophore: computational clues for a trap state and a lever effect in Rhodopsin catalysis. *Theoretical Chemistry Accounts* **118,** 173–183 (2007) (see p. 23).

29. Walczak, E., Szefczyk, B. & Andruniów, T. Geometries and Vertical Excitation Energies in Retinal Analogues Resolved at the CASPT2 Level of Theory: Critical Assessment of the Performance of CASSCF, CC2, and DFT Methods. *Journal of Chemical Theory and Computation* **9,** 4915–4927 (2013) (see pp. 23, 41, 157).

30. Hayashi, S. & Ohmine, I. Proton Transfer in Bacteriorhodopsin: Structure, Excitation, IR Spectra, and Potential Energy Surface Analyses by an ab Initio QM/MM Method. *The Journal of Physical Chemistry B* **104,** 10678–10691 (2000) (see pp. 23, 153, 155).

31. Ferré, N. & Olivucci, M. Probing the Rhodopsin Cavity with Reduced Retinal Models at the CASPT2//CASSCF/AMBER Level of Theory. *Journal of the American Chemical Society* **125,** 6868–6869 (2003) (see pp. 23, 28, 41, 72).

32. Wanko, M., Hoffmann, M., Strodel, P., Koslowski, A., Thiel, W., Neese, F., Frauenheim, T. & Elstner, M. Calculating absorption shifts for retinal proteins: computational challenges. *The Journal of Physical Chemistry B* **109,** 3606–3615 (2005) (see p. 23).

33. Fujimoto, K., Hayashi, S., Hasegawa, J.-y. & Nakatsuji, H. Theoretical Studies on the Color-Tuning Mechanism in Retinal Proteins. *Journal of Chemical Theory and Computation* **3,** 605–618 (2007) (see pp. 23, 149, 155).

34. Bravaya, K., Bochenkova, A., Granovsky, A. & Nemukhin, A. An opsin shift in rhodopsin: retinal S0-S1 excitation in protein, in solution, and in the gas phase. *Journal of the American Chemical Society* **129,** 13035–13042 (2007) (see p. 23).

35. Hayashi, S., Tajkhorshid, E. & Schulten, K. Molecular Dynamics Simulation of Bacteriorhodopsin's Photoisomerization Using Ab Initio Forces for the Excited Chromophore. *Biophys. J. Vol.* **85,** 1440–1449 (2003) (see pp. 23, 153).

36. Röhrig, U. F., Guidoni, L. & Rothlisberger, U. Solvent and Protein Effects on the Structure and Dynamics of the Rhodopsin Chromophore. *ChemPhysChem* **6,** 1836–1847 (2005) (see p. 23).

37. Frutos, L. M., Andruniow, T., Santoro, F., Ferre, N. & Olivucci, M. Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 7764–7769 (2007) (see pp. 23, 92).

38. Rivalta, I., Nenov, A. & Garavelli, M. Modelling retinal chromophores photoisomerization: from minimal models in vacuo to ultimate bidimensional spectroscopy in rhodopsins. *Physical Chemistry Chemical Physics* **16,** 16865 (2014) (see pp. 23, 25, 27).

39. Punwong, C., Owens, J. & Martínez, T. J. Direct QM/MM Excited-State Dynamics of Retinal Protonated Schiff Base in Isolation and Methanol Solution. *The Journal of Physical Chemistry B* **119,** 704–714 (2015) (see pp. 23, 85).

40. Roos, B. O. in (ed Lawley, K. P.) 399–445 (Wiley, 1987) (see pp. 23, 41, 46, 69, 101, 156).

41. Andersson, K., Malmqvist, P. & Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics* **96,** 1218–1226 (1992) (see pp. 23, 41, 52, 101, 156).

42. Garavelli, M., Vreven, T., Celani, P., Bernardi, F., Robb, M. A. & Olivucci, M. Photoisomerization Path for a Realistic Retinal Chromophore Model: The Nonatetraeniminium Cation. *Journal of the American Chemical Society* **120,** 1285–1288 (1998) (see p. 23).

43. Garavelli, M., Negri, F. & Olivucci, M. Initial Excited-State Relaxation of the Isolated 11-cis Protonated Schiff Base of Retinal: Evidence for in-Plane Motion from ab Initio Quantum Chemical Simulation of the Resonance Raman Spectrum. *Journal of the American Chemical Society* **121,** 1023–1029 (1999) (see p. 23).

44. Garavelli, M. Computational organic photochemistry: strategy, achievements and perspectives. *Theoretical Chemistry Accounts* **116,** 87–105 (2006) (see pp. 23, 25, 36, 41).

45. González-Luque, R., Garavelli, M., Bernardi, F., Merchán, M., Robb, M. A. & Olivucci, M. Computational evidence in favor of a two-state, two-mode model of the retinal chromophore photoisomerization. *Proceedings of the National Academy of Sciences* **97,** 9379–9384 (2000) (see pp. 23, 41).

46. Haran, G., Morlino, E. A., Matthes, J., Callender, R. H. & Hochstrasser, R. M. Femtosecond Polarized Pump-Probe and Stimulated Emission Spectroscopy of the Isomerization Reaction of Rhodopsin. *The Journal of Physical Chemistry A* **103,** 2202–2207 (1999) (see p. 24).

47. Vreven, T., Bernardi, F., Garavelli, M., Olivucci, M., Robb, M. A. & Schlegel, H. B. Ab Initio Photoisomerization Dynamics of a Simple Retinal Chromophore Model. *Journal of the American Chemical Society* **119,** 12687–12688 (1997) (see p. 24).

48. Cembran, A., Bernardi, F., Olivucci, M. & Garavelli, M. The retinal chromophore/chloride ion pair: structure of the photoisomerization path and interplay of charge transfer and covalent states. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6255–6260 (2005) (see pp. 26, 95).

49. Cembran, A., Bernardi, F., Olivucci, M. & Garavelli, M. Counterion Controlled Photoisomerization of Retinal Chromophore Models: a Computational Investigation. *Journal of the American Chemical Society* **126,** 16018–16037 (2004) (see pp. 26, 27).

50. Tomasello, G., Olaso-González, G., Altoè, P., Stenta, M., Serrano-Andrés, L., Merchán, M., Orlandi, G., Bottoni, A. & Garavelli, M. Electrostatic control of the photoisomerization efficiency and optical properties in visual pigments: on the role of counterion quenching. *Journal of the American Chemical Society* **131,** 5172–5186 (2009) (see pp. 28, 29, 95).

51. Okada, T., Sugihara, M., Bondar, A.-N., Elstner, M., Entel, P. & Buss, V. The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure. *Journal of Molecular Biology* **342,** 571–583 (2004) (see p. 28).

52. Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 A resolution. *Journal of Molecular Biology* **291,** 899–911 (1999) (see pp. 28, 152, 155).

53. Martínez, T. J. Insights for Light-Driven Molecular Devices from Ab Initio Multiple Spawning Excited-State Dynamics of Organic and Biological Chromophores. *Accounts of Chemical Research* **39,** 119–126 (2006) (see p. 28).

54. Ferré, N., Cembran, A., Garavelli, M. & Olivucci, M. Complete-active-space self-consistent-field/Amber parameterization of the Lys296–retinal–Glu113 rhodopsin chromophore-counterion system. *Theoretical Chemistry Accounts* **112,** 335–341 (2004) (see pp. 28, 41).

55. Weingart, O., Altoè, P., Stenta, M., Bottoni, A., Orlandi, G. & Garavelli, M. Product formation in rhodopsin by fast hydrogen motions. *Physical Chemistry Chemical Physics* **13,** 3645–3648 (2011) (see p. 29).

56. Warshel, A., Chu, Z. T. & Hwang, J. .-.-K. The dynamics of the primary event in rhodopsins revisited. *Chemical Physics* **158,** 303–314 (1991) (see p. 29).

57. Warshel, A. Bicycle-pedal model for the first step in the vision process. *Nature* **260,** 679–683 (1976) (see p. 29).

58. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103,** 227–249 (1976) (see pp. 29, 56).

59. Altoè, P., Stenta, M., Bottoni, A. & Garavelli, M. A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. *Theoretical Chemistry Accounts* **118,** 219–240 (2007) (see pp. 29, 38, 56–58, 69, 101, 156).

60. Andruniów, T., Ferré, N. & Olivucci, M. Structure, initial excited-state relaxation, and energy storage of rhodopsin resolved at the multiconfigurational perturbation theory level. *Proceedings of the National Academy of Sciences* **101,** 17908–17913 (2004) (see p. 29).

61. Weingart, O. The role of HOOP-modes in the ultrafast photo-isomerization of retinal models. *Chemical Physics. Electron Correlation and Molecular Dynamics for Excited States and Photochemistry* **349,** 348–355 (2008) (see p. 29).

62. Kukura, P., McCamant, D. W., Yoon, S., Wandschneider, D. B. & Mathies, R. A. Structural Observation of the Primary Isomerization in Vision with Femtosecond-Stimulated Raman. *Science* **310,** 1006–1009 (2005) (see p. 29).

63. Born, M. & Huang, K. *Dynamical Theory of Crystal Lattices.* (Oxford University Press, 1968) (see p. 35).

64. Malhado, J. P., Bearpark, M. J. & Hynes, J. T. Non-adiabatic dynamics close to conical intersections and the surface hopping perspective. *Frontiers in Chemistry* **2.** doi:10.3389/fchem.2014.00097 (2014) (see pp. 35, 39).

65. Martinez, T. J. Physical chemistry: Seaming is believing. *Nature* **467,** 412–413 (2010) (see p. 37).

66. Keal, T. W., Koslowski, A. & Thiel, W. Comparison of algorithms for conical intersection optimisation using semiempirical methods. *Theoretical Chemistry Accounts* **118,** 837–844 (2007) (see p. 38).

67. Bearpark, M. J., Robb, M. A. & Schlegel, H. B. A direct method for the location of the lowest energy point on a potential surface crossing. *Chemical Physics Letters* **223,** 269–274 (1994) (see pp. 38, 71, 102).

68. Tully, J. C. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics* **93,** 1061–1071 (1990) (see pp. 39, 40, 71).

69. Tully, J. C. Perspective: Nonadiabatic dynamics theory. *The Journal of Chemical Physics* **137,** 22A301 (2012) (see p. 39).

70. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136,** B864–B871 (1964) (see p. 41).

71. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140,** A1133–A1138 (1965) (see p. 41).

72. Runge, E. & Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **52,** 997–1000 (1984) (see p. 41).

73. Jacquemin, D., Perpète, E. A., Scuseria, G. E., Ciofini, I. & Adamo, C. TD-DFT Performance for the Visible Absorption Spectra of Organic Dyes: Conventional versus Long-Range Hybrids. *Journal of Chemical Theory and Computation* **4,** 123–135 (2008) (see p. 41).

74. Valsson, O., Filippi, C. & Casida, M. E. Regarding the use and misuse of retinal protonated Schiff base photochemistry as a test case for time-dependent density-functional theory. *The Journal of Chemical Physics* **142,** 144104 (2015) (see p. 41).

75. Demoulin, B., El-Tahawy, M. M. T., Nenov, A., Garavelli, M. & Bahers, T. L. Intramolecular photo-induced charge transfer in visual retinal chromophore mimics: electron density-based indices at the TD-DFT and post-HF levels. *Theoretical Chemistry Accounts* **135,** 96 (2016) (see p. 41).

76. Le Bahers, T., Adamo, C. & Ciofini, I. A Qualitative Index of Spatial Extent in Charge-Transfer Excitations. *Journal of Chemical Theory and Computation* **7,** 2498–2506 (2011) (see p. 41).

77. Roos, B. O., Lindh, R., Malmqvist, P.-Å., Veryazov, V. & Widmark, P.-O. *Multiconfigurational Quantum Chemistry* (John Wiley & Sons, Inc., 2016) (see pp. 41, 46).

78. Roos, B. O., Andersson, K., Fulscher, M. P., Malmqvist, P.-A., Serrano-Andres, L., Pierloot, K. & Merchan, M. in (eds Prigogine, I. & Rice, S. A.) 219–330 (John Wiley & Sons, Inc., 1996) (see pp. 41, 52).

79. Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J. & Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *The Journal of Physical Chemistry* **94,** 5483–5488 (1990) (see pp. 41, 52).

80. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry* (Dover, 1982) (see pp. 42, 43, 51).

81. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics* **23,** 69–89 (1951) (see p. 44).

82. Roos, B. O., Taylor, P. R. & Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics* **48,** 157–173 (1980) (see p. 46).

83. Veryazov, V., Malmqvist, P. Å. & Roos, B. O. How to select active space for multiconfigurational quantum chemistry? *International Journal of Quantum Chemistry* **111,** 3329–3338 (2011) (see p. 47).

84. Malmqvist, P. A., Rendell, A. & Roos, B. O. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *The Journal of Physical Chemistry* **94,** 5477–5482 (1990) (see p. 47).

85. Ma, D., Li Manni, G. & Gagliardi, L. The generalized active space concept in multiconfigurational self-consistent field methods. *The Journal of Chemical Physics* **135,** 044128 (2011) (see p. 47).

86. Löwdin, P.-O. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Physical Review* **97,** 1474–1489 (1955) (see p. 48).

87. Cohen-Tannoudji, C. *Mécanique quantique* (Hermann, 1997) (see p. 49).

88. Forsberg, N. & Malmqvist, P.-Å. Multiconfiguration perturbation theory with imaginary level shift. *Chemical Physics Letters* **274,** 196–204 (1997) (see p. 53).

89. Ghigo, G., Roos, B. O. & Malmqvist, P.-Å. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chemical Physics Letters* **396,** 142–149 (2004) (see p. 54).

90. Zobel, J. P., Nogueira, J. J. & González, L. The IPEA dilemma in CASPT2. *Chemical Science* **8,** 1482–1499 (2017) (see pp. 54, 70).

91. Finley, J., Malmqvist, P.-Å., Roos, B. O. & Serrano-Andrés, L. The multi-state CASPT2 method. *Chemical Physics Letters* **288,** 299–306 (1998) (see p. 54).

92. Brunk, E. & Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chemical Reviews* **115,** 6217–6263 (2015) (see p. 54).

93. Ochsenfeld, C., Kussmann, J. & Koziol, F. Ab Initio NMR Spectra for Molecular Systems with a Thousand and More Atoms: A Linear-Scaling Method. *Angewandte Chemie International Edition* **43,** 4485–4489 (2004) (see p. 54).

94. Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **48,** 1198–1229 (2009) (see pp. 55–58).

95. Sherwood, P. *Hybrid Quantum Mechanics/Molecular Mechanics Approaches* in *NIC Series.* **3:** *Modern Methods and Algorithms of Quantum Chemistry* (ed Grotendorst, J.) (John von Neumann Institute for Computing, Jülich, 2000), 285–305 (see pp. 55, 57).

96. Ponder, J. W. & Case, D. A. in *Advances in Protein Chemistry* 27–85 (Academic Press, 2003). doi:10.1016/S0065-3233(03)66002-X (see p. 55).

97. Théry, V., Rinaldi, D., Rivail, J.-L., Maigret, B. & Ferenczy, G. G. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *Journal of Computational Chemistry* **15,** 269–282 (1994) (see p. 56).

98. Pu, J., Gao, J. & Truhlar, D. G. Generalized Hybrid Orbital (GHO) Method for Combining Ab Initio Hartree-Fock Wave Functions with Molecular Mechanics. *The Journal of Physical Chemistry A* **108,** 632–650 (2004) (see p. 56).

99. Singh, U. C. & Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl- exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry* **7,** 718–730 (1986) (see p. 56).

100. Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **11,** 700–733 (1990) (see p. 56).

101. Lin, H. & Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **117,** 185–199 (2007) (see pp. 56, 57).

102. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25,** 1157–1174 (2004) (see pp. 57, 62, 69, 100).

103. Vreven, T. & Morokuma, K. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Annual Reports in Computational Chemistry* **2,** 35–51 (2006) (see p. 57).

104. Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W. & Ren, P. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **9,** 4046–4063 (2013) (see p. 58).

105. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe,

D. R., Mathiews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A. & Kollman, P. A. *AMBER 12* San Francisco, 2012 (see pp. 62, 69).

106. Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X. & Kollman, P. A. *AmberTools 14* San Francisco, 2014 (see pp. 62, 100).

107. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25,** 247–260 (2006) (see pp. 62, 69, 100).

108. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. E. & Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65,** 712–725 (2006) (see pp. 62, 69, 70, 100, 156).

109. Aquilante, F., Autschbach, J., Carlson, R. K., Chibotaru, L. F., Delcey, M. G., De Vico, L., Galván, I. F., Ferré, N., Frutos, M. L., Gagliardi, L., Garavelli, M., Giussiani, A., Hoyer, C. E., Li Manni, G., Lischka, H., Ma, D., Malmqvist, P. ., Müller, T., Nenov, A., Olivucci, M., Pedersen, T. B., Peng, D., Plasser, F., Pritchard, B., Reiher, M., Rivalta, I., Schapiro, I., Segarra-Martí, J., Stenrup, M., Truhlar, D. G., Ungur, L., Valentini, A., Vancoillie, S., Veryazov, V., Vysotskiy, V. P., Weingart, O., Zapata, F. & Lindh, R. MOLCAS 8: New Capabilities for Multiconfigurational Quantum Chemical Calculations across the Periodic Table. *Journal of Computational Chemistry* **37,** 506–541 (2016) (see pp. 62, 70, 102, 156).

110. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R.,

Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J. & Fox, D. J. *Gaussian 09 Revision A.1* 2009 (see pp. 62, 70).

111. Haran, G., Wynne, K., Xie, A., He, Q., Chance, M. & Hochstrasser, R. M. Excited state dynamics of bacteriorhodopsin revealed by transient stimulated emission spectra. *Chemical Physics Letters* **261,** 389–395 (1996) (see pp. 67, 83, 153).

112. Hasson, K. C., Gai, F. & Anfinrud, P. A. The photoisomerization of retinal in bacteriorhodopsin: Experimental evidence for a three-state model. *Biophysics (Oxf)*. **93,** 15124–15129 (1996) (see pp. 67, 83, 153).

113. Ruhman, S., Hou, B., Friedman, N., Ottolenghi, M. & Sheves, M. Following Evolution of Bacteriorhodopsin in Its Reactive Excited State via Stimulated Emission Pumping. *Journal of the American Chemical Society* **124,** 8854–8858 (2002) (see pp. 67, 83, 153, 163).

114. Wand, A., Loevsky, B., Friedman, N., Sheves, M. & Ruhman, S. Probing Ultrafast Photochemistry of Retinal Proteins in the Near-IR: Bacteriorhodopsin and Anabaena Sensory Rhodopsin vs Retinal Protonated Schiff Base in Solution. *The Journal of Physical Chemistry B* **117,** 4670–4679 (2013) (see pp. 67, 163).

115. Hamm, P., Zurek, M., Röschinger, T., Patzelt, H., Oesterhelt, D. & Zinth, W. Femtosecond spectroscopy of the photoisomerisation of the protonated Schiff base of all-trans retinal. *Chemical Physics Letters* **263,** 613–621 (1996) (see p. 67).

116. Zgrablić, G., Voïtchovsky, K., Kindermann, M., Haacke, S. & Chergui, M. Ultrafast Excited State Dynamics of the Protonated Schiff Base of All-trans Retinal in Solvents. *Biophysical Journal* **88,** 2779–2788 (2005) (see p. 67).

117. Bismuth, O., Friedman, N., Sheves, M. & Ruhman, S. Photochemical dynamics of all-trans retinal protonated Schiff-base in solution: Excitation wavelength dependence. *Chemical Physics* **341,** 267–275 (2007) (see pp. 67, 83).

118. Zgrablić, G., Haacke, S. & Chergui, M. Heterogeneity and Relaxation Dynamics of the Photoexcited Retinal Schiff Base Cation in Solution. *The Journal of Physical Chemistry B* **113,** 4384–4393 (2009) (see p. 67).

119. Zgrablić, G., Novello, A. M. & Parmigiani, F. Population branching in the conical intersection of the retinal chromophore revealed by multipulse ultrafast optical spectroscopy. *Journal of the American Chemical Society* **134,** 955–961 (2012) (see p. 67).

120. Bismuth, O., Friedman, N., Sheves, M. & Ruhman, S. Photochemistry of a Retinal Protonated Schiff-Base Analogue Mimicking the Opsin Shift of Bacteriorhodopsin. *The Journal of Physical Chemistry B* **111,** 2327–2334 (2007) (see p. 67).

121. Sovdat, T., Bassolino, G., Liebel, M., Schnedermann, C., Fletcher, S. P. & Kukura, P. Backbone modification of retinal induces protein-like excited state dynamics in solution. *Journal of the American Chemical Society* **134,** 8318–8320 (2012) (see pp. 67, 68, 87, 178).

122. Bassolino, G., Sovdat, T., Liebel, M., Schnedermann, C., Odell, B., Claridge, T. D. W., Kukura, P. & Fletcher, S. P. Synthetic control of retinal photochemistry and photophysics in solution. *Journal of the American Chemical Society* **136,** 2650–2658 (2014) (see pp. 67, 68).

123. Bassolino, G., Sovdat, T., Soares Duarte, A., Lim, J. M., Schnedermann, C., Liebel, M., Odell, B., Claridge, T. D. W., Fletcher, S. P. & Kukura, P. Barrierless Photoisomerization of 11-cis Retinal Protonated Schiff Base in Solution. *Journal of the American Chemical Society* **137,** 12434–12437 (2015) (see pp. 67, 68).

124. Rintelman, J. M., Adamovic, I., Varganov, S. & Gordon, M. S. Multireference second-order perturbation theory: How size consistent is “almost size consistent”. *The Journal of Chemical Physics* **122,** 044105 (2005) (see p. 82).

125. Altoè, P., Cembran, A., Olivucci, M. & Garavelli, M. Aborted double bicycle-pedal isomerization with hydrogen bond breaking is the primary event of bacteriorhodopsin proton pumping. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20172–20177 (2010) (see pp. 86, 90, 153, 157, 160).

126. Lindvold, L. & Ramanujam, P. The use of bacteriorhodopsin in optical processing: A review. *Journal of Scientific and Industrial Research* **54,** 55–66 (1995) (see p. 95).

127. Sasaki, J., Brown, L., Chon, Y., Kandori, H., Maeda, A., Needleman, R. & Lanyi, J. Conversion of bacteriorhodopsin into a chloride ion pump. *Science* **269** (1995) (see pp. 95, 163).

128. Inoue, K., Nomura, Y. & Kandori, H. Asymmetric Functional Conversion of Eubacterial Light-driven Ion Pumps. *Journal of Biological Chemistry* **291,** 9883–9893 (2016) (see p. 95).

129. Kawanabe, A., Furutani, Y., Jung, K.-H. & Kandori, H. Engineering an Inward Proton Transport from a Bacterial Sensor Rhodopsin. *Journal of the American Chemical Society* **131,** 16439–16444 (2009) (see p. 95).

130. Inoue, K., Tsukamoto, T., Shimono, K., Suzuki, Y., Miyauchi, S., Hayashi, S., Kandori, H. & Sudo, Y. Converting a Light-Driven Proton Pump into a Light-Gated Proton Channel. *Journal of the American Chemical Society* **137,** 3291–3299 (2015) (see p. 95).

131. Vasileiou, C., Vaezeslami, S., Crist, R. M., Rabago-Smith, M., Geiger, J. H. & Borhan, B. Protein design: reengineering cellular retinoic acid binding protein II into a rhodopsin protein mimic. *Journal of the American Chemical Society* **129,** 6140–6148 (2007) (see p. 95).

132. Lee, K. S. S., Berbasova, T., Vasileiou, C., Jia, X., Wang, W., Choi, Y., Nossoni, F., Geiger, J. H. & Borhan, B. Probing Wavelength Regulation with an Engineered Rhodopsin Mimic and a C15-Retinal Analogue. *Chempluschem* **77,** 273–276 (2012) (see p. 96).

133. Wang, W., Nossoni, Z., Berbasova, T., Watson, C. T., Yapici, I., Lee, K. S. S., Vasileiou, C., Geiger, J. H. & Borhan, B. Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science* **338,** 1340–1343 (2012) (see pp. 96, 98).

134. Cheng, C., Kamiya, M., Uchida, Y. & Hayashi, S. Molecular Mechanism of Wide Photoabsorption Spectral Shifts of Color Variants of Human Cellular Retinol Binding Protein II. *Journal of the American Chemical Society* **137,** 13362–13370 (2015) (see pp. 96, 103, 147).

135. Rivalta, I., Nenov, A., Cerullo, G., Mukamel, S. & Garavelli, M. Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach. *International Journal of Quantum Chemistry* **114,** 85–93 (2014) (see pp. 97, 99).

136. Huntress, M. M., Gozem, S., Malley, K. R., Jailaubekov, A. E., Vasileiou, C., Vengris, M., Geiger, J. H., Borhan, B., Schapiro, I., Larsen, D. S. & Olivucci, M. Toward an understanding of the retinal chromophore in rhodopsin mimics. *The Journal of Physical Chemistry B* **117,** 10053–10070 (2013) (see p. 103).

137. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **9,** 3878–3888 (2013) (see p. 104).

138. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32,** 2319–2327 (2011) (see p. 104).

139. Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., Domański, J., Dotson, D. L., Buchoux, S., Kenney, I. M. & Beckstein, O. *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations* in *Proceedings of the 15th Python in Science Conference* (eds Benthall, S. & Rostrup, S.) (2016), 98–105 (see p. 104).

140. Melaccio, F., del Carmen Marín, M., Valentini, A., Montisci, F., Rinaldi, S., Cherubini, M., Yang, X., Kato, Y., Stenrup, M., Orozco-Gonzalez, Y., Ferré, N., Luk, H. L., Kandori, H. & Olivucci, M. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *Journal of Chemical Theory and Computation* **12,** 6020–6034 (2016) (see p. 148).

141. Lanyi, J. K. Understanding Structure and Function in the Light-Driven Proton Pump Bacteriorhodopsin. *Journal of Structural Biology* **124,** 164–178 (1998) (see p. 151).

142. Kandori, H. Role of internal water molecules in bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* doi:10.1016/S0005-2728(00)00138-9 (2000) (see pp. 153, 163).

143. Mikihiro, S., Tanimoto, T. & Kandori, H. Water Molecules in the Schiff Base Region of Bacteriorhodopsin. *Journal of the American Chemical Society* **125,** 13312–13313 (2003) (see p. 153).

144. Tanimoto, T., Furutani, Y. & Kandori, H. Structural changes of water in the Schiff base region of bacteriorhodopsin: Proposal of a hydration switch model. *Biochemistry* **42,** 2300–2306 (2003) (see pp. 153, 174).

145. Kandori, H. Hydration switch model for the proton transfer in the Schiff base region of bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* doi:10.1016/j.bbabio.2004.03.015 (2004) (see pp. 153, 163).

146. Shibata, M. & Kandori, H. FTIR Studies of Internal Water Molecules in the Schiff Base Region of Bacteriorhodopsin. *Biochemistry* **44,** 7406–7413 (2005) (see p. 153).

147. Schenkl, S., Portuondo, E., Zgrabli, G., Chergui, M., Haacke, S., Friedman, N. & Sheves, M. Ultrafast energy relaxation in bacteriorhodopsin studied by time-integrated fluorescence. *Phys. Chamistry Chem. Phys.* **4,** 5020–5024 (2002) (see p. 153).

148. Ye, T., Friedman, N., Gat, Y., Atkinson, G. H., Sheves, M., Ottolenghi, M. & Ruhman, S. On the Nature of the Primary Light-Induced Events in Bacteriorhodopsin: Ultrafast Spectroscopy of Native and C 13 =C 14 Locked Pigments. *The Journal of Physical Chemistry B* **103,** 5122–5130 (1999) (see p. 153).

149. Szymczak, J. J., Barbatti, M. & Lischka, H. Is the Photoinduced Isomerization in Retinal Protonated Schiff Bases a Single-or Double-Torsional Process? *The Journal of Physical Chemistry A* **113,** 11907–11918 (2009) (see p. 153).

150. Luecke, H., Schobert, B., Stagno, J., Imasheva, E. S., Wang, J. M., Balashov, S. P. & Lanyi, J. K. Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16561–16565 (2008) (see pp. 153, 154, 156).

151. Zhu, J., Gdor, I., Smolensky, E., Friedman, N., Sheves, M. & Ruhman, S. Photoselective Ultrafast Investigation of Xanthorhodopsin and Its Carotenoid Antenna Salinixanthin. *The Journal of Physical Chemistry B* **114,** 3038–3045 (2010) (see p. 153).

152. Koganov Smolensky, E., Brumfeld, V., Friedman, N. & Sheves, M. Origin of Circular Dichroism of Xanthorhodopsin. A Study with Artificial Pigments. *The Journal of Physical Chemistry B* **119,** 456–464 (2014) (see p. 153).

153. Hayashi, S., Tajkhorshid, E., Pebay-Peyroula, E., Royant, A., Landau, E. M., Navarro, J. & Schulten, K. Structural Determinants of Spectral Tuning in Retinal Proteins – Bacteriorhodopsin vs Sensory Rhodopsin II. *The Journal of Physical Chemistry B* **105,** 10124–10131 (2001) (see p. 155).

154. Hayashi, S., Tajkhorshid, E., Kandori, H. & Schulten, K. Role of hydrogen-bond network in energy storage of bacteriorhodopsin's light-driven proton pump revealed by ab initio normal-mode analysis. *Journal of the American Chemical Society.* doi:10.1021/ja047506s (2004) (see p. 155).

155. Luecke, H., Richter, H.-T. & Lanyi, J. K. Proton Transfer Pathways in Bacteriorhodopsin at 2.3 Angstrom Resolution. *Science* **280,** 1934–1937 (1998) (see p. 155).

156. Grigorieff, N., Ceska, T., Downing, K., Baldwin, J. & Henderson, R. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *Journal of Molecular Biology* **259,** 393–421 (1996) (see p. 155).

157. Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J. P., Landau, E. M. & Pebay-Peyroula, E. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* **7,** 909–917 (1999) (see p. 155).

158. Sasaki, J., Lanyi, J. K., Needleman, R., Yoshizawa, T. & Maeda, A. Complete Identification of C=O Stretching Vibrational Bands of Protonated Aspartic Acid Residues in the Difference Infrared Spectra of M and N Intermediates versus Bacteriorhodopsint. *Biochemistry* **33,** 178–3 (1994) (see p. 155).

159. Brown, L. S., Sasaki, J., Kandori, H., Maeda, A., Needleman, R. & Lanyi, J. K. Glutamic acid 204 is the terminal proton release group at the extracellular surface of bacteriorhodopsin. *Journal of Biological Chemistry* **270,** 27122–6 (1995) (see p. 155).

160. Furutani, Y. & Kandori, H. Hydrogen-bonding changes of internal water molecules upon the actions of microbial rhodopsins studied by FTIR spectroscopy. *Biochim. Biophys. Acta - Bioenerg.* **1837,** 598–605 (2014) (see p. 163).

161. Rothschild, K., Gray, D., Mogi, T., Marti, T., Braiman, M., Stern, L. & Khorana, H. Vibrational spectroscopy of bacteriorhodopsin mutants: Chromophore isomerization perturbs tryptophan-86. *Biochemistry* **28,** 7052–7059 (1989) (see p. 163).

162. Needleman, R., Chang, M., Ni, B., Váró, G., Fornés, J., White, S. H. & Lanyi, J. K. Properties of Asp212 to Asn Bacteriorhodopsin Suggest That Asp212 and Asp85 Both Participate in a Counterion and Proton Acceptor Complex near the Schiff Base. *The Journal of Biological Chemistry* **266,** 11478–11484 (1991) (see p. 163).

163. Subramanian, S., Greenhalgh, D. A. & Khorana, H. G. Aspartic Acid 85 in Bacteriorhodopsin Functions Both as Proton Acceptor and Negative Counterion to the Schiff Base. *Journal of Biological Chemistry* **267,** 25730–25733 (1992) (see p. 163).

164. Shibata, M., Yoshitsugu, M., Mizuide, N., Ihara, K. & Kandori, H. Halide inding by the D212N Mutant of Bacteriorhodopsin Affects Hydrogen Bonding of Water in the Active Site. *Biochemistry* **46,** 7525–7535 (2007) (see p. 163).